Journal of Medical Internet Research

Journal Impact Factor (JIF) (2023): 5.8
Volume 23 (2021), Issue 4 ISSN 1438-8871 Editor in Chief: Gunther Eysenbach, MD, MPH

Contents

Reviews

Viewpoints

Tutorial

Original Papers

Corrigenda and Addendas

Corrigenda and Agenda

Short Paper

Review

Social Media and Health Care, Part I: Literature Review of Social Media Use by Health Care Providers

Deema Farsi[1], BDS, MSci

Department of Pediatric Dentistry, Faculty of Dentistry, King Abdulaziz University, Jeddah, Saudi Arabia

Corresponding Author:
Deema Farsi, BDS, MSci
Department of Pediatric Dentistry
Faculty of Dentistry
King Abdulaziz University
PO Box 80200
Jeddah, 21589
Saudi Arabia
Phone: 966 16402000 ext 20388
Email: dfarsi@kau.edu.sa

Related Article:

Companion article: https://www.jmir.org/2022/1/e30379

Abstract

Background: As the world continues to advance technologically, social media (SM) is becoming an essential part of billions of people's lives worldwide and is affecting almost every industry imaginable. As the world is becoming more digitally oriented, the health care industry is increasingly visualizing SM as an important channel for health care promotion, employment, recruiting new patients, marketing for health care providers (HCPs), building a better brand name, etc. HCPs are bound to ethical principles toward their colleagues, patients, and the public in the digital world as much as in the real world.

Objective: This review aims to shed light on SM use worldwide and to discuss how it has been used as an essential tool in the health care industry from the perspective of HCPs.

Methods: A literature review was conducted between March and April 2020 using MEDLINE, PubMed, Google Scholar, and Web of Science for all English-language medical studies that were published since 2007 and discussed SM use in any form for health care. Studies that were not in English, whose full text was not accessible, or that investigated patients' perspectives were excluded from this part, as were reviews pertaining to ethical and legal considerations in SM use.

Results: The initial search yielded 83 studies. More studies were included from article references, and a total of 158 studies were reviewed. SM uses were best categorized as health promotion, career development or practice promotion, recruitment, professional networking or destressing, medical education, telemedicine, scientific research, influencing health behavior, and public health care issues.

Conclusions: Multidimensional health care, including the pairing of health care with SM and other forms of communication, has been shown to be very successful. Striking the right balance between digital and traditional health care is important.

(J Med Internet Res 2021;23(4):e23205) doi:10.2196/23205

KEYWORDS

social media; social networking; internet; health care; COVID-19; research activity; medical education; telemedicine; mobile phone

Introduction

Background

A key characteristic of being human is the ability and desire for social networking. Over the ages, humanity has thrived in social communities in which members shared knowledge, opinions, and experiences, empowered by a sense of belonging. As the world continues to advance in terms of technology, social media (SM)—defined as "a group of Internet-based applications (apps) that allow the creation and exchange of user-generated content"—is becoming an essential part of billions of people's lives worldwide and is affecting almost every industry imaginable [1]. The definition of SM (the "read, write web," "Web 2.0," or "social networking") is constantly evolving [2]. The Merriam-Webster Dictionary defines it as "any form of electronic communication through which users create web-based communities to share information, personal messages, ideas, and other content such as photos and videos" [3]. SM is considered one of the most powerful communication tools of the 21st century. There has been a proliferation of SM tools in recent years, creating new opportunities to communicate, connect, create, and share information, without requiring exceptional coding skills to create or retrieve content [4].

Specifically, SM is increasingly becoming an augmenting tool in health care by enabling its users to acquire and share information; connect with others in the field; and communicate with colleagues, patients, or the public regarding health topics. Furthermore, SM supports patient empowerment by expanding the knowledge of the patients and placing them in a position where they can take control of their own health care needs [5]. This review is based on numerous studies and reviews that have investigated the different uses of SM in health care and its limitations and shortcomings. Consequently, this narrative is comprehensive and up to date, including the recent use of SM during the COVID-19 pandemic. The topic is relevant in today's scenario because the use of SM and social networking sites (SNSs) is increasing worldwide, especially in the health care industry. The findings presented in this review have strong implications for health care professionals, educators, and researchers.

Objectives

This review aims to shed light on SM use worldwide and discuss how SM has been an essential tool in the health care industry from the perspective of health care providers (HCPs). The review will be continued in Part II, where the use of SM from the perspective of patients will be discussed.

Methods

Search Strategy and Information Sources

Between March and April 2020, a comprehensive search on 4 databases (MEDLINE, PubMed, Google Scholar, and Web of Science) was conducted for all English-language medical studies that were published since 2007 and discussed SM use in any form for health care. A combination of the following keywords was used to search for titles and abstracts: "social media" (MeSH term) OR "social networking" OR "internet" (MeSH term) OR "WhatsApp" OR "Instagram" OR "Facebook" OR "YouTube" OR "Twitter" OR "LinkedIn" AND "healthcare" OR "health" (MeSH term) OR "medicine" (MeSH term) OR "physician" (MeSH term) OR "nursing" (subheading) OR "dentistry" (MeSH term) OR "telemedicine" (MeSH term), "recruitment" OR "education" (subheading) OR "career" OR "behavior" (MeSH term) OR "research" (MeSH term). Each of the 9 words in the first set was separately searched with each of the 12 words in the second set using "AND."

Screening Process

The articles were entered into an EndNote library, and duplicate publications were removed. Articles published before 2007 were excluded, as the words *social* and *media* at that time did not represent the current definition of SM. Titles and abstracts were assessed for eligibility. Studies that were not in English were excluded, along with those with inaccessible full text after unsuccessful attempts to access them. Irrelevant studies, such as studies that were not related to health care, studies whose primary outcome was not the use of SM in health care, or studies that discussed the negative impact of SM on health, were also excluded. Dissertations were also excluded from the study. The full texts of the studies were then appraised. Several relevant studies investigating SM use from patients' perspectives were found. Reviews on legal and ethical issues pertaining to the use of SM in health care were also obtained, following which, the publications were divided into 4 groups: *HCP*, *patient* or *the public*, *ethics* and *legal considerations*, and *shortcomings*. A decision was made to defer reviewing the last 3 groups and focus on this review on SM use by HCPs.

Categorization

After accessing the complete texts of the articles of interest, their reference lists were searched for additional studies, and the cited studies were also located. Thereafter, the articles were comprehensively reviewed. On the basis of the key findings, articles were initially grouped as follows: *sharing information*, *recruitment*, *education*, and *marketing*. As the review proceeded and more information was obtained, the groups were modified. *Sharing information* was divided into 2 groups: *health promotion*, focusing on HCPs sharing scientific information with the public, and *critical public health care issues*, which focuses on health announcements in crisis, especially COVID-19–related publications that warrant special attention. *Recruitment* was also divided into 2 groups: *recruitment*, which included job employment and residency program enrollment, and *scientific research*, in which studies discussed recruiting research participants and analyzing SM data. *Education* was renamed *professional medical education*, as this name specifies medical education. Studies related to continuous education were added to *marketing*, and the group was renamed *career development* and *practice promotion*. Another group was created—*professional networking* and *destressing*—which included findings from *sharing information* that discussed peer-to-peer communication and those from *education* that did not reflect professional education or career development. Finally, an additional group was created, *telemedicine*, as studies on this subject were abundant.

Results

Summary and Characteristics of Included Studies

The search yielded 5683 titles that were scanned with their abstracts. After exclusion of duplicates and noneligible studies, the initial sample comprised 73 publications. The full-text papers were retrieved. Additional studies from the article references or those emerged from the review but were not identified earlier were also added. This was because of variation in the keywords with respect to spellings (eg, behavior and behaviour), terminology (social networking and social network), and synonyms (eg, recruitment and employment) that were not accounted for in the initial search. A total of 142 articles (63 original studies) and 3 textbook chapters were reviewed (Figure 1).

The studies were conducted in the United States (61), Canada (12), Brazil (2), the United Kingdom (12), Europe (22), the Middle East (9), India (9), Asia (8), and Australia (7). The earliest study was published in 2008, and the latest studies were published in 2020, with most of them being published after 2014 (Figure 2).

Figure 1. Flowchart of the selection procedure.

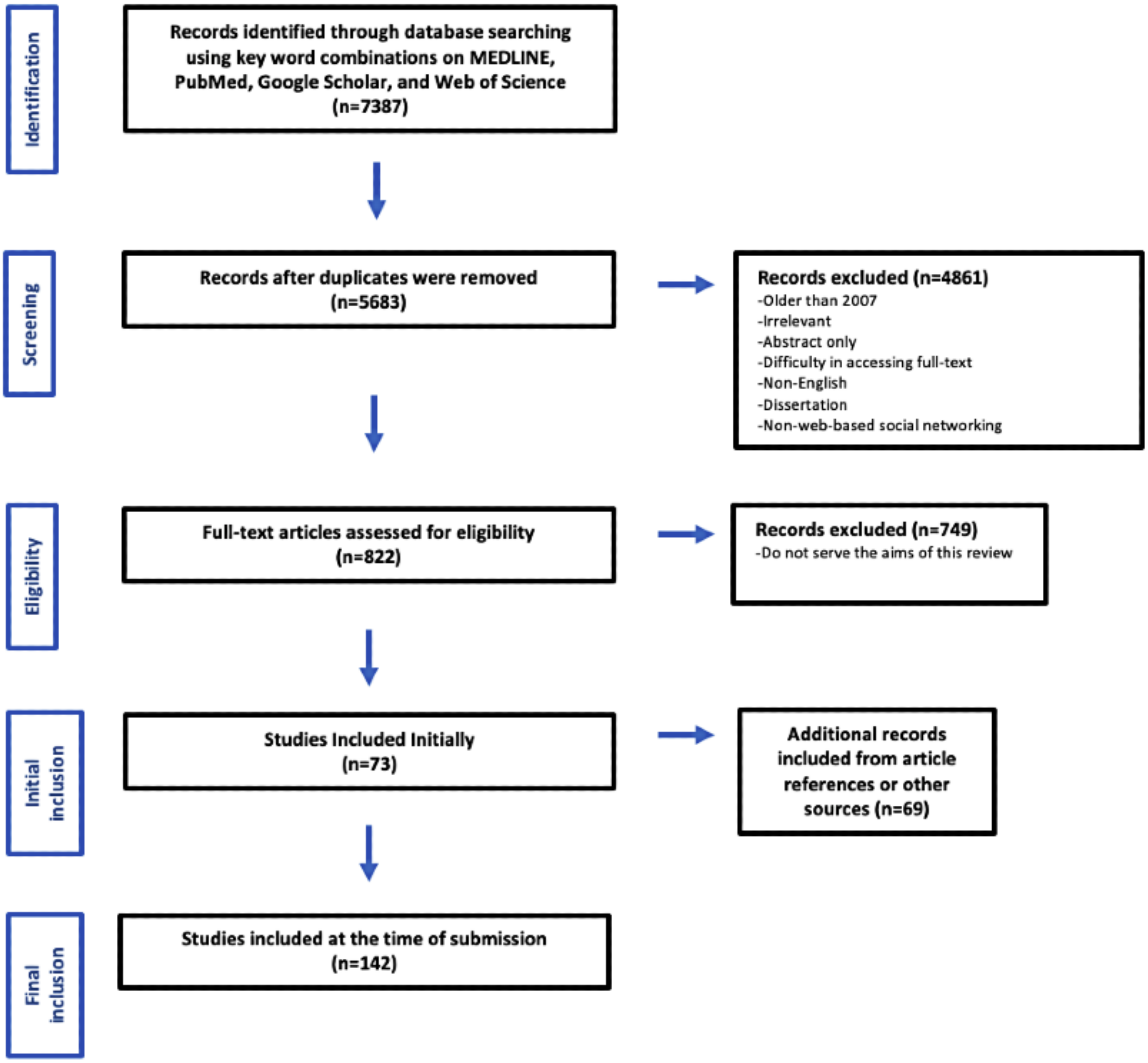
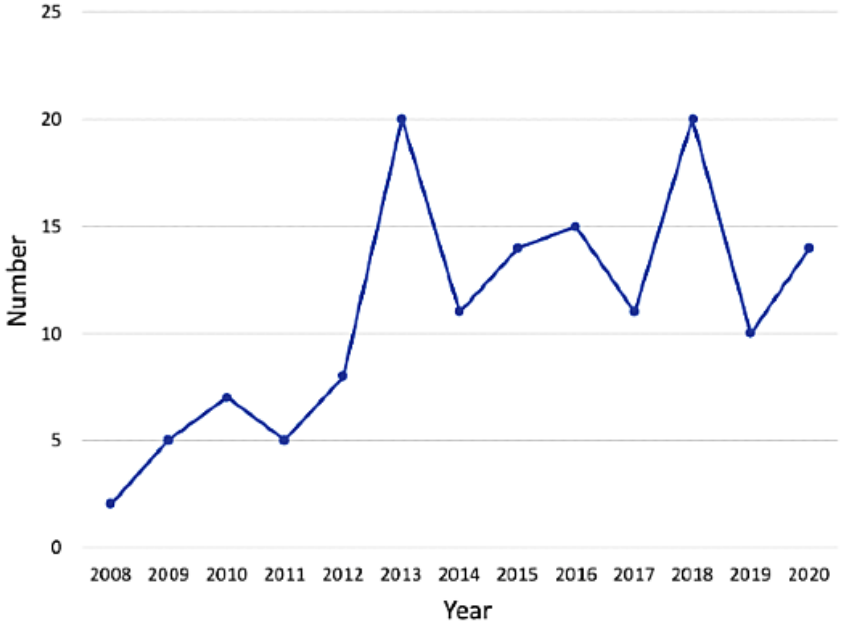


Figure 2. Number of included publications per year.



SM Platforms

Most reviews discussed SM in general and did not specify a particular platform; however, some original studies investigated specific platforms. The platforms investigated were WhatsApp/WeChat (15), Facebook (8), Twitter/Weibo (9), YouTube (4), Instagram (1), LinkedIn (1), Podcast (1), and Wikipedia (1).

Medical Specialties

Some reviews discussed SM use in a particular medical specialty, whereas others reviewed studies from diverse or unspecified specialties. Researchers from a variety of medical specialties investigated SM use in their original studies. These specialties were surgery (1), plastic surgery (4), neurosurgery (1), maxillofacial surgery (2), medicine (2), emergency medicine (2), psychiatry (3), orthopedics (3), otolaryngology (1), immunology (2), dermatology (1), radiology (1), urology (2), anesthesia (1), cardiology (1), pediatrics (1), oncology (3), nursing (5), dentistry (11), and pharmacy (1).

Discussion

Principal Findings

This literature review aimed to examine SM use in the modern world and how it has been recently incorporated into health care. Most of the reviewed articles were published in the past decade, suggesting that this review is both relevant and contemporary. It is evident from the published studies that SM has broad applications in modern health care. As discussed in the subsequent sections, HCPs (the term is used in this review as including physicians, dentists, nurses, medical and dental allied personnel, and health care organizations) not only use SM to provide care to their patients but also for personal development and destressing.

SM Use

SM use is one of the most common web-based activities, with an estimated 2.9 billion users worldwide as of 2019, a number that is projected to increase to 3.4 billion by 2023 [6]. With this, digital networking has witnessed a massive growth, and social communities have become boundless. Facebook, Twitter, Instagram, WhatsApp, and Google are relatively new platforms, but they are being used every day by millions of people worldwide. SM platforms are among the most commonly used sources for acquiring and disseminating information [7,8]. They are not only used for socialization, knowledge acquisition, and entertainment, but they have also been linked to significant political events led by young users [2].

Many SM tools have been introduced, and they continue to evolve. They may be categorized as tools for social networking (eg, Facebook and Instagram), professional networking (eg, Doximity and Sermo), media sharing (eg, YouTube and WhatsApp), content production (eg, Twitter), and blogs [9].

In terms of technological knowledge, SM users belong to 1 of the following 2 groups, as classified by Prensky [10]: digital natives and digital immigrants. Digital natives are those born after 1980, who are skilled in using technology, and who rely heavily on technology and social networking. Most digital natives were introduced to technology at an early age. Conversely, digital immigrants are those who acquired technological skills and adopted technology later in their careers [9]. SM use is generally high among digital natives, who explicitly prefer it over traditional media [11]. Some researchers believe that there is no dichotomous divide between internet users and nonusers. Although the terms are commonly used, Prensky's model and its usefulness have been challenged [12]. First, basic digital skills are not difficult to acquire, especially with repetitive use. With practice, a person born in the 1960s can become as digitally fluent as a millennial. Second, the distinction between both generations implies that digital immigrants can never completely acquire digital abilities and

that digital natives are automatically technologically skilled. This approach is neither scientific nor based on any empirical evidence. Third, the model overlooks the fact that age is not the only factor in determining digital skills. Socioeconomic and cultural factors of digital capability must not be ignored. For instance, a millennial who lacks access to technology is not a digital native.

Regarding SM demographics, its use is prevalent across all ages and professions [13]. However, different SM platforms differ in their demographics. The demographics of some of the most commonly used SM platforms worldwide can be further explored. Facebook has 2.7 billion monthly active users. According to a 2020 report, the highest number of Facebook users were aged between 18 to 29 years and 30 to 49 years, with more female than male users and more urban than suburban or rural users [13]. India had the largest number of users, followed by the United States, Indonesia, and Brazil. Regarding Instagram, there are 1 billion users globally. As of August 2020, there were more female than male users, and the United States had the highest number of users [13]. Users aged between 25 and 34 years represented the largest group of users [14]. Twitter had over 330 million users, who were predominantly male [13]. The top 3 countries for Twitter use were the United States, Japan, and India [13]. Approximately 30% of all users were aged between 25 and 34 years [15]. The Chinese Twitter-like SM platform is Weibo, and it had an estimated 480 million users [16]. WhatsApp is a mobile messaging app that is used by 2 billion users in 180 countries and in 60 different languages [17]. WhatsApp is more commonly used by younger people [17]. WhatsApp's direct Chinese competitor is WeChat, and it has about 1.17 billion users [18]. YouTube is commonly used worldwide, particularly in the United States. It is estimated that it has approximately 2 billion users. The users are more commonly male than female, and its use is prevalent in urban, suburban, and rural locations [13]. Finally, there were about 46 million students and recent college graduates on LinkedIn out of 675 million users [19]. Male users constituted 67% of the total users, and the United States had the highest number of users, followed by India, China, and Brazil.

Although most SM platforms share common features such as free registration, public and private communication, and fast content upload and retrieval, each platform is unique and has distinctive uses. It is common for users to have different accounts across multiple platforms, using each platform for different purposes. Facebook is an SNS that can be accessed from any internet-enabled device, such as personal computers and smartphones. Registration on Facebook is free, and users can create a profile that reveals selective information about themselves [20]. Users can post text, photos, and multimedia that become available to any user in their *friend list*. Users generally begin by adding family members and friends to their friend list, which can be expanded to include colleagues, acquaintances, and strangers with or without common interests. Apart from being able to share public comments and *likes*, a few years after Facebook was founded, a messaging feature was added that allows users to send private messages to individuals and groups. Users can use a variety of embedded apps; join and create *groups* and *pages*; play games; and receive updates regarding the activities of their friends, pages, and groups. Although the platform was initially limited to students in certain American universities, Facebook now has users worldwide. Instagram is a newer SM platform owned by Facebook and is designed primarily for free photo and video sharing [21]. By modifying their privacy preferences, Instagram users can opt to have either public accounts or limit their content to users that they accepted as *followers*. The platform also allows viewing, commenting on, and *liking* posts shared by users that they follow as well as private messaging between users. WhatsApp, which was acquired by Facebook in 2014, is a text and voice messaging app that has become incredibly popular owing to its features, flexibility, and compatibility with various phone and computer operating systems [22]. Although a free service, WhatsApp allows exchange of messages and calls on both desktop and mobile devices, in addition to media sharing and group features. WhatsApp's objective was to provide an alternative to SMS. Using WhatsApp, billions of users across the globe can simultaneously and instantly connect with others.

"Twitter is what's happening in the world and what people are talking about right now"—this is how Twitter describes itself [23]. It is a microblogging platform that allows users to post and access short text, image, or video posts called tweets. Although tweets were originally limited to 140 characters, the limit was increased to 280 characters in 2018, along with permitting the sharing of website links and multimedia. Twitter's mission is to provide users with the ability to create and share ideas and information instantly and without barriers. Users follow other personal, official, or organizational accounts. They can either create their own tweets or *retweet* those by others to their followers. YouTube is a web-based video-sharing platform that allows users to upload, view, share, rate, report, comment on videos, and subscribe to other users [24]. Its mission is to provide users with a voice through video sharing, stemming from the belief that the world would be a better place when people listen, share, and build a community through their stories. The name of this platform is straightforward: *You* represents content that is user-generated and not created by the site itself, and *Tube* is an older term for television. Most YouTube content is uploaded by individuals, but some media corporations have established partnerships with YouTube to offer some of their materials on this platform. LinkedIn, acquired by Microsoft in 2016, is a business and employment-oriented SM service that operates as both a website and mobile app. LinkedIn is mainly used for professional networking, allowing employers to post about job openings and seekers to share their curricula vitae [25]. Using the platform, users can build strategic professional relationships rather than expand their friend circle. LinkedIn's vision is to provide professional opportunities to its users, and its mission is to connect professionals worldwide. It originated in the living room of one of its cofounders a year before its launch in 2003. LinkedIn today has a diversified business model that has generated successful recruitments.

SM users have claimed that they have more digital friends and connections than real-world ones, which highlights the transformation of the ways in which people connect with each other and the importance of web-based relationships in today's world [26]. Over time, social networking platforms have targeted

different age cohorts, making SM use widespread among the general population. For example, as of April 2020, men aged between 25 and 34 years constituted the largest demographic group of Facebook users, and those aged 65 years and older were the fastest-growing group [27,28]. Similarly, the largest group of Twitter users were people aged 25 to 34 years, whereas 15% of users were older than 50 years [15].

The public attitude toward SM use has drastically changed over the years as it became more accessible and diverse in its offerings. Consequently, SM has become a universal communication channel, and responses in reality and on the web have increasingly become intertwined and concurrent [29,30]. Furthermore, SM offers lucrative opportunities to disseminate information and thoughts directly to the public, share experiences, build communities, and connect people with common interests, something unthinkable 20 years ago [31].

SM Use in Health Care

The more digitally oriented the world becomes, the more the health care industry visualizes SM as an important channel for health care promotion, employment, recruitment of new clients or patients, marketing for HCPs, and building a captivating brand name. HCPs have realized that SM is not just a platform to post vacation photos and interact with followers. Perhaps the 4 most common areas where SM plays a major role in the health care industry are health promotion, research, marketing and branding for individuals and practices, and recruitment. It has been evident that web-based content can be spread to offline environments, such as classrooms and meeting rooms. Furthermore, SM has undeniably changed patient-practitioner relationships because of patients' better understanding of health information and their more active role in health maintenance [32].

The effect of behavioral and social factors on health outcomes has evolved significantly in recent decades [33]. HCPs continually search for new and more efficient methods to reach larger populations, especially those who were inaccessible via traditional methods. It is incumbent upon them to use every available tool to reach their intended audience. Thus, HCPs and health organizations should capitalize on the opportunities provided by SM and update strategies to reach communities and age cohorts at a relatively low cost [7,34]. In other words, SM brings a new dimension to health care and is changing the nature and speed of health-related interactions between individuals and health organizations. For example, communicating through photos and videos along with text is part of the mobile revolution, and messaging apps are now regarded as a viable medium for sharing knowledge and discussing clinical cases [35]. In summary, using SM could be a key strategy in addressing some of the challenges and limitations often faced by HCPs in traditional health communication through faster and cheaper dissemination, more accessibility, better interaction, and increased patient empowerment [7]. Moreover, information can now be easily brought to audiences with special needs or low literacy [36].

In the fast-paced modern world, time constraints are common in medical practice, and when combined with the demands of individuals with chronic conditions or unplanned emergency situations, it is challenging for HCPs to dedicate extra time to patients. SM can provide efficient and easy-to-use platforms that encourage patient-practitioner interaction and facilitate necessary actions from both ends [37]. In fact, as of April 2018, there were more than 200,000 health apps, including social networks for people living with a specific medical condition; approximately 19% of smartphone users have at least one health app on their phones [38]. Furthermore, in a 2013 survey of more than 4000 physicians, 65% stated that they used SM for professional reasons [39]. SM use has not been limited to HCPs. Patients have also incorporated SM in their daily lives, which encourages HCPs to explore different ways of making their mark in this growing market [31].

Technology is evolving very rapidly [40]. Competition produces better services, and the diversity of options enables users to choose a tool that best matches their individual needs. Although different platforms often have different target demographics, audience overlapping may occur and should be considered by health organizations when devising their SM health promotion strategies. In health care, SM tools can be used for different purposes (health promotion; dissemination of health information; education; professional development; recruitment; communication with the public, colleagues, and patients; and research) and in diverse medical specialties (cardiology, nursing, radiology, dentistry, surgery, pathology, pediatrics, pharmacy, emergency, and critical and palliative care) [41]. As of August 4, 2020, 27,546 results appeared when searching for *social media* on PubMed, demonstrating the growing interest in SM within the health care industry.

Health care systems, especially in times of crisis and outbreaks, require the dissemination of information to practitioners, patients, and the general public rapidly and effectively [34]. Health organizations and officials, by taking upon a more active SM presence, gain access to vast global networks capable of quickly spreading information and promptly mobilizing large numbers of people toward public health goals [2,42]. Never before has the entire world united as it has in recent months in its fight against the disease caused by SARS-CoV-2, more commonly known as COVID-19. Searching *COVID-19* on PubMed on August 4, 2020, yielded 37,576 results, which exposes the abundance of information and data that has become available in 7 months since the beginning on the pandemic [8]. However, the World Health Organization (WHO) has expressed concerns about fighting 2 battles at once: the pandemic and the *infodemic*—the latter refers to a rapid and far-reaching spread of both accurate and inaccurate information about the disease [43].

It is important to mention that the popularity of SM is directly linked to its many advantages. Advantages of using SM in health care include its expressive nature, accessibility from a smartphone, prompt content sharing and response generation, improved and two-sided communication, reduction of consultation time, smoothing of hierarchy, more efficient teamwork, ability to forge connections between people, and ability to reach large masses [44,45]. Furthermore, SM facilitates the access to health information for extended population groups, regardless of geographic location, age, or education, compared with traditional communication methods [41]. However, the

most influential advantage of SM remains its cost-benefit feature: it can reach an increasing number of people without the high cost of traditional means and the information remains available 24 hours a day, 7 days a week.

Similar to most technologies, SM has its disadvantages. In the health care industry, these include increased workload, risk of unprofessional behavior, disparity in the sense of urgency, a demanding sense of needing to stay connected 24 hours a day, difficulty in obtaining discussion records, worries about leading to or identifying patients, privacy breech, change of patient-practitioner relationship from a professional to a personal one, and the risk of reducing the autonomy of junior doctors [45]. Those who choose to use SM should be aware of the potential risks and problems that they could encounter but should not shy away from using SM because it can greatly increase the reach and impact of HCPs' work and improve patients' health [34]. In the following section, the specific uses of SM in health care are discussed in more depth.

SM Use From the Perspective of HCPs

The literature review yielded an abundance of information. The studies were categorized as follows to best present the findings: health promotion, career development or practice promotion, recruitment, professional networking and destressing, professional medical education, telemedicine, scientific research, and critical public health care issues.

Health Promotion

Producing and disseminating information has played a pivotal role in the history of humanity. Over the years, an increasing number of public health organizations, medical institutes, and HCPs are using SM tools to disseminate visually rich public health messages to the general public. The primary goal is to share solid, evidence-based, and up-to-date health information that educates and affects millions of SM users and to dispel common misconceptions and counterbalance inaccurate material rapidly spreading through SM [2,32]. Examples of how SM can increase the accessibility of a massive number of recipients to health care information around the world include concise educational tweets on Twitter, a pediatric dentistry group on Facebook where fellow specialists discuss anonymized cases and share ideas, photos of a salvation mission to an underprivileged community on Instagram, and the results of a clinical study broadcasted via WhatsApp. These are all examples of how SM can not only increase accessibility, but it can do so at a faster rate than any other means, and perhaps in the cheapest way possible.

Access to oral health care services is limited by a lack of universal coverage. SM, which is a method of mass communication, offers an alternative to traditional communication, which extends to reach underprivileged and underserved communities. The WHO and the United States Centers for Disease Control and Prevention (CDC) are among many other public health institutions that use SM to communicate with the public during public health crises and natural disasters [8,29,35]. Physicians also use SM to promote patient health care education on a smaller scale within their networks. Research has shed light on the many tools that have been used for this purpose. For example, HCPs can tweet, record videos, and participate in health-related discussion forums, which provides an opportunity for physicians to share scientific information and broaden their knowledge [46,47]. Furthermore, information from international conferences and findings from the latest research and clinical trials can be presented in mainstream media to be shared with millions of people [48].

Sharing such information not only helps improve knowledge but can also improve attitudes and practices related to health. For example, in dentistry, SM has played a role in helping patients cope with challenges such as dental anxiety and in presenting dental management options in a convenient and nonthreatening manner [42,49]. Evidence now shows that SM-based interventions are linked to healthy practices such as tobacco cessation, increased physical activity, and diversion from risky sexual behaviors [39,50].

In conclusion, there is evidence that SM helps to improve access to health information. When designing SM campaigns and interventions to disseminate health information, it is important to develop messages that may be more likely to resonate with and elicit reactions from individuals [2]. Messages tailored to certain population segments are more effective than generic messages, as tailored messages address the specific needs of their recipients [51]. Furthermore, interactive (two-way) communication is more effective than linear (one way) communication [7]. Importantly, SM must complement rather than replace traditional health promotion. More research is needed to investigate strategies that can increase access to health information for minorities and marginalized communities and for populations deprived of internet access.

Career Development or Practice Promotion

One of the measures of the success of HCPs is their ability to attract and retain patients. This will not only maximize income but will also boost reputation. SM has played an important role in enhancing practice or practitioner ranking on search engines, even more than academic pedigree and experience [52]. As search engines generally direct patients' traffic, a strong presence on the web can be crucial to attract patients to a practice.

HCPs at all stages of their careers can use SM to brand their name. SM aids in developing their name, expanding their network, and learning about career-enhancing opportunities [53]. It can also be used as a marketing strategy to attract patients of various demographics and has been proven to be effective in engaging and obtaining new patients [54]. In a survey conducted in 2013, 12.5% of health care organizations reported attracting new patients through SM [39]. Moreover, a 2012 study of dental practices in the United States revealed that 51% of the practices used SM, of which 91% used it for marketing purposes and 73% used it to increase their presence on the web [55].

With the extended use of SM among patients and HCPs, practitioners must now compete for patients' attention and need to be strategic regarding the content they share and platforms they use [7]. HCPs should advertise their professional trajectories, areas of experience, and treatment outcomes by

focusing on information tailored to the target audience in an educational manner that does not typify commodification or unfair competition. The eagerness to achieve popularity and to attract new customers or compete with colleagues results in some HCPs thinking only with a short-term approach and prioritizing greater financial gains. For example, some orthodontists and plastic surgeons post before-and-after photographs with drastic improvements without explaining that biological variations among patients, differences in response to treatment, and other external factors may affect the course and outcome of the intervention. Without such information, patients tend to have unrealistic expectations and end up being disappointed [56]. Unfortunately, some practitioners tend to digitally modify images to accentuate treatment-led improvements. This misuse of technology could lead to serious reputational damage for the practitioner and the profession in general in addition to unfavorable court decisions [56].

HCPs build their status using SM in diverse ways. They begin by creating a profile page on one or multiple platforms, which allows them to create and upload content. By connecting with colleagues, they can begin to establish a digital social network. Moreover, groups based on common interests further expand their social network and raise the practitioner's name in bigger circles. These processes can create a haven for viral marketing, which can be leveraged to create a name, develop a digital voice, and disseminate health information in a timely and cost-efficient way [7,30,42]. Moreover, for newly qualified practitioners, contributing to discussions on forums and virtual meetings raises their profile among more experienced practitioners who may be geographically distant. This can leverage word-of-mouth referrals and attract fellow researchers to collaborate.

In terms of cost-effectiveness, a 2017 study on the SM return on investment (ROI) showed an upward trend that represented stable growth for Facebook, whereas Instagram demonstrated substantial ROI. It was concluded that SM resources were superior to standard internet-based resources. When all SM platforms were combined into one graph, there was a consistent finding of growth associated with all SM sources over time [57]. As the following quote summarizes, "physicians have to realize that our patients are doing it (SM), so this is where we need to be" [58].

In summary, there is no one-size-fits-all SM platform, and there is no single way to share content that is superior to all others. It is essential for an HCP to emphasize the importance of their specialty; present the strengths in their particular practice; understand the features and user demographics for SM platforms; and, most importantly, know their target audience. For example, a plastic surgeon or orthodontist may find it beneficial to share pre- and posttreatment or procedure photos of anonymized patients, for which Instagram may be the ideal platform. A practice that is community friendly may invest in a Facebook page to keep the audience updated on offers and services. A family medicine office may share announcements regarding the arrival of a flu-vaccine and post photos of staff members vaccinating themselves to motivate people. Twitter may be useful for posting specialty-related educational messages or sharing information on health-related matters to make HCPs more visible. A pediatric dentist may use YouTube to share videos of tricks used in the office to make the experience less threatening for children. More training courses and talks on how to leverage SM to establish a presence and build a name may be beneficial for HCPs who are not SM savvy.

Recruitment

SM is making great strides not only in the modern world of technology but also in the workplace: it is transforming the way people find and engage in work. It seems that the conventional channels for recruiting employees are not as effective as they once were. Instead, we are shifting toward SM not only as a platform for social interaction, photograph exhibition, and creative expression but also as a space for far-reaching, low-cost job searches. Regarding employment, the interest in SM is bidirectional. On the one hand, employers are often keen to know more about a candidate applying for a position than what is stated in their resume´. On the other hand, employees, especially millennials, will first want to know more about the dynamics of a firm and the personalities of their future boss and coworkers before they commit to the job. Recruitment in the medical field requires more than an application. In this section, the recruitment of HCPs for employment or students for residency programs is discussed.

Human capital is of major importance to any organization because humans produce income and are a source of competitive advantage [59]. Recruitment of qualified employees who are fit for the job is not a simple, one-way decision as it used to be. Performing due diligence in hiring a new employee is more essential than ever and is a multidimensional process, including at least one interview, drug screenings, and background checks. More recently, employers have turned to nontraditional methods and to SM to further analyze potential candidates [60]. SM prescreening may have the potential to offer information about the applicant above and beyond what is stated in the resume´ and can be assessed in a more traditional screening [61]. For job seekers, because of the large number of SM users and the relatively low cost of setting up, SM platforms are ideal for finding employment. Furthermore, many organizations are now investing in SM to display their employer brand and, in return, attract qualified applicants [59,62]. Organizations aiming to attract applicants, especially in fields where competition and demand are high, such as in nursing, must make every effort to promote a unique brand image and attract potential candidates [59,63].

There is evidence that recruiters can accurately determine productivity-related traits solely on the basis of personal information about a candidate available on SNSs [64]. Baert [65] found that personal photographs have become more effective as objects of communication than of memory. This research described interesting theoretical mechanisms that underlie better labor market outcomes for more attractive people. For example, it proposed that self-confidence from good looks could drive productivity, leading to emotional stability, and, consequently, labor market success [66]. The study also found a higher impact of face pictures seen on Facebook's profile photo compared with those attached to a resume´ [65].

From a job seeker's point of view, SM makes it possible to apply to hundreds of jobs, even globally, at once. Through SM,

job applicants can increase their presence on the web to grab the attention of employers [60]. In the health care industry, it is advisable for job seekers to be active in various medical societies to expand their connections and to make a positive impression on future employers. Similar to the real world, it is advisable for applicants to attend virtual conferences and discussion boards and to introduce themselves to others at every reasonable opportunity [67]. It is important to note that employers do not use only professional platforms such as LinkedIn for hiring. In fact, they check many SM platforms when screening for prospective employees [60]. It is not a bad idea that employers and applicants conduct periodic searches for their own names to ensure that their SM persona projects a professional image [2].

Several studies have investigated the effects of SM on recruitment in health care. It was found that a hospital's profile on SM can shape employer brand perceptions and attract nurses. In addition, nurses who visited the hospital's Facebook page were more attracted to work there [59]. In another study, over 92% of employers stated that they were planning to use SM for recruiting [68]. Moreover, a study conducted in 2012 found that the recruitment and screening costs were reduced by 50% by using SM and that 65% of employers were evaluating the integrity and character of potential employees based on their SM profiles [60]. As shown in a review by Davison et al [61], a study found that 20% of the organizations surveyed were planning to use SM for applicant screening. Furthermore, LinkedIn was the most commonly used SM platform for screening applicants, whereas the use of Twitter for screening purposes increased from 11% to 31% between 2011 and 2013 [61]. Interestingly, a study found that there were 38% more job interview invitations to candidates with the most beneficial Facebook pictures [65]. In addition, a Microsoft survey revealed that 79% of employers searched for web-based information regarding prospective employees, but only 7% of the candidates were aware of this possibility [69,70].

Regarding residency programs, SM is a mechanism to accentuate the programs' visibility on the web and to screen residency applicants [71]. It is important to note that it is not only credentials and high scores that secure a spot in a specialty program; personality traits and characters are becoming more significant than ever. Program directors (PDs) now want to know applicants on a personal level. As much of this information would be illegal to obtain in a traditional interview, they may search for it on SM [72]. Admission officers and PDs are now capitalizing on the abundance of information and the popularity of SM [72]. They may encounter content that seems unprofessional or exposes negative character traits that are useful in making decisions about applicants. Many residency programs now search Facebook and other personal SM platforms to screen applicants [73]. Even residents are now using SM platforms to obtain information on possible postgraduate opportunities [71].

There is an abundance of research on the use of SM for applicant selection in residency programs. In one study, 17% of PDs screened applicants on SM, 33% gave lower rankings to applicants based on SM findings, and 69% stated that they will continue to use SM for applicant screening [74]. In another survey, most school children who were interested in studying medicine felt that behaviors on SM should be considered for admission to medical schools [75]. Furthermore, a study conducted in 2016 found that 18% of PDs visited the SM profiles of residency applicants, 10% gave a lower rank or completely disqualified an applicant because of negative web-based behavior, and 10% took formal disciplinary action against a resident because of negative web-based behavior, with Facebook being the platform used by most PDs [76,77]. Another survey found that 97% and 90% of PDs agreed that candidates should be held accountable for illegal acts and unprofessional behavior on the web, respectively, whereas 89% of them agreed that information voluntarily published on the web is fair to use in judging character and professionalism. Furthermore, 82.4% of PDs indicated that they would favor the candidate with a sterile Facebook profile if they were choosing between 2 mock candidates [78]. Moreover, student pharmacists demonstrated a general attitude that web-based personas on SM should not be used to judge professional attitudes and abilities [79]. Although most medical school PDs believed that screening applicants on SM does not constitute a violation of the applicants' privacy, the topic remains controversial and views regarding the appropriateness of using SM profiles to judge character and professionalism vary [78,80]. There is general agreement that SM information is open for judgment by others, especially among older PDs [78].

Professionalism is advocated by the American College of Surgeons as a quality that extends beyond the clinic, operating room, and hospital and into the community in the real world and on the web [81]. There are some issues associated with using SM to judge a possible employee's or resident's professionalism and character. First, screening is usually done by a single person without a standardized scoring rubric. Second, content is unstandardized among the different SM platforms, and the information displayed differs across platforms; for instance, it would not be fair to compare someone's Facebook photo album of a Spring Break trip with someone's contribution to a medical discussion on Twitter. Moreover, screening SM content showed poor test-retest reliability, especially as the content could change rapidly. Interrater reliability is potentially affected by the content being rated and the characteristics of the rater. Construct validity also seems to be weak as no specific construct is usually in mind; instead, a rater casually scans profiles to make a judgment on an applicant or screen potential new hires. Finally, there is a problem with generalizability across platforms. It is suggested that personality traits should be judged from platforms with flexible formats (eg, Facebook), whereas professional traits and experiences should be judged from more structured platforms (eg, LinkedIn) [61].

In summary, although e-professionalism is a new topic, it is receiving considerable attention from recruiters and is being taught as a part of medical curricula [76,79,82,83]. It is necessary in this age to educate job or residency candidates about their digital voice and persona management [65]. Job or residency candidates should consider their publicly available web-based information as an extension of their resume´ and should be aware that many employees use SM to investigate applicants. Therefore, candidates should ensure that their public SM profiles include nothing unprofessional about themselves

[65,67]. Finally, the establishment of clear and equitable guidelines for searching candidates on SM is essential to prevent potential bias.

Professional Networking and Destressing

Professional connections represent important channels through which HCPs exchange knowledge, share expertise, refer patients, seek a second opinion, collaborate on research, hire and employ, provide social support, and improve health care outcomes. In the last few years, work-related communication has changed considerably with the advent of electronic communication tools, especially with the aid of instant messaging on smartphones [30,84]. Virtual professional communities can enable members to quickly access evidence-based information and disseminate work, which can lead to increased immediate impact [85,86]. Most SM platforms are found to be easy to implement, effective, quick, and low cost [87]. In a recent systematic review, positive predictors for using SM among HCPs were identified to be younger age, lower rank, and fewer years of experience, and the most commonly used platforms for communicating with colleagues were Facebook, Twitter, LinkedIn, and WhatsApp [88].

Owing to the exceedingly large number of HCPs on SM, platforms that are designed only for medical personnel have been introduced. Digital communication and social interactions occur between people who may or may not be known to each other [44]. In addition to medical issues, discussions usually address diverse subjects such as politics, practice management, career enhancement opportunities, and even dating in a medical environment [2,89]. Sermo, the world's largest virtual doctors' lounge, is a leading social network for physicians that is now available in 30 countries [90]. Doximity is a newer physician-only social networking platform with more than 500,000 members as of 2020 [91]. In addition, there is the Medical Directors Forum, which is an SNS exclusively for medical directors that provides a secure environment for peer-to-peer interaction [92]. Studies on HCPs' preferred SM platform showed that Facebook was used most frequently (86%). Other commonly used platforms were Medscape Physician Connect (52%), Sermo (44%), LinkedIn (42%), YouTube (40%), Blogging (25%), and Twitter (20%) [26]. These statistics have been confirmed in subsequent studies [32]. For health-related reasons, physicians primarily used LinkedIn (70.7%) and Twitter (51.2%) [32]. Another study reported that HCPs spent an average of 11 to 13 hours per week on medical professional networking sites [26].

WhatsApp has been used as an intradepartmental, patient-related communication method because of its instant and more efficient handovers [93]. When physical proximity was a barrier, physicians preferred to use WhatsApp to exchange work-related knowledge over traditional text messages [84]. The American Academy of Pediatric Dentistry has a private group on Facebook with fewer than 3800 members who share clinical cases, clinical experiences, research results, new products, and relevant events [9]. Facebook has many other groups for dentists that are open to the public to view and join [30]. Microblogs such as Twitter allow a dynamic and concise exchange of information that is instantly accessible by an increasingly large number of readers [89]. Furthermore, the dissemination of scientific literature on SM (eg, Twitter) has increased the number of citations and downloads of published articles [94-96]. LinkedIn serves as a professional space for HCPs to demonstrate their expertise and capabilities; 54% of physicians have used it to communicate with colleagues [32,42].

SM also has a positive impact on students. The sense of belonging is crucial for undergraduate training. By being part of a well-respected learning environment, students benefit educationally and socially [97]. Moreover, having guidance and support on a 24/7 basis can ease their transition from university to more independent training centers [30].

In recent months, SM platforms have become helpful in maintaining communication with friends and family and reducing isolation and sense of loneliness, which could have a negative psychological impact [98]. Amid the COVID-19 pandemic, many practices have been affected and many jobs have been lost worldwide [99]. The sense of unity and comradery introduced by SM among users has helped countless individuals overcome hardships, including HCPs. In the first half of 2020, HCPs were deployed into unfamiliar environments because of the COVID-19 pandemic, worked beyond their areas of expertise and over long hours, and had to involuntarily isolate themselves from their families. This crisis has been anxiety inducing and stressful for HCPs, who often resorted to SM to voice their frustrations, experiences, and opinions not only with family and friends but also with the global network of frontline staff enduring similar challenges. The unique virtual siblinghood united the global health care community like never before. A Facebook page was created to facilitate the renting of recreational vehicles for HCPs to self-isolate outside their homes [100]. The public played an important role in paying due respect to HCPs, who were often referred to as heroes, both literally and figuratively. Illustrations portraying their pivotal role were shared on every known SM platform. Videos showing countless people applauding for HCPs at certain hours of the day went viral. Many HCPs engaged in what was labeled as *COVID-19–free zones* to escape, even if momentarily, from the pandemic stress. Clinicians from all specialties in diverse locations joined forces against a single enemy. Their voices echoed louder when they addressed lawmakers demanding improved access to personal protective equipment (PPE), increased testing for COVID-19, reduced reimbursement barriers to telemedicine, and improved mental health care.

To conclude, SM plays an important role in the lives of HCPs at a personal level. Whether SM is used for amusement, *zoning out*, or commiserating, it provides a safe haven for HCPs to put off their metaphoric heroes' capes and find comfort in their humanhood again. Future research should investigate the role of SM in helping HCPs individually and collectively tackle the challenges resulting from the COVID-19 pandemic.

Professional Medical Education

Millennial students of health professions are increasing in numbers each year. They possess qualities consistent with being lifelong learners [71]. As Prensky [10] discussed, traditional education systems are no longer suitable for contemporary students. Millennials and younger generations process

information in a fundamentally different manner from their predecessors. SM offers formal and informal educational opportunities and has the ability to remove physical barriers that could otherwise impede access to educational resources [51]. Not only is SM rich in educational resources but coupling the information with the interactive exchange of ideas and the live discussions has also made it a valuable educational tool. When SM was integrated into teaching, students were motivated by content obtained from SM, and positive behavioral changes were promoted [101]. Moreover, when SM was incorporated into clinical education, students perceived better collaboration with their peers, improved professional and career development, and larger supportive learning communities [102].

Social networks are an underutilized educational resource, not only for trainees but also for experienced clinicians. A large array of educational material is abundant on SM from seniors with advanced experience and from fellow trainees as well, usually at no cost to the user [30]. The differences between experts and novices are slowly diminishing because of novel forms of peer learning and knowledge production facilitated by SM [35]. Most platforms are frequently used to engage learners. YouTube in particular is more commonly used to teach technical skills and has been acknowledged by dentists as a convenient educational platform [30,71,103,104]. YouTube can also be used in classrooms to forge discussion, illustrate a procedure, or reinforce information, which promotes critical thinking and problem-solving skills [70]. A study conducted in Saudi Arabia found that YouTube was the most commonly used SM platform in medical education [105].

Evidence suggests that SM has a place in health care education. Universities use SM to create virtual classrooms and increase access to academic libraries [70]. In the United States, 95% of medical schools have some Facebook presence, and 71% of them have student groups [106]. In a study on nursing students, Twitter was used to view videos of clinical scenarios, and students tweeted their observations for instructor feedback [70]. Internet and SM content have been successfully used to train older caregivers to improve the caregivers' and patients' quality of life [5]. In a unique experience, the University of Rhode Island managed to connect students to geriatric patients on Facebook. It promoted students' empathy and communication skills while helping patients advance their SM skills to battle loneliness [2]. In addition, students in an oral and maxillofacial radiology course perceived using Twitter as a helpful learning tool that enhanced access to faculty [107]. Moreover, medical students and professionals in cardiology reported the use of social networks for education and professional training [103].

Learning may be considered a social activity [106]. The more senses the students stimulate in their learning process, the more likely the information acquired is to be retained. Thus, the greater the engagement and contribution of the students, the better the learning outcomes. SM provides a medium for active collaboration rather than passive learning. In nursing, 92.4% of students perceived a positive learning impact from the podcasting of lecture materials [108]. In another study, students who were more heavily engaged in blog-based discussion of relevant learning material had higher grades than peers who had contributed less to the discussion [109]. Passing an examination was significantly associated with combining discussion on a WhatsApp group with the web-based question bank, and so were their higher grades [110]. Medical students who used Wikipedia had superior short-term knowledge acquisition compared with those who used a digital textbook, which suggested a potential role for Wikipedia in medical education [111]. In addition, medical students who integrated the use of SM in 2 elective courses were satisfied with the new approach [112]. Twitter and Instagram have also been described as helpful tools in radiology education [113]. Participants in a study reported that SM was perceived as helpful and very helpful for improving knowledge, creativity, decision making, critical skills, and problem-solving abilities [114].

In a study by Alsuraihi et al [105], YouTube, Facebook, and Twitter were among the most commonly used resources for learning. Although 95.8% of the students believed that SM was beneficial for learning, 40% thought it might be distracting [105]. In a review by Chan et al [85], it was found that multiple residencies used SM to broaden the horizons of trainees and facilitate engagement in journal clubs using virtual classrooms; a wide range of SM platforms were featured, including Facebook, blogs, Wikipedia, and podcasts. Specifically, dermatologists in a study agreed that WhatsApp discussions enriched their scientific knowledge of clinical cases and promoted learning about relevant references and upcoming meetings [115]. Participants of a 2015 study on surgical teams expressed that WhatsApp helped to flatten the hierarchy among students, residents, and experienced consultants, enabling them all to actively contribute to discussions without inhibition. This comfortable environment is especially helpful for shy and marginalized students [116].

Once students move beyond structured, supervised learning environments, they must recognize their own gaps in knowledge and skills over time and make every effort to fill them, adopting skills for lifelong learning [106]. The increasing mutability of knowledge in the digital age and its exchangeability and accessibility on mobile phones make learning thorough SM platforms a common practice for many medical students [35,117]. However, learning cannot be done through SM alone but is used to augment learning from textbooks, peer-reviewed research publications, and mentors, and just like with other sources of information, critical appraisal to information retrieved from SM must be applied; this is what lays the foundation for a future competent web-based learner [85,106,118]. It is important that students understand that educational material shared through SM cannot be accepted as is without a great degree of skepticism and objective evaluation.

To summarize, today's students are unique in how they learn and acquire skills. Current educational systems must adapt to the needs and qualities of modern students and augment, if not replace, the traditional teaching methods with more digital means. It is essential for educators to put every effort in determining the best means of presenting information to their students and guiding them in their information search and appraisal. Retrospective research can be planned to compare the performances, learning outcomes, and teaching strategies between 2 student cohorts: those that were taught in a traditional manner and those that relied on SM.

Telemedicine

As people are becoming increasingly fluent in using novel technologies, health care has recently changed when, where, and how patients and HCPs communicate [119]. Telemedicine is the use of communication technologies and electronic information to provide health care support to patients or health care workers who are physically distant from HCPs [120]. Many branches of medicine are now adopting electronically mediated care; terms such as teledentistry and telepsychiatry are not uncommon, and publications related to telemedicine have been increasing [121-123]. Among the specialties that use telemedicine are pediatrics, psychiatry, diabetes, dentistry, nursing, palliative care, and allergies [124-130]. HCPs can now overcome their limited clinical time by communicating with their patients remotely. With the aid of novice technology, they provide a more convenient type of care for patients, especially for following up patients with chronic health care needs [131].

As young and highly qualified HCPs prefer staying in urban communities, telemedicine significantly augments clinical care, especially in underprivileged and underserved communities in rural areas [2,82,132,133]. Furthermore, as health care costs continue to increase, organizations are aiming to reduce costs without jeopardizing the quality of care being provided [134]. Recruitment and workforce sustainability are often an issue, and some countries with large, sparsely populated rural areas have grappled with how to overcome medical and dental provider shortages in these rural areas. Telemedicine and teledentistry can be of great use to ensure that new practitioners appointed in rural locations are not secluded and have the advice and support they need to promote their clinical work and psychological well-being [135]. It may also be used to connect HCPs in third-world countries with specialists in more medically advanced regions; for example, surgical procedures may be streamed live, and questions can be asked in real time [69].

Smartphones are fast, portable, and simple to use; mobile apps now seem to be ideal for quick learning or communication between colleagues or HCPs and the public or patients. Mobile apps are among the most commonly used tools for telemedicine [82]. Globally, but particularly in low- and middle-income countries, communication among HCPs is facilitated via WhatsApp, providing faster diagnosis and immediate management of acute findings [136-138]. A systematic review on telehealth concluded that 74% of the studies reported economic benefits of eHealth interventions for different medical conditions [134].

Programs for electronically monitoring intensive care units allowed HCPs to remotely monitor the conditions of up to 100 patients in multiple hospitals [139]. Pandemics and natural disasters pose challenges to effective and prompt health care delivery. Although telemedicine and eHealth might not solve them all, they can aid HCPs in providing the necessary management in scenarios in which the infrastructure is intact. In recent weeks, the distant triage that allowed patients to be efficiently screened for COVID-19 was patient centered and in compliance with self-quarantine; thus, it protected patients, clinicians, and the community from exposure to the virus [139]. For instance, replacing scheduled office visits with telemedicine visits in case HCPs were quarantined, absent, or sick was a productive initiative at Jefferson Health, Philadelphia, Pennsylvania [139]. An interesting model of telemedicine was explained by Baker and Stanley [40], in which patients use an app to navigate to a specific medical site, answer a few triage questions about their medical condition, wait in a virtual queue to be connected by video to an HCP, and discuss their condition or concern.

Sending clinical photographs privately between colleagues for a second opinion or to enrich discussion is not uncommon. A comprehensive review by Boulos et al [35] shares findings from multiple studies on the use of WhatsApp and Instagram in those contexts. One study found excellent inter- and intraobserver agreement in the assessment images of tibial fractures using WhatsApp [140]. There was a report of a life-saving use of WhatsApp in a resource-limited situation in which the life of a critically ill patient was saved by sending clinical monitor images with electrocardiogram changes and receiving feedback from an expert consultant who was 40 km away from the center where the patient was admitted [141]. Moreover, evaluating maxillofacial computed tomography scans using WhatsApp has been reported to be easy and rapid [142]. WhatsApp was also useful for communication between emergency department consultants when they were not onsite [143]. In dentistry, a study showed that 67.32% of dentists used WhatsApp to send clinical images to colleagues seeking second opinions, and 35.29% of them did so on a weekly basis. About 60.29% of the dentists received a prompt response, whereas 38.23% received delayed responses. In addition, about 98.52% of dentists sent radiographs on WhatsApp for a second opinion [144].

In conclusion, it is noteworthy that telemedicine is not *a practice in and of itself* [145]. It is not the most suitable model of care for every patient and is not the preferred approach when physicians cannot meet clinical standards of care. Patients using telemedicine must also have access to traditional emergency care, if needed. Although these innovations have significant potential benefits, the electronic exchange of health information and care may pose risks to patients' privacy, confidentiality, and safety and to quality and continuity of care. Furthermore, the limitations of electronically mediated physical examination may weaken the relationship between patients and HCPs, thereby jeopardizing care [119]. High-quality research is needed to improve the utilization of telemedicine, and more well-designed studies comparing telemedicine with traditional patient care are essential.

Scientific Research

The perceived benefits of using SM in health care include the ability to connect with geographically distant researchers and to build and foster research communities [4]. SM is a potential tool to revolutionize health research, as it has fewer temporal and spatial limitations and can overcome boundaries between research communities and the public [146,147]. SM can aid research in several ways: by recruiting participants, disseminating surveys, connecting with fellow researchers, identifying research opportunities, sharing study findings, and gaining access to published work.

There are conveniences in taking scientific research to the digital world. Publishing study findings on SM provides enhanced dissemination of research and increases the access to valid evidence-based information for patients. Furthermore, because not all studies end in a publication in a traditional journal, their findings can thus be shared via SM to a wider audience and be of substantial value to a broader research community [4]. Another advantage of SM for scientific research became evident during the COVID-19 pandemic, which made it possible to break geographical barriers and arrange collaborative research projects, surveys, and multicenter studies [8]. Sites such as Google Scholar and ResearchGate create communities for researchers to network, collaborate with each other, and promote publications [53,148]. The anonymity of posts, not having to answer questions in the presence of others and acquiring large samples that attenuate the effect of false information or extreme views were viewed as advantages unique to SM surveys and possible factors that improve research accuracy [146]. Content posted on the web may be used as data for research without interacting with the authors of the content, and perhaps without even considering them to be *human subjects* [146]. Moreover, compared with traditional recruitment methods, web-based surveys have the ability to store large numbers of responses, which can be easily accessed for analysis [5].

Recruiting research participants on SM has gained popularity in recent years. In a review by Lafferty and Manca [4], it was found that the most common tools used for recruiting participants were Facebook, Twitter, and a combination of both. Snowballing sampling method involves participants themselves recruiting more participants by contacting people in their networks [149]. A study on 8252 participants found that web-based recruitment was more efficient and had lower costs per recruited participant compared with traditional methods [150].

Disseminating surveys on the web is now a common practice. One study chose SM platforms to send its survey because it was cost effective, time saving, and easily accessible [151]. In another dentist or patient study, the survey for dentists was distributed via a dentist-only Facebook group that had more than 4500 members; for patients, the survey distribution was mainly through Facebook, LinkedIn, and Twitter, and the recipients were asked to share it with their connections [42]. Furthermore, in a study involving health care quality personnel, the survey was distributed through WhatsApp [114]. In a study in Saudi Arabia, the link to the web-based questionnaire was made available through Twitter and Facebook, the 2 most popular SNSs in the country [152]. Over half of university students strongly or somewhat liked using Facebook for research conducted by university researchers [153]. Zaballos et al [154] developed a web-based multiplatform that integrated WhatsApp and emails to assess the quality of life of individuals with hearing loss issues; the tool facilitated data collection in an easy-to-use platform.

A review by Topolovec-Vranic and Natarajan [155] showed that 40% of the studies found SM to be the most effective recruitment method, whereas 50% of them stated that their target population was *hard to reach*. Approximately 43% of the studies reported cost-effectiveness [155]. In addition, SM helped in recruiting a large number of individuals and reached challenging populations such as adolescents and young adults. Another review found that traditional recruitment methods tend to underrepresent users of marijuana, ecstasy, cocaine, or alcohol or people with at-risk sexual behavior; in comparison, Facebook recruitment yielded more representative results [156].

Researchers who plan to recruit participants on SM must consider their target populations' SM use patterns and preferences. For example, a study on sexual health might consider dating sites for recruitment, whereas Facebook may be more suitable for a nonsexual health study [157]. To best tailor recruitment campaigns, the selection of hashtags or keywords that reflect the interests of the target population might be useful [45].

Regarding shortcomings, it is important to note that research participants recruited from web-based environments may not truly represent the population of interest as a whole, suggesting that SM should only augment traditional recruitment methods [4]. A study suggested that people with disabilities may disproportionately be living in conditions with lower standards of living and may not have access to the internet [158]. In another study, subjects recruited from SM were largely middle class, whereas those recruited at a local hospital were more disadvantaged [159]. A review by Whitaker et al [156] showed an overrepresentation of young White women resulting from web-based recruitment.

Other limitations of using SM for research include that researchers have little control over distractions, the research idea may be copied, or participants may share research information with other participants, which puts the scientific integrity of the study at risk [4]. In a review by Denecke et al [5], the most reported ethical concerns for using SM for research recruitment were self-selection—that is, users with an interest in the study area will be recruited preferentially, which will affect the representativeness of the sample—and a skew toward well-educated and higher socioeconomic status cohorts on the web [5].

Ethical and privacy concerns regarding SM for research recruitment must be addressed because tracking, profiling, and targeting of users are common in the digital world [45]. Bender et al [160] proposed privacy-enhanced SM recruitment guidelines, including proactive measures to protect privacy and declaration of potential risks. Vulnerable groups such as children and teenagers, homosexuals in regions where homosexuality is illegal, and individuals with mental illnesses require extra emphasis on respect, confidentiality, and caution in obtaining consent [146].

To summarize, there is growing evidence to suggest that SM is a useful research tool that enables researchers to connect with each other, recruit participants, and share their findings with the public. Moreover, the data obtained from SM can be investigated. Nevertheless, researchers must not overlook the shortcomings of SM that may ultimately debilitate the integrity of the study. Privacy concerns and ethical considerations must also be considered. The development of guidelines for ethical conduct in web-based research should be based on the best available practices and should be comprehensive and

standardized to minimize a study's error margin. Future studies that compare different recruitment methods and varying participant demographics recruited using various methods should be encouraged. Research investigating the cost-effectiveness of SM research and those with large sample sizes that enable the generalizability of findings is also recommended.

Critical Public Health Care Issues

SM can be used by emergency notification systems to mass communicate information to large groups in a fast and low-cost manner. Studies have shown that SM can be a source of data to detect outbreaks, infection distribution, and areas of acute health care needs [29]. It can also help understand the public's knowledge, fears, attitudes, and behaviors during a crisis [161-163]. For example, the Red Cross tracks Twitter posts during natural disasters, such as hurricanes and earthquakes, to assess where the greatest needs lie [50,164]. Perhaps one of the first publications investigating SM use during a pandemic is a study that analyzed tweets posted during the 2009 H1N1 outbreak; this study found that SM can be a useful tool for disseminating information and for the public to share their opinions and experiences [165]. Twitter posts were also helpful in monitoring disease activity during the cholera and influenza outbreaks [166,167]. When interaction and collaboration were essential, as with the influenza A-H1N1 pandemic, SM provided an unmatched opportunity to engage the public and was used by prominent health organizations such as the WHO [7,41]. However, coordination between web-based and real-world response activities is also important [29].

Perhaps there is no more powerful example of SM use during a health crisis than what has happened during the COVID-19 pandemic. The dissemination of information during a pandemic has never been this quick and effective in the past. Information on the virus spread as quickly as the virus itself and dominated conversations on SM. On March 11, 2020, there were more than 20 million mentions of coronavirus-related terms on SM [168]. Since the beginning of the outbreak, SM has been one of the most commonly used communication channels by international health organizations such as the WHO and the CDC to possibly disseminate information to every person on earth with access to SM. Thousands of smaller health authorities may have also used SM to communicate with local communities. Although traditional access to medical guidelines and policies often requires some form of affiliation or membership, it is available to internet users today with a tap on a keyboard or a finger slide on a smartphone. The distribution of PPE, sharing treatment protocols, clinical trial results, and allocation of medical resources have been efficient with the aid of SM [8]. A recent study evaluated the 100 most viewed *coronavirus* videos on YouTube; as of March 5, 2020 (very early in the pandemic), these videos had 165 million views [169]. Another study in China collected data from 250 million Weibo users, a Twitter-like SM platform. Posts mentioning symptoms or diagnoses significantly predicted daily case counts ahead of the statistics announced by officials in Hubei Province, the epicenter of the initial outbreak, and the rest of China [170].

Perhaps the founders of Twitter did not expect it to become a helpful tool in the fight against COVID-19. For example, using Twitter, a cardiologist was able to expedite the delivery of a drug to a COVID-19 patient within just 6 hours of his tweet [171]. The American Heart Association launched a registry on Twitter to aggregate COVID-19 cases to better understand risk factor profiles and treatment algorithms [171]. Hashtags such as #GetUsPPE highlighted the scarcity of PPE, resulting in technology pioneers ramping up their production of PPE [171]. After calls were raised on Twitter and other SM platforms, HCPs flew to other parts of their countries that were in crisis, retired clinicians volunteered to rejoin the work force in several countries, and those who were unable to be present helped colleagues through telemedicine. Another example of SM use during the pandemic is the COVIDBRONCH Initiative—an international network of airway specialists who foster rapid acquisition and dissemination of knowledge regarding airway procedures during the pandemic [172].

Despite its catastrophic impact and the substantial loss of lives, humans will overcome the existential threat brought by COVID-19 and will also likely overcome future pandemics. Over time, humans have survived environmental, biological, and man-made calamities because of their innate adaptability, resilience, innovativeness, and persistence. Today, humans use SM to disseminate information quickly and to a large number of people, thus eliciting an almost immediate response. More research is already taking place and will continue to investigate the key role of SM in the fight against pandemics, not only from a medical perspective but also from social and economic viewpoints.

Conclusions

This review provided an overview of the different uses of SM in health care. It is evident that SM use indicates not a trend but a fundamental shift in how people communicate today. Multidimensional health care, which includes SM and other forms of communication, has been shown to be highly successful. Not only can SM be used to improve direct patient care but it can also be used to increase the public's knowledge, facilitate research, connect HCPs, improve medical education, and combat public health crises. However, striking the right balance between digital and traditional health care is imperative. As SM is a relatively recent phenomenon, further research is needed to determine its long-term effectiveness and to identify the best strategies for maximizing its advantages and limiting risks. This review will be continued in the second part, in which the use of SM from patients' perspectives will be discussed. This discussion will be supplemented with specific barriers, ethical considerations, and disadvantages reported in the extant literature.

Acknowledgments

The author acknowledges the steadfast support and telementoring of Professor Paul Casamassimo.

Conflicts of Interest

None declared.

References

1. Kaplan AM, Haenlein M. Users of the world, unite! The challenges and opportunities of Social Media. Business Horizons 2010 Jan;53(1):59-68. [doi: 10.1016/j.bushor.2009.09.003]
2. Ventola CL. Social media and health care professionals: benefits, risks, and best practices. P T 2014 Jul;39(7):491-520 [FREE Full text] [Medline: 25083128]
3. Dictionary M. Merriam-Webster. Social Media. URL: https://www.merriam-webster.com/dictionary/social%20media
4. Lafferty NT, Manca A. Perspectives on social media in and as research: A synthetic review. International Review of Psychiatry 2015 Mar 05;27(2):85-96. [doi: 10.3109/09540261.2015.1009419]
5. Denecke K, Bamidis P, Bond C, Gabarron E, Househ M, Lau AYS, et al. Ethical Issues of Social Media Usage in Healthcare. Yearb Med Inform 2018 Mar 10;24(01):137-147. [doi: 10.15265/iy-2015-001]
6. Statista. Number of social network users worldwide from 2017 to 2025. Statista. URL: https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/ [accessed 2020-07-23]
7. Levac JJ, O'Sullivan T. Social Media and its Use in Health Promotion. RISS-IJHS 2010 Feb 10;1(1):47. [doi: 10.18192/riss-ijhs.v1i1.1534]
8. González-Padilla DA, Tortolero-Blanco L. Social media influence in the COVID-19 Pandemic. Int. braz j urol 2020 Jul;46(suppl 1):120-124. [doi: 10.1590/s1677-5538.ibju.2020.s121]
9. Dhar V. Social Media and Its Implications in Pediatric Dentistry. Pediatr Dent 2018 May 15;40(3):174-176. [Medline: 29793562]
10. Prensky M. Digital Natives, Digital Immigrants. Prensky. 2001 Oct 23. URL: https://www.marcprensky.com/writing/Prensky%20-%20Digital%20Natives,%20Digital%20Immigrants%20-%20Part1.pdf [accessed 2020-10-23]
11. Tkalac Verčič A, Verčič D. Digital natives and social media. Public Relations Review 2013 Dec;39(5):600-602. [doi: 10.1016/j.pubrev.2013.08.008]
12. Helsper EJ, Eynon R. Digital natives: where is the evidence? British Educational Research Journal 2010 Jun;36(3):503-520. [doi: 10.1080/01411920902989227]
13. Sprout S. Social media demographics to inform your brand's strategy in 2020. Sprout. URL: https://sproutsocial.com/insights/new-social-media-demographics/ [accessed 2020-10-17]
14. Statista. Distribution of Instagram users worldwide as of July 2020, by age and gender. Statista. URL: https://www.statista.com/statistics/248769/age-distribution-of-worldwide-instagram-users/ [accessed 2020-10-17]
15. Statista. Distribution of Twitter users worldwide as of April 2020, by age group. Statista. URL: https://www.statista.com/statistics/283119/age-distribution-of-global-twitter-users/ [accessed 2020-07-23]
16. Statista. Number of Sina Weibo users in China from 2017 to 2021. Statista. 2002. URL: https://www.statista.com/statistics/941456/china-number-of-sina-weibo-users/ [accessed 2020-10-23]
17. Oberlo. 10 WhatApp Statistics. Oberlo. URL: https://www.oberlo.com/blog/whatsapp-statistics [accessed 2020-10-17]
18. Business Of Apps. WeChat Revenue and Usage Statistics (2020). Business Of Apps. URL: https://www.businessofapps.com/data/wechat-statistics/ [accessed 2020-10-22]
19. Omnicore. Linkedin by the Numbers: Stats, Demographics & Fun Facts. Omnicore. URL: https://www.omnicoreagency.com/linkedin-statistics/ [accessed 2020-10-23]
20. Facebook. About Facebook. Facebook. URL: https://about.fb.com/ [accessed 2020-10-23]
21. Instagram. About Instagram. Instagram. URL: https://about.instagram.com/ [accessed 2020-10-23]
22. WhatsApp. About WhatsApp. WhatsApp. URL: https://www.whatsapp.com/about/
23. Twitter. About Twitter. Twitter. URL: https://about.twitter.com/ [accessed 2020-10-24]
24. YouTube. About YouTube. YouTube. URL: https://www.youtube.com/about/ [accessed 2020-10-23]
25. LinkedIn. About LinkedIn. LinkedIn. URL: https://about.linkedin.com/ [accessed 2020-10-22]
26. Barreto JE, Whitehair CL. Social Media and Web Presence for Patients and Professionals: Evolving Trends and Implications for Practice. PM&R 2017 May 18;9:S98-S105. [doi: 10.1016/j.pmrj.2017.02.012]
27. Sprout S. 18 Facebook statistics every marketer should know in 2020. Sprout. URL: https://sproutsocial.com/insights/facebook-stats-for-marketers/ [accessed 2020-07-23]
28. Statista. Distribution of Facebook users worldwide as of April 2020, by age and gender. Statista. URL: https://www.statista.com/statistics/376128/facebook-global-user-age-distribution/ [accessed 2020-07-23]
29. Abd-Alrazaq A, Alhuwail D, Househ M, Hamdi M, Shah Z. Top Concerns of Tweeters During the COVID-19 Pandemic: Infoveillance Study. J Med Internet Res 2020 Apr 21;22(4):e19016. [doi: 10.2196/19016]
30. Bhola S, Hellyer P. The risks and benefits of social media in dental foundation training. Br Dent J 2016 Nov 18;221(10):609-613. [doi: 10.1038/sj.bdj.2016.854]
31. Nayak LM, Linkov G. Social Media Marketing in Facial Plastic Surgery. Facial Plastic Surgery Clinics of North America 2019 Aug;27(3):373-377. [doi: 10.1016/j.fsc.2019.04.002]

32. Antheunis ML, Tates K, Nieboer TE. Patients' and health professionals' use of social media in health care: Motives, barriers and expectations. Patient Education and Counseling 2013 Sep;92(3):426-431. [doi: 10.1016/j.pec.2013.06.020]
33. Edington DW, Schultz AB, Pitts JS, Camilleri A. The Future of Health Promotion in the 21st Century. American Journal of Lifestyle Medicine 2015 Sep 22;10(4):242-252. [doi: 10.1177/1559827615605789]
34. Chan TM, Stukus D, Leppink J, Duque L, Bigham BL, Mehta N, et al. Social Media and the 21st-Century Scholar: How You Can Harness Social Media to Amplify Your Career. Journal of the American College of Radiology 2018 Jan;15(1):142-148. [doi: 10.1016/j.jacr.2017.09.025]
35. Kamel Boulos M, Giustini D, Wheeler S. Instagram and WhatsApp in Health and Healthcare: An Overview. Future Internet 2016 Jul 26;8(3):37. [doi: 10.3390/fi8030037]
36. Adams SA. Revisiting the online health information reliability debate in the wake of "web 2.0": An inter-disciplinary literature and website review. International Journal of Medical Informatics 2010 Jun;79(6):391-400. [doi: 10.1016/j.ijmedinf.2010.01.006]
37. Gabarron E, Bradway M, Fernandez-Luque L, Chomutare T, Hansen AH, Wynn R, et al. Social media for health promotion in diabetes: study protocol for a participatory public health intervention design. BMC Health Serv Res 2018 Jun 5;18(1):1-5. [doi: 10.1186/s12913-018-3178-7]
38. We Go Health. 50 Social Media Healthcare Statistics to Watch. We Go Health. 2018 Jul 23. URL: https://www.wegohealth.com/2018/04/02/social-media-healthcare-statistics-to-watch/ [accessed 2020-07-23]
39. Househ M. The use of social media in healthcare: organizational, clinical, and patient perspectives. Stud Health Technol Inform 2013;183:244-248. [Medline: 23388291]
40. Baker J, Stanley A. Telemedicine Technology: a Review of Services, Equipment, and Other Aspects. Curr Allergy Asthma Rep 2018 Sep 26;18(11):1-8. [doi: 10.1007/s11882-018-0814-6]
41. Moorhead SA, Hazlett DE, Harrison L, Carroll JK, Irwin A, Hoving C. A New Dimension of Health Care: Systematic Review of the Uses, Benefits, and Limitations of Social Media for Health Communication. J Med Internet Res 2013 Apr 23;15(4):e85. [doi: 10.2196/jmir.1933]
42. Parmar N, Dong L, Eisingerich AB. Connecting With Your Dentist on Facebook: Patients' and Dentists' Attitudes Towards Social Media Usage in Dentistry. J Med Internet Res 2018 Jun 29;20(6):e10109. [doi: 10.2196/10109]
43. World Health Organization. Working together to tackle the Infodemic? World Health Organization. URL: https://www.euro.who.int/en/health-topics/Health-systems/e-health/news/news/2020/6/working-together-to-tackle-the-infodemic [accessed 2020-07-23]
44. Neville P, Waylen A. Social media and dentistry: some reflections on e-professionalism. Br Dent J 2015 Apr 24;218(8):475-478. [doi: 10.1038/sj.bdj.2015.294]
45. De Benedictis A, Lettieri E, Masella C, Gastaldi L, Macchini G, Santu C, et al. WhatsApp in hospital? An empirical investigation of individual and organizational determinants to use. PLoS ONE 2019 Jan 11;14(1):e0209873. [doi: 10.1371/journal.pone.0209873]
46. Dizon DS, Graham D, Thompson MA, Johnson LJ, Johnston C, Fisch MJ, et al. Practical Guidance: The Use of Social Media In Oncology Practice. JOP 2012 Sep;8(5):e114-e124. [doi: 10.1200/jop.2012.000610]
47. Chretien KC, Kind T. Social Media and Clinical Care. Circulation 2013 Apr 02;127(13):1413-1421. [doi: 10.1161/circulationaha.112.128017]
48. Stukus DR. Using Twitter to expand the reach and engagement of allergists. The Journal of Allergy and Clinical Immunology: In Practice 2016 Mar;4(2):345-346.e1. [doi: 10.1016/j.jaip.2015.11.002]
49. Gao X, Hamzah S, Yiu CKY, McGrath C, King NM. Dental Fear and Anxiety in Children and Adolescents: Qualitative Study Using YouTube. J Med Internet Res 2013 Feb 22;15(2):e29. [doi: 10.2196/jmir.2290]
50. George D. R., L.S. Rovniak, and J. L. Kraschnewski, Dangers and opportunities for social media in medicine. Clin Obstet Gynecol 2013;56(3):453-462. [doi: 10.1097/grf.0b013e318297dc38]
51. Stellefson M, Paige SR, Chaney BH, Chaney JD. Evolving Role of Social Media in Health Promotion: Updated Responsibilities for Health Education Specialists. IJERPH 2020 Feb 12;17(4):1153. [doi: 10.3390/ijerph17041153]
52. Dorfman RG, Mahmood E, Ren A, Turin SY, Vaca EEE, Fine NA, et al. Google Ranking of Plastic Surgeons Values Social Media Presence Over Academic Pedigree and Experience. Aesthet Surg J 2019;39(4):447-451. [doi: 10.1093/asj/sjy285]
53. Escoffery C, Kenzig M, Hyden C, Hernandez K. Capitalizing on Social Media for Career Development. Health Promotion Practice 2017 Sep 29;19(1):11-15. [doi: 10.1177/1524839917734522]
54. Alalawi A, Aljuaid H, Natto ZS. The Effect Of Social Media On The Choice Of Dental Patients: A Cross-Sectional Study In The City Of Jeddah, Saudi Arabia . PPA 2019 Oct;Volume 13:1685-1692. [doi: 10.2147/ppa.s213704]
55. Henry RK, Molnar A, Henry JC. A survey of US dental practices' use of social media. J Contemp Dent Pract 2012 Mar 01;13(2):137-141. [doi: 10.5005/jp-journals-10024-1109] [Medline: 22665737]
56. Simplício AHDM. Social media and Dentistry: ethical and legal aspects. Dental Press J. Orthod 2019 Dec;24(6):80-89. [doi: 10.1590/2177-6709.24.6.080-089.sar]
57. Gould D, Nazarian S. Social Media Return on Investment: How Much is it Worth to My Practice? Aesthet Surg J 2018;38(5):565-574. [doi: 10.1093/asj/sjx152]

58. Thompson MA, Younes A, Miller RS. Using social media in oncology for education and patient engagement. Oncology (Williston Park) 2012;26(9):84-85.
59. Carpentier M, Van Hoye G, Stockman S, Schollaert E, Van Theemsche B, Jacobs G. Recruiting nurses through social media: Effects on employer brand and attractiveness. J Adv Nurs 2017 Jun 13;73(11):2696-2708. [doi: 10.1111/jan.13336]
60. Parks-Yancy R, Cooley D. The Impact of Traditional and Internet/Social Media Screening Mechanisms on Employers? Perceptions of Job Applicants. Social Media in Society 2016;5(3):151-186.
61. Davison K, Bing M, Kluemper D, Roth P. Social media as a personnel selection and hiring resource: Reservations and recommendations. Social media in employee selection and recruitment: Theory, practice, and current challenges. Social Media as a Personnel Selection and Hiring Resource: Reservations and Recommendations 2016:15-42. [doi: 10.1007/978-3-319-29989-1_2]
62. Nikolaou I. Social Networking Web Sites in Job Search and Employee Recruitment. Int J Select Assess 2014 Apr 23;22(2):179-189. [doi: 10.1111/ijsa.12067]
63. Collins C, Kanar A. The oxford handbook of recruitment Employer brand equity and recruitment research. The oxford handbook of recruitment Employer brand equity and recruitment research 2014:15-42. [doi: 10.1093/oxfordhb/9780199756094.013.0016]
64. Kluemper DH, Rosen PA. Future employment selection methods: evaluating social networking web sites. Journal of Managerial Psych 2009 Aug 14;24(6):567-580. [doi: 10.1108/02683940910974134]
65. Baert S. Facebook profile picture appearance affects recruiters' first hiring decisions. New Media & Society 2017 Feb 01;20(3):1220-1239. [doi: 10.1177/1461444816687294]
66. Baert S, Decuypere L. Better sexy than flexy? A lab experiment assessing the impact of perceived attractiveness and personality traits on hiring decisions. Applied Economics Letters 2014 Mar 05;21(9):597-601. [doi: 10.1080/13504851.2013.877564]
67. Harolds J. Tips for a physician in getting the right job part III: Networking and social media, references, and recruiters. Clin Nucl Med 2013;38(10):805-807. [doi: 10.1097/rlu.0b013e31829e8146]
68. Jobvite. Social recruiting survey. Jobvite. 2014. URL: https://www.jobvite.com/wp-content/uploads/2014/10/Jobvite_SocialRecruiting_Survey2014.pdf [accessed 2020-07-23]
69. MacMillan C. Social media revolution and blurring of professional boundaries. Imprint 2013;60(3):44-46. [Medline: 23821819]
70. Peck JL. Social Media in Nursing Education: Responsible Integration for Meaningful Use. J Nurs Educ 2014 Feb 19;53(3):164-169. [doi: 10.3928/01484834-20140219-03]
71. Sterling M, Leung P, Wright D, Bishop TF. The Use of Social Media in Graduate Medical Education. Academic Medicine 2017;92(7):1043-1056. [doi: 10.1097/acm.0000000000001617]
72. Wells DM. When Faced With Facebook: What Role Should Social Media Play in Selecting Residents? J Grad Med Educ 2015;7(1):14-15. [doi: 10.4300/jgme-d-14-00363.1]
73. Chauhan B, George R, Coffin J. Social media and you: what every physician needs to know. J Med Pract Manage 2012;28(3):206-209. [doi: 10.4324/9780080460574-19]
74. Go PH, Klaassen Z, Chamberlain RS. Attitudes and Practices of Surgery Residency Program Directors Toward the Use of Social Networking Profiles to Select Residency Candidates: A Nationwide Survey Analysis. Journal of Surgical Education 2012 May;69(3):292-300. [doi: 10.1016/j.jsurg.2011.11.008]
75. Foley NM, Maher BM, Corrigan MA. Social media and tomorrow's medical students--how do they fit? J Surg Educ 2014;71(3):385-390. [doi: 10.1016/j.jsurg.2013.10.008] [Medline: 24797855]
76. Go P, Klaassen Z, Chamberlain R. Residency selection: do the perceptions of US programme directors and applicants match? Med Educ 2012;46(5):491-500. [doi: 10.1111/j.1365-2923.2012.04257.x]
77. Langenfeld SJ, Cook G, Sudbeck C, Luers T, Schenarts PJ. An Assessment of Unprofessional Behavior Among Surgical Residents on Facebook: A Warning of the Dangers of Social Media. Journal of Surgical Education 2014 Nov;71(6):e28-e32. [doi: 10.1016/j.jsurg.2014.05.013]
78. Cain J, Scott D, Smith K. Use of social media by residency program directors for resident selection. Am J Health Syst Pharm 2010;67(19):1635-1639. [doi: 10.2146/ajhp090658]
79. Cain J, Scott DR, Akers P. Pharmacy Students' Facebook Activity and Opinions Regarding Accountability and E-Professionalism. Am J Pharm Educ 2009 Sep;73(6):104. [doi: 10.5688/aj7306104]
80. Schulman CI, Kuchkarian FM, Withum KF, Boecker FS, Graygo JM. Influence of social networking websites on medical school and residency selection process. Postgrad Med J 2012 Nov 08;89(1049):126-130. [doi: 10.1136/postgradmedj-2012-131283]
81. McGinnis L. Presidential address: Professionalism in the 21st century. Bull Am Coll Surg 2009;94(12):8-18. [doi: 10.1142/9789812799647_0017]
82. Gorrindo T, Gorrindo PC, Groves JE. Intersection of Online Social Networking with Medical Professionalism: Can Medicine Police the Facebook Boom? J GEN INTERN MED 2008 Oct 2;23(12):2155-2155. [doi: 10.1007/s11606-008-0810-y]
83. Chretien KC. Online Posting of Unprofessional Content by Medical Students. JAMA 2009 Sep 23;302(12):1309. [doi: 10.1001/jama.2009.1387]

84. Mascia D, Rinninella E, Pennacchio NW, Cerrito L, Gasbarrini A. It's how we communicate! Exploring face-to-face versus electronic communication networks in multidisciplinary teams. Health Care Manage Rev 2019 Apr 18;46(2):153-161. [doi: 10.1097/hmr.0000000000000246]
85. Chan TM, Dzara K, Dimeo SP, Bhalerao A, Maggio LA. Social media in knowledge translation and education for physicians and trainees: a scoping review. Perspect Med Educ 2019 Dec 13;9(1):20-30. [doi: 10.1007/s40037-019-00542-7]
86. Ting DK, Thoma B, Luckett-Gatopoulos S, Thomas A, Syed S, Bravo M, et al. CanadiEM: Accessing a Virtual Community of Practice to Create a Canadian National Medical Education Institution. AEM Education and Training 2018 Nov 02;3(1):86-91. [doi: 10.1002/aet2.10199]
87. Wani SA, Rabah SM, AlFadil S, Dewanjee N, Najmi Y. Efficacy of communication amongst staff members at plastic and reconstructive surgery section using smartphone and mobile WhatsApp. Indian J Plast Surg 2019 Oct 07;46(03):502-505. [doi: 10.4103/0970-0358.121990]
88. Chan WS, Leung AY. Use of Social Network Sites for Communication Among Health Professionals: Systematic Review. J Med Internet Res 2018 Mar 28;20(3):e117. [doi: 10.2196/jmir.8382]
89. Grajales III FJ, Sheps S, Ho K, Novak-Lauscher H, Eysenbach G. Social Media: A Review and Tutorial of Applications in Medicine and Health Care. J Med Internet Res 2014 Feb 11;16(2):e13. [doi: 10.2196/jmir.2912]
90. Sermo Expands Global Network to Include 30 Countries. Sermo. 2016 Jul 24. URL: https://www.sermo.com/press-releases/sermo-expands-global-network-to-include-30-countries/ [accessed 2020-07-24]
91. Doximity. Doximity. Doximity. URL: https://www.doximity.com/ [accessed 2020-07-24]
92. Medical Directors Forum. Medical Directors Forum. Medical Directors Forum. URL: https://medicaldirectorsforum.com/ [accessed 2020-07-24]
93. Khanna V, Sambandam SN, Gul A, Mounasamy V. "WhatsApp"ening in orthopedic care: a concise report from a 300-bedded tertiary care teaching center. Eur J Orthop Surg Traumatol 2015 Jan 30;25(5):821-826. [doi: 10.1007/s00590-015-1600-y]
94. Eysenbach G. Can Tweets Predict Citations? Metrics of Social Impact Based on Twitter and Correlation with Traditional Metrics of Scientific Impact. J Med Internet Res 2011 Dec 16;13(4):e123. [doi: 10.2196/jmir.2012]
95. Allen HG, Stanton TR, Di Pietro F, Moseley GL. Social Media Release Increases Dissemination of Original Articles in the Clinical Pain Sciences. PLoS ONE 2013 Jul 17;8(7):e68914. [doi: 10.1371/journal.pone.0068914]
96. Teoh JY, Mackenzie G, Tortolero L, Rivas JG. Social Media Analytics: What You Need to Know as a Urologist. European Urology Focus 2020 May;6(3):434-436. [doi: 10.1016/j.euf.2019.08.005]
97. Radford DR, Hellyer P. Belongingness in undergraduate dental education. Br Dent J 2016 May 27;220(10):539-543. [doi: 10.1038/sj.bdj.2016.379] [Medline: 27228935]
98. Brooks SK, Webster RK, Smith LE, Woodland L, Wessely S, Greenberg N, et al. The psychological impact of quarantine and how to reduce it: rapid review of the evidence. The Lancet 2020 Mar;395(10227):912-920. [doi: 10.1016/s0140-6736(20)30460-8]
99. Kaiser health News. Dentists, Physicians Offices Hit Hardest During April's Loss Of 1.4M Health Care Jobs. Kaiser health News. URL: https://khn.org/morning-breakout/dentists-physicians-offices-hit-hardest-during-aprils-loss-of-1-4m-health-care-jobs/ [accessed 2020-07-24]
100. Lee A. A Facebook group matches RVs that are sitting idle with health care workers who need a place to isolate after long hospital shifts. CNN.com. 2018. URL: https://edition.cnn.com/2020/04/01/us/coronavirus-rvs-4-mds-healthcare-isolate-trnd/index.html [accessed 2020-07-24]
101. de Peralta TL, Farrior OF, Flake NM, Gallagher D, Susin C, Valenza J. The Use of Social Media by Dental Students for Communication and Learning: Two Viewpoints. Journal of Dental Education 2019 Jun 01;83(6):663-668. [doi: 10.21815/jde.019.072]
102. Cheston CC, Flickinger TE, Chisolm MS. Social Media Use in Medical Education. Academic Medicine 2013;88(6):893-901. [doi: 10.1097/acm.0b013e31828ffc23]
103. Camm CF, Sunderland N, Camm AJ. A Quality Assessment of Cardiac Auscultation Material on YouTube. Clin Cardiol 2012 Nov 21;36(2):77-81. [doi: 10.1002/clc.22080]
104. Knösel M, Jung K, Bleckmann A. YouTube, dentistry, and dental education. J Dent Educ 2011 Dec;75(12):1558-1568. [Medline: 22184594]
105. Alsuraihi AK, Almaqati AS, Abughanim SA, Jastaniah NA. Use of social media in education among medical students in Saudi Arabia. Korean J Med Educ 2016 Dec;28(4):343-354. [doi: 10.3946/kjme.2016.40]
106. Kind T, Genrich G, Sodhi A, Chretien KC. Social media policies at US medical schools. Medical Education Online 2010 Sep 15;15(1):5324. [doi: 10.3402/meo.v15i0.5324]
107. Gonzalez SM, Gadbury-Amyot CC. Using Twitter for Teaching and Learning in an Oral and Maxillofacial Radiology Course. Journal of Dental Education 2016 Feb 01;80(2):149-155. [doi: 10.1002/j.0022-0337.2016.80.2.tb06070.x]
108. Forbes M, Hickey M. Podcasting: implementation and evaluation in an undergraduate nursing program. Nurse Educ 2008;33(5):224-227. [doi: 10.1097/01.nne.0000334775.98018.e8]
109. Carvas M, Imamura M, Hsing W, Dewey-Platt L, Fregni F. An innovative method of global clinical research training using collaborative learning with Web 2.0 tools. Med Teach 2010;32(3):270. [doi: 10.4018/978-1-61520-963-7.ch012]

110. Oyewole BK, Animasahun VJ, Chapman HJ. A survey on the effectiveness of WhatsApp for teaching doctors preparing for a licensing exam. PLoS ONE 2020 Apr 2;15(4):e0231148. [doi: 10.1371/journal.pone.0231148]
111. Scaffidi MA, Khan R, Wang C, Keren D, Tsui C, Garg A, et al. Comparison of the Impact of Wikipedia, UpToDate, and a Digital Textbook on Short-Term Knowledge Acquisition Among Medical Students: Randomized Controlled Trial of Three Web-Based Resources. JMIR Med Educ 2017 Oct 31;3(2):e20. [doi: 10.2196/mededu.8188]
112. George D, Dellasega C. Social media in medical education: two innovative pilot studies. Med Educ 2011;45(11):1158-1159. [doi: 10.1111/j.1365-2923.2011.04124.x]
113. Ranginwala S, Towbin AJ. Use of Social Media in Radiology Education. Journal of the American College of Radiology 2018 Jan;15(1):190-200. [doi: 10.1016/j.jacr.2017.09.010]
114. Alanzi T, Al-Habib DK. The Use of Social Media by Healthcare Quality Personnel in Saudi Arabia. Journal of Environmental and Public Health 2020 May 22;2020:1-9. [doi: 10.1155/2020/1417478]
115. Kaliyadan F, Ashique K, Jagadeesan S, Krishna B. What's up dermatology? A pilot survey of the use of WhatsApp in dermatology practice and case discussion among members of WhatsApp dermatology groups. Indian J Dermatol Venereol Leprol 2016;82(1):67. [doi: 10.4103/0378-6323.171638]
116. Johnston MJ, King D, Arora S, Behar N, Athanasiou T, Sevdalis N, et al. Smartphones let surgeons know WhatsApp: an analysis of communication in emergency surgical teams. The American Journal of Surgery 2015 Jan;209(1):45-51. [doi: 10.1016/j.amjsurg.2014.08.030]
117. Nettle M. Social media in medical education: Can you trust it? Yes. Emergency Medicine Australasia 2018 May 06;30(3):416-417. [doi: 10.1111/1742-6723.13100]
118. Chan R. Social media in medical education: Can you trust it? No. Emergency Medicine Australasia 2018 May 06;30(3):418-419. [doi: 10.1111/1742-6723.13101]
119. Chaet D, Clearfield R, Sabin JE, Skimming K. Ethical practice in Telehealth and Telemedicine. J Gen Intern Med 2017 Jun 26;32(10):1136-1140. [doi: 10.1007/s11606-017-4082-2]
120. Shah MN, McDermott R, Gillespie SM, Philbrick EB, Nelson D. Potential of Telemedicine to Provide Acute Medical Care for Adults in Senior Living Communities. Acad Emerg Med 2013 Feb 13;20(2):162-168. [doi: 10.1111/acem.12075]
121. AlShaya MS, Assery MK, Pani SC. Reliability of mobile phone teledentistry in dental diagnosis and treatment planning in mixed dentition. J Telemed Telecare 2018 Aug 22;26(1-2):45-52. [doi: 10.1177/1357633x18793767]
122. Klaassen B, van Beijnum BJF, Hermens HJ. Usability in telemedicine systems-A literature survey. Int J Med Inform 2016 Sep;93:57-69. [doi: 10.1016/j.ijmedinf.2016.06.004] [Medline: 27435948]
123. Turvey C, Coleman M, Dennison O, Drude K, Goldenson M, Hirsch P, et al. ATA Practice Guidelines for Video-Based Online Mental Health Services. Telemedicine and e-Health 2013 Sep;19(9):722-730. [doi: 10.1089/tmj.2013.9989]
124. Burke BL, Hall RW. Telemedicine: Pediatric Applications. Pediatrics 2015 Jun 29;136(1):e293-e308. [doi: 10.1542/peds.2015-1517]
125. Fortney JC, Pyne JM, Kimbrell TA, Hudson TJ, Robinson DE, Schneider R, et al. Telemedicine-Based Collaborative Care for Posttraumatic Stress Disorder. JAMA Psychiatry 2015 Jan 01;72(1):58. [doi: 10.1001/jamapsychiatry.2014.1575]
126. Estai M, Kanagasingam Y, Huang B, Checker H, Steele L, Kruger E, et al. The efficacy of remote screening for dental caries by mid-level dental providers using a mobile teledentistry model. Community Dent Oral Epidemiol 2016 Apr 25;44(5):435-441. [doi: 10.1111/cdoe.12232]
127. Lee JY, Lee SWH. Telemedicine Cost–Effectiveness for Diabetes Management: A Systematic Review. Diabetes Technology & Therapeutics 2018 Jul;20(7):492-500. [doi: 10.1089/dia.2018.0098]
128. Simon P. [Telemedicine in current nursing practices]. Soins 2016 Nov;61(810):24-27 [FREE Full text] [doi: 10.1016/j.soin.2016.09.003] [Medline: 27894474]
129. Portnoy JM, Waller M, Farrell K. Telemedicine Improves Allergy Care for Hospitalized Patients. The Journal of Allergy and Clinical Immunology: In Practice 2018 Nov;6(6):2041-2042. [doi: 10.1016/j.jaip.2018.05.016]
130. Worster B, Swartz K. Telemedicine and Palliative Care: an Increasing Role in Supportive Oncology. Curr Oncol Rep 2017 Apr 18;19(6):1-5. [doi: 10.1007/s11912-017-0600-y]
131. Dixon RF, Rao L. Asynchronous Virtual Visits for the Follow-Up of Chronic Conditions. Telemedicine and e-Health 2014 Jul;20(7):669-672. [doi: 10.1089/tmj.2013.0211]
132. van Velthoven MH, Li Y, Wang W, Du X, Wu Q, Chen L, et al. mHealth Series: mHealth project in Zhao County, rural China – Description of objectives, field site and methods. J Glob Health 2013 Dec;3(2):1-22. [doi: 10.7189/jogh.03.020401]
133. Leung R, Guo H, Pan X. Social Media Users' Perception of Telemedicine and mHealth in China: Exploratory Study. JMIR Mhealth Uhealth 2018 Sep 25;6(9):e181. [doi: 10.2196/mhealth.7623]
134. van den Heuvel JF, Groenhof TK, Veerbeek JH, van Solinge WW, Lely AT, Franx A, et al. eHealth as the Next-Generation Perinatal Care: An Overview of the Literature. J Med Internet Res 2018 Jun 05;20(6):e202. [doi: 10.2196/jmir.9262]
135. Irving M, Stewart R, Spallek H, Blinkhorn A. Using teledentistry in clinical practice as an enabler to improve access to clinical care: A qualitative systematic review. J Telemed Telecare 2017 Jan 16;24(3):129-146. [doi: 10.1177/1357633x16686776]

136. Joshi SS, Murali-Krishnan S, Patankar P, Choudhari KA. Neurosurgical referral service using smartphone client WhatsApp: preliminary study at a tertiary referral neurosurgical unit. British Journal of Neurosurgery 2018 Jul 13;32(5):553-557. [doi: 10.1080/02688697.2018.1490944]
137. Dhiliwal S, Salins N. Smartphone applications in palliative homecare. Indian J Palliat Care 2015;21(1):88. [doi: 10.4103/0973-1075.150199]
138. Choudhari P. Study on effectiveness of communication amongst members at department of orthopedics surgery unit 3 using smartphone and mobile WhatsApp. Int Surg J 2014;1(1):9. [doi: 10.5455/2349-2902.isj20140504]
139. Hollander JE, Carr BG. Virtually Perfect? Telemedicine for Covid-19. N Engl J Med 2020 Apr 30;382(18):1679-1681. [doi: 10.1056/nejmp2003539]
140. Giordano V, Koch HA, Mendes CH, Bergamin A, de Souza FS, do Amaral NP. WhatsApp messenger is useful and reproducible in the assessment of tibial plateau fractures: Inter- and intra-observer agreement study. International Journal of Medical Informatics 2015 Feb;84(2):141-148. [doi: 10.1016/j.ijmedinf.2014.11.002]
141. Thota R, Divatia J. WhatsApp: What an App!. Indian J Crit Care Med 2015;19(6):363-365. [doi: 10.4103/0972-5229.158288]
142. Kelahmetoglu O, Firincioglullari R, Yagmur C. Efficient Utility of WhatsApp: From Computer Screen to the Surgeon's Hand to Determine Maxillofacial Traumas. J Craniofac Surg 2015;26(4):1437. [doi: 10.1097/scs.0000000000001627]
143. Gulacti U, Lok U, Hatipoglu S, Polat H. An Analysis of WhatsApp Usage for Communication Between Consulting and Emergency Physicians. J Med Syst 2016 Apr 15;40(6):1-7. [doi: 10.1007/s10916-016-0483-8]
144. Sarode S, Sarode G, Gaikwad T, Patekar, D, Gadbail A, Gondivkar S, et al. Usage Analysis of WhatsApp for Dentistry-related Purposes among General Dental Practitioners. J Contemp Dent Pract 2018;19(10):1267-1272. [doi: 10.5005/jp-journals-10024-2415]
145. Merrell RC, Doarn CR. Ethics in Telemedicine Research. Telemedicine and e-Health 2009 Mar;15(2):123-124. [doi: 10.1089/tmj.2009.9992]
146. Golder S, Ahmed S, Norman G, Booth A. Attitudes Toward the Ethics of Research Using Social Media: A Systematic Review. J Med Internet Res 2017 Jun 06;19(6):e195. [doi: 10.2196/jmir.7082]
147. Fordis M, Street RL, Volk RJ, Smith Q. The Prospects for Web 2.0 Technologies for Engagement, Communication, and Dissemination in the Era of Patient-Centered Outcomes Research: Selected Articles Developed From the Eisenberg Conference Series 2010 Meeting. Journal of Health Communication 2011 Jul 29;16(sup1):3-9. [doi: 10.1080/10810730.2011.598398]
148. Thelwall M, Kousha K. ResearchGate: Disseminating, communicating, and measuring Scholarship? J Assn Inf Sci Tec 2014 May 23;66(5):876-889. [doi: 10.1002/asi.23236]
149. Child RJH, Mentes JC, Pavlish C, Phillips LR. Using Facebook and participant information clips to recruit emergency nurses for research. Nurse Researcher 2014 Jul 25;21(6):16-21. [doi: 10.7748/nr.21.6.16.e1246]
150. Christensen T, Riis AH, Hatch EE, Wise LA, Nielsen MG, Rothman KJ, et al. Costs and Efficiency of Online and Offline Recruitment Methods: A Web-Based Cohort Study. J Med Internet Res 2017 Mar 01;19(3):e58. [doi: 10.2196/jmir.6716]
151. Sudha B, Samuel AJ, Narkeesh K. Feasibility online survey to estimate physical activity level among the students studying professional courses: a cross-sectional online survey. J Exerc Rehabil 2018 Feb 26;14(1):58-63. [doi: 10.12965/jer.1835130.565]
152. AlShahrani I, Tikare S, Togoo R, AlAsere Y, AlAsmari A. Self-perception of personal oral health in Saudi population: a social media approach. Easter Mediterr Health J 2015 May 01;21(5):342-348. [doi: 10.26719/2015.21.5.342]
153. Moreno MA, Grant A, Kacvinsky L, Moreno P, Fleming M. Older Adolescents' Views Regarding Participation in Facebook Research. Journal of Adolescent Health 2012 Nov;51(5):439-444. [doi: 10.1016/j.jadohealth.2012.02.001]
154. Pérez Zaballos M, Ramos Macías, Pérez Placencia D, Borkoski Barreiro S, Ramos de Miguel. LifeQuestionnaire. A new tool for the evaluation of quality of life in patients with hearing loss-using WhatsApp. European Annals of Otorhinolaryngology, Head and Neck Diseases 2016 Jun;133:S44-S49. [doi: 10.1016/j.anorl.2016.04.020]
155. Topolovec-Vranic J, Natarajan K. The Use of Social Media in Recruitment for Medical Research Studies: A Scoping Review. J Med Internet Res 2016 Nov 07;18(11):e286. [doi: 10.2196/jmir.5698]
156. Whitaker C, Stevelink S, Fear N. The Use of Facebook in Recruiting Participants for Health Research Purposes: A Systematic Review. J Med Internet Res 2017 Aug 28;19(8):e290. [doi: 10.2196/jmir.7071]
157. Holloway IW, Rice E, Gibbs J, Winetrobe H, Dunlap S, Rhoades H. Acceptability of Smartphone Application-Based HIV Prevention Among Young Men Who Have Sex With Men. AIDS Behav 2013 Dec 1;18(2):285-296. [doi: 10.1007/s10461-013-0671-1]
158. Morris R. 'Unjust, inhumane and highly inaccurate': the impact of changes to disability benefits and services – social media as a tool in research and activism. Disability & Society 2013 Jul;28(5):724-728. [doi: 10.1080/09687599.2013.808093]
159. Balfe M, Doyle F, Conroy R. Using Facebook to recruit young adults for qualitative research projects: how difficult is it? Comput Inform Nurs 2012;30(10):511-515. [doi: 10.1097/nxn.0b013e31826e4fca]
160. Bender JL, Cyr AB, Arbuckle L, Ferris LE. Ethics and Privacy Implications of Using the Internet and Social Media to Recruit Participants for Health Research: A Privacy-by-Design Framework for Online Recruitment. J Med Internet Res 2017 Apr 06;19(4):e104. [doi: 10.2196/jmir.7029]

161. Jordan S, Hovet S, Fung I, Liang H, Fu K, Tse Z. Using Twitter for Public Health Surveillance from Monitoring and Prediction to Public Response. Data 2018 Dec 29;4(1):6. [doi: 10.3390/data4010006]
162. Shah Z, Surian D, Dyda A, Coiera E, Mandl KD, Dunn AG. Automatically Appraising the Credibility of Vaccine-Related Web Pages Shared on Social Media: A Twitter Surveillance Study. J Med Internet Res 2019 Nov 4;21(11):e14007. [doi: 10.2196/14007]
163. Shah Z, Dunn AG. Event detection on Twitter by mapping unexpected changes in streaming data into a spatiotemporal lattice. IEEE Trans. Big Data 2019:1-1. [doi: 10.1109/tbdata.2019.2948594]
164. Fogelson N, Rubin Z, Ault A. Beyond likes and tweets: an in-depth look at the physician social media landscape. Clin Obstet Gynecol 2013;56(3):a-508. [doi: 10.1097/grf.0b013e31829e7638]
165. Chew C, Eysenbach G. Pandemics in the Age of Twitter: Content Analysis of Tweets during the 2009 H1N1 Outbreak. PLoS ONE 2010 Nov 29;5(11):e14118. [doi: 10.1371/journal.pone.0014118]
166. Collier N, Son N, Nguyen N. OMG U got flu? Analysis of shared health messages for bio-surveillance. J Biomed Sem 2011;2(Suppl 5):S9. [doi: 10.1186/2041-1480-2-s5-s9]
167. Denecke K, Krieck M, Otrusina L, Smrz P, Dolog P, Nejdl W, et al. How to Exploit Twitter for Public Health Monitoring? Methods Inf Med 2018 Jan 20;52(04):326-339. [doi: 10.3414/me12-02-0010]
168. Molla R. How coronavirus took over social media. vox.com. URL: https://www.vox.com/recode/2020/3/12/21175570/coronavirus-covid-19-social-media-twitter-facebook-google [accessed 2020-10-22]
169. Basch CH, Hillyer GC, Meleo-Erwin ZC, Jaime C, Mohlman J, Basch CE. Preventive Behaviors Conveyed on YouTube to Mitigate Transmission of COVID-19: Cross-Sectional Study. JMIR Public Health Surveill 2020;6(2):e18807. [doi: 10.2196/preprints.18807]
170. Shen C, Chen A, Luo C, Zhang J, Feng B, Liao W. Using Reports of Symptoms and Diagnoses on Social Media to Predict COVID-19 Case Counts in Mainland China: Observational Infoveillance Study. J Med Internet Res 2020;22(5):e19421. [doi: 10.2196/preprints.19421]
171. Aggarwal NR, Alasnag M, Mamas MA. Social media in the era of COVID-19. Open Heart 2020 Sep 29;7(2):e001352. [doi: 10.1136/openhrt-2020-001352]
172. Mulrennan S, Colt H. Medical information and social media in the time of covid‐19. Respirology 2020 Apr 28;25(6):578-579. [doi: 10.1111/resp.13832]

Abbreviations

CDC: Centers for Disease Control and Prevention
HCP: health care provider
PD: program director
PPE: personal protective equipment
ROI: return on investment
SM: social media
SNS: social networking site
WHO: World Health Organization

Edited by R Kukafka; submitted 04.08.20; peer-reviewed by C Escobar-Viera, E Said-Hung, O Anikeeva; comments to author 23.09.20; revised version received 05.11.20; accepted 04.03.21; published 05.04.21.

Please cite as:
Farsi D
Social Media and Health Care, Part I: Literature Review of Social Media Use by Health Care Providers
J Med Internet Res 2021;23(4):e23205
URL: https://www.jmir.org/2021/4/e23205
doi:10.2196/23205
PMID:33664014

Review

Impact of Big Data Analytics on People's Health: Overview of Systematic Reviews and Recommendations for Future Studies

Israel Júnior Borges do Nascimento[1,2], ClinPath, PharmB; Milena Soriano Marcolino[3,4], MD, MSc, PhD; Hebatullah Mohamed Abdulazeem[5], MBBS; Ishanka Weerasekara[6,7], PhD; Natasha Azzopardi-Muscat[8], MD, MPH, MSc, PhD; Marcos André Gonçalves[9], PhD; David Novillo-Ortiz[8], MLIS, MSc, PhD

[1]School of Medicine, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil

[2]Department of Medicine, School of Medicine, Medical College of Wisconsin, Wauwatosa, WI, United States

[3]Department of Internal Medicine, University Hospital, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil

[4]School of Medicine and Telehealth Center, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil

[5]Department of Sport and Health Sciences, Technical University Munich, Munich, Germany

[6]School of Health Sciences, Faculty of Health and Medicine, The University of Newcastle, Callaghan, Australia

[7]Department of Physiotherapy, Faculty of Allied Health Sciences, University of Peradeniya, Peradeniya, Sri Lanka

[8]Division of Country Health Policies and Systems, World Health Organization, Regional Office for Europe, Copenhagen, Denmark

[9]Department of Computer Science, Institute of Exact Sciences, Federal University of Minas Gerais, Belo Horizonte, Minas Gerais, Brazil

Corresponding Author:
David Novillo-Ortiz, MLIS, MSc, PhD
Division of Country Health Policies and Systems
World Health Organization, Regional Office for Europe
Marmorej 51
Copenhagen, 2100
Denmark
Phone: 45 61614868
Email: dnovillo@who.int

Abstract

Background: Although the potential of big data analytics for health care is well recognized, evidence is lacking on its effects on public health.

Objective: The aim of this study was to assess the impact of the use of big data analytics on people's health based on the health indicators and core priorities in the World Health Organization (WHO) General Programme of Work 2019/2023 and the European Programme of Work (EPW), approved and adopted by its Member States, in addition to SARS-CoV-2–related studies. Furthermore, we sought to identify the most relevant challenges and opportunities of these tools with respect to people's health.

Methods: Six databases (MEDLINE, Embase, Cochrane Database of Systematic Reviews via Cochrane Library, Web of Science, Scopus, and Epistemonikos) were searched from the inception date to September 21, 2020. Systematic reviews assessing the effects of big data analytics on health indicators were included. Two authors independently performed screening, selection, data extraction, and quality assessment using the AMSTAR-2 (A Measurement Tool to Assess Systematic Reviews 2) checklist.

Results: The literature search initially yielded 185 records, 35 of which met the inclusion criteria, involving more than 5,000,000 patients. Most of the included studies used patient data collected from electronic health records, hospital information systems, private patient databases, and imaging datasets, and involved the use of big data analytics for noncommunicable diseases. "Probability of dying from any of cardiovascular, cancer, diabetes or chronic renal disease" and "suicide mortality rate" were the most commonly assessed health indicators and core priorities within the WHO General Programme of Work 2019/2023 and the EPW 2020/2025. Big data analytics have shown moderate to high accuracy for the diagnosis and prediction of complications of diabetes mellitus as well as for the diagnosis and classification of mental disorders; prediction of suicide attempts and behaviors; and the diagnosis, treatment, and prediction of important clinical outcomes of several chronic diseases. Confidence in the results was rated as "critically low" for 25 reviews, as "low" for 7 reviews, and as "moderate" for 3 reviews. The most frequently identified challenges were establishment of a well-designed and structured data source, and a secure, transparent, and standardized database for patient data.

Conclusions: Although the overall quality of included studies was limited, big data analytics has shown moderate to high accuracy for the diagnosis of certain diseases, improvement in managing chronic diseases, and support for prompt and real-time analyses of large sets of varied input data to diagnose and predict disease outcomes.

Trial Registration: International Prospective Register of Systematic Reviews (PROSPERO) CRD42020214048; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=214048

(J Med Internet Res 2021;23(4):e27275) doi:10.2196/27275

KEYWORDS

public health; big data; health status; evidence-based medicine; big data analytics; secondary data analysis; machine learning; systematic review; overview; World Health Organization

Introduction

Big data analytics tools handle complex datasets that traditional data processing systems cannot efficiently and economically store, manage, or process. Through the application of artificial intelligence (AI) algorithms and machine learning (ML), big data analytics has potential to revolutionize health care, supporting clinicians, providers, and policymakers for planning or implementing interventions [1], faster disease detection, therapeutic decision support, outcome prediction, and increased personalized medicine, resulting in lower-cost, higher-quality care with better outcomes [1,2].

In 2018, the World Health Organization (WHO) proposed the expedited 13th General Programme of Work (GPW13), which was approved and adopted by its 194 Member States, focusing on measurable impacts on people's health at the state level to transform public health with three core features: enhanced universal health coverage, health emergencies protection, and better health and well-being [3]. Forty-six outcome target indicators emerged from the GPW13, covering a range of health issues [3]. Big data analytics may help to support health policy decision-making, accelerate the achievement of the GPW13 core priorities and targets, and guide the roadmap for the European region based on the European Programme of Work (EPW) 2020/2025 [4,5].

Therefore, the aim of this study was to provide an overview of systematic reviews that assessed the effects of the use of big data analytics on people's health according to the WHO core features defined in the GPW13 and the EPW. We included complex reviews that assessed multiple interventions, different populations, and differing outcomes resulting from big data analytics on people's health, and identified the challenges, opportunities, and best practices for future research.

Methods

Study Design

This study was designed to provide an overview of systematic reviews in accordance with guidelines from the Cochrane Handbook for Systematic Reviews of Interventions, along with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) and the QUOROM (Quality of Reporting of Meta-analyses) guidelines [6-8]. The study protocol is published on PROSPERO (CRD42020214048).

Search Strategy

To identify records assessing the effect of big data analytics on people's health, aligned with the WHO health indicators defined in the GPW13 (Textbox 1), a comprehensive and systematic search was performed using six multidisciplinary databases from their inception to September 21, 2020. The search strategy was designed in collaboration with a senior librarian and is described in detail in Multimedia Appendix 1.

References were imported into reference management software (EndNote X9) and duplicates were removed. Unique records were uploaded onto the Covidence Platform (Veritas Health Innovation) for screening, data extraction, and quality assessment. A manual search of reference lists was performed to supplement the search.

Textbox 1. List of 46 World Health Organization health indicators defined at the Thirteenth General Programme of Work.

- Number of persons affected by disasters (per 100,000 population)
- Domestic general government health expenditure (% of general government expenditure)
- Prevalence of stunting in children under 5 (%)
- Prevalence of wasting in children under 5 (%)
- Prevalence of overweight in children under 5 (%)
- Maternal mortality ratio (per 100,000 live births)
- Proportion of births attended by skilled health personnel (%)
- Under 5 mortality rate (per 1000 live births)
- Neonatal mortality rate (per 1000 live births)
- New HIV infections (per 1000 uninfected population)
- Tuberculosis incidence (per 100,000 population)
- Malaria incidence (per 1000 population at risk)
- Hepatitis B incidence (measured by surface antigen [HBsAg] prevalence among children under 5 years)
- Number of people requiring interventions against neglected tropical diseases (NTDs)
- Probability of dying from any of cardiovascular disease (CVD), cancer, diabetes, chronic renal disease (CRD) (aged 30-70 years) (%)
- Suicide mortality rate (per 100,000 population)
- Coverage of treatment interventions for substance-use disorders (%)
- Total alcohol per capita consumption in adults aged >15 years (liters of pure alcohol)
- Road traffic mortality rate (per 100,000 population)
- Proportion of women (aged 15-49 years) having need for family planning satisfied with modern methods (%)
- Universal Health Coverage (UHC) Service Coverage Index
- Population with household expenditures on health >10% of total household expenditure or income (%)
- Mortality rate attributed to air pollution (per 100,000 population)
- Mortality rate attributed to exposure to unsafe water, sanitation, and hygiene (WASH) services (per 100,000 population)
- Mortality rate from unintentional poisoning (per 100,000 population)
- Prevalence of tobacco use in adults aged ≥15 years (%)
- Proportion of population covered by all vaccines included in national programs (diphtheria-tetanus-pertussis vaccine, measles-containing-vaccine second dose, pneumococcal conjugated vaccine) (%)
- Proportion of health facilities with essential medicines available and affordable on a sustainable basis (%)
- Density of health workers (doctors, nurse and midwives, pharmacists, dentists per 10,000 population)
- International Health Regulations capacity and health emergency preparedness
- Proportion of bloodstream infections due to antimicrobial-resistant organisms (%)
- Proportion of children under 5 years developmentally on track (health, learning, and psychosocial well-being) (%)
- Proportion of women (aged 15-49 years) subjected to violence by current or former intimate partner (%)
- Proportion of women (aged 15-49 years) who make their own decisions regarding sexual relations, contraceptive use, and reproductive health care (%)
- Proportion of population using safely managed drinking-water services (%)
- Proportion of population using safely managed sanitation services and hand-washing facilities (%)
- Proportion of population with primary reliance on clean fuels (%)
- Annual mean concentrations of fine particulate matter (PM2.5) in urban areas ($\mu g/m^3$)
- Proportion of children (aged 1-17 years) experiencing physical or psychological aggression (%)
- Vaccine coverage for epidemic-prone diseases
- Proportion of vulnerable people in fragile settings provided with essential health services (%)

- Prevalence of raised blood pressure in adults aged ≥18 years
- Effective policy/regulation for industrially produced trans-fatty acids
- Prevalence of obesity (%)
- Number of cases of poliomyelitis caused by wild poliovirus
- Patterns of antibiotic consumption at the national level

Study Selection

Peer-reviewed publications categorized as systematic reviews assessing the effects of big data analytics on any of the GPW13 and EPW health indicators and core priorities were included, regardless of language and study design. We only considered studies in which the search was performed in at least two databases, and included a description of the search strategy and the methodology used for study selection and data extraction. We only included studies that evaluated concrete relationships between the use of big data analytics and its effect on people's lives, according to the WHO strategic priorities and indicators. Along with the 46 indicators listed in Textbox 1, we also included studies evaluating the use of big data during the COVID-19 pandemic. To identify gaps, we included reviews focusing on challenges, best practices, and short- and long-term opportunities related to big data technologies. Nonsystematic reviews, primary studies, opinions, short communications, nonscientific articles, conference abstracts, and reviews with big data inappropriately defined were excluded.

Although big data analysis is capable of handling large volumes of data, rather than focusing on the data volume/size, we focused on the process that defines big data analytics, which includes the following phases [9]: (1) data selection, (2) data preprocessing, (3) data transformation, (4) AI/expert systems, and (5) understanding/assessment. The first three phases include subtasks such as: (i) feature selection and extraction, (ii) data cleaning, and (iii) data integration from multiple sources. The included studies covered all phases of the process. Title, abstract, and full-text screening were independently performed by two reviewers using the inclusion criteria. Any disagreements were resolved by a third independent investigator.

Data Extraction

The following data were extracted from the retrieved articles: publication information, journal name and impact factor, study characteristics, big data characteristics, outcomes, lessons and barriers for implementation, and main limitations. Data were individually extracted by team members and cross-checked for accuracy by a second investigator.

Assessment of Methodological Quality of Included Reviews

Two researchers independently assessed the studies using the AMSTAR 2 (A Measurement Tool to Assess Systematic Reviews 2) checklist, which includes the following critical domains, assessed in 16 items: protocol registered prior to review, adequacy of literature search, justification for excluded studies, risk of bias in included studies, appropriateness of meta-analytic methods, consideration of bias risk when interpreting results, and assessing the presence and likely impact of publication bias [10]. Appropriateness to each appraisal feature was rated as yes, no, partial yes, not applicable, or unclear. Any conflict was resolved by a third party. Studies with a review protocol tracking number were analyzed. A final summary score was given to each included record, rated as "critically low," "low," "moderate," or "high" [10].

Data Synthesis

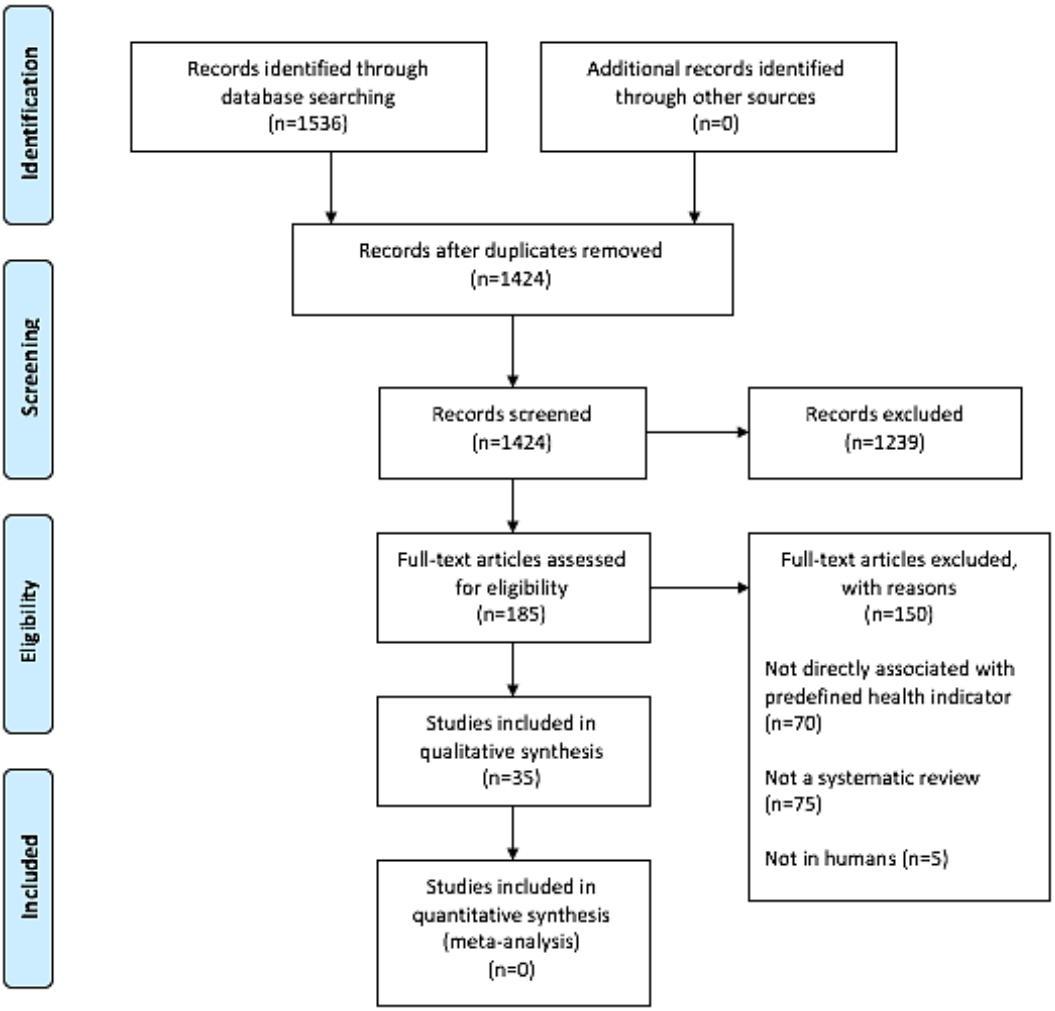
Results are reported in summary tables and through a narrative synthesis, grouping studies assessing the same disease or condition, and identifying challenges and opportunities. We also schematically represent the evidence and gaps from these reviews as an overall synthesis.

Results

Reviews Retrieved

The search retrieved 1536 publications, 112 of which were duplicates. Most of the studies were excluded after title and abstract analysis (n=1237), leaving 185 selected for full-text screening, and 35 [11-45] were ultimately included in the final analysis after applying the eligibility criteria according to the QUOROM guidelines [8] (Figure 1). Reference list screening did not retrieve any additional review. One study under "awaiting classification" could not be retrieved.

Figure 1. Flow chart of the different phases of article retrieval.



Quality of Evidence in Individual Systematic Reviews

Multimedia Appendix 2 shows the detailed results of the quality assessment of the 35 systematic reviews. Overall, most of the reviews (n=25) were rated with "critically low" confidence in the results using the AMSTAR 2 criteria, with 7 rated "low" and 3 rated as "moderate." None of the reviews achieved a "high" rating. Common methodological drawbacks included omission of prospective protocol submission or publication, inappropriate search strategy, lack of independent and dual literature screening and data extraction, absence of explanation for heterogeneity among the studies, unclear or no reasons for study exclusion, and lack of risk of bias assessment.

No standard critical appraisal tools were mentioned. Among the 12 reviews that performed any quality assessment, the Quality Assessment of Diagnostic Accuracy Studies 2 tool was used in four reviews demonstrating an overall low risk of bias [14,16,27,28], whereas other tools assessed the risk of bias in studies not specifically aiming at diagnostic accuracy features. El Idrissi et al [18] used their own quality assessment tool and Luo et al [34] used an adapted version of the Critical Appraisal Skills Programme. Appraisal of the quality of evidence aligned with the Grading of Recommendations Assessment, Development and Evaluation method was reported in only one review [17]. Many reviews did not evaluate bias.

Characteristics of Included Reviews

Summary features and main findings of the 35 systematic reviews are presented in Multimedia Appendix 3 and Multimedia Appendix 4, respectively. The included reviews were published in 34 different journals from 2007 to 2020. Most were published in English in a first-quartile journal with an impact factor ranging from 0.977 to 17.679. They covered over 2501 primary studies, involving at least 5,000,000 individuals. Only three reviews included meta-analyses, and one included a randomized clinical trial; the others were based on cohort studies.

Data Sources and Purposes of Included Studies

Many reviews included data collected from electronic medical records, hospital information systems, or any databank that used individual patient data to create predictive models or evaluate collective patterns [12,13,16-21,24-27,30,33-35,37,38,40,42-45]. Additionally, four reviews included primary studies based on imaging datasets and databanks, assessing different parameters of accuracy [15,29,31,36]. Other reviews focused on genetic databases [28,35], data from assisted reproductive technologies [30], or publicly available data [11,14,22,32]. Four studies lacked precision about the origin of the datasets used in their analysis or did not specifically use patient data in the investigation [23,37,39,41].

The purposes of the reviews varied broadly. Generally, they (1) outlined AI applications in different medical specialties; (2) analyzed features for the detection, prediction, or diagnosis of multiple diseases or conditions; or (3) pinpointed challenges and opportunities.

WHO Indicators and Core Priorities

Most of the studies assessed the effects of big data analytics on noncommunicable diseases [12-15,17,21,22,24,27,31,32,34,36,38,40-44]. Furthermore, three reviews covered mental health, associated with the indicator "suicide mortality rate" [19,25,45]; three studies were related to the indicator "probability of dying from any of cardiovascular, cancer, diabetes, or chronic renal disease" [16,18,20,28,29]; and two studies were related to the indicator "proportion of bloodstream infections due to antimicrobial-resistant organisms" [26,33]. One study described technology use in disaster management and preparedness, covering the "number of persons affected by disasters" indicator [11], and one study was associated with the indicator "maternal mortality ratio" [30]. Overlap made precise classification into WHO health indicators challenging, and four studies could not be categorized because they mainly described challenges or opportunities in big data analytics [23,39] or because they were related to the COVID-19 pandemic [35,37].

Diseases or Conditions Assessed

Diabetes Mellitus

AI tools associated with big data analytics in the care of patients with diabetes mellitus (DM) were assessed in six reviews that included 345 primary studies [15,20,32,38,40]. Three studies reviewed AI in screening and diagnosing type 1 or type 2 DM, providing varied ranges of accuracy, sensitivity, and specificity [20,32,40]. Variables included systolic blood pressure, body mass index, triglyceride levels, and others. Two reviews covered DM control and the clinical management of DM patients [32,40]. One noted that techniques for diabetes self-management varied among the tools evaluated and reported mean values for its robust metrics [18]. The other evaluated the use of data-driven tools for predicting blood glucose dynamics and the impact of ML and data mining [20], describing the input parameters used among data-driven analysis models. However, the authors of these reviews concluded that achieving a methodologically precise predictive model is challenging and must consider multiple parameters.

Various studies assessed the ability of big data analytics to predict individual DM complications such as hypoglycemia, nephropathy, and others [15,32,38]. Supervised ML methods, decision trees, deep neural networks, random forests (RF) learning, and support vector machine (SVM) reportedly had the best outcomes for assessing complications. One review assessed deep learning–based algorithms in screening patients for diabetic retinopathy. Of 11 studies, 8 reported sensitivity and specificity of 80.3% to 100% and 84%% to 99%, respectively; two reported accuracies of 78.7% and 81%; and one reported an area under the receiver operating curve (AUC) of 0.955 [15].

Mental Health

Five reviews reported on AI, data mining, and ML in psychiatry/psychology [12,14,19,25,45], most commonly assessing these techniques in the diagnosis of mental disorders. Two reviews assessed the use of ML algorithms for predicting suicidal behaviors. High levels of risk classification accuracy (typically higher than 90%) were reported in two reviews, either for adult primary care patients or teenagers [19,25]. Although the review authors stated the potential of ML techniques in daily clinical practice, limitations were highlighted, including no external validation and reporting inconsistencies.

The use of ML algorithms for early detection of psychiatric conditions was also reported [12,45]. ML was used to develop prediagnosis algorithms for constructing risk models to signal a patient's predisposition or risk for a psychiatric/psychological health issue, for predicting a diagnosis of newly identified patients, and to differentiate mental conditions with overlapping symptomatology. For studies using structural neuroimaging to classify bipolar diseases and other diagnoses, the accuracy ranged from 52.13% to 100%, whereas studies using serum biomarkers reported an accuracy ranging from 72.5% to 77.5%.

Only one review used social media to generate analyzable data on the prevention, recognition, and support for severe mental illnesses [14]. The study included broad descriptions of ML techniques and data types for detection, diagnosis, prognosis, treatment, support, and resulting public health implications. The authors highlighted the potential for monitoring well-being, and providing an ecologically and cost-efficient evaluation of community mental health through social media and electronic records.

COVID-19

Two reviews reported the application of big data analytics and ML to better understand the current novel coronavirus pandemic [35,37]. One assessed data mining and ML techniques in diagnosing COVID-19 cases. Although the study did not define the best methodology to evaluate and detect potential cases, the authors noted an elevated frequency of decision tree models, naïve Bayes classifiers, and SVM algorithms used during previous pandemics.

Another review focused on SARS-CoV-2 immunization, and proposed that AI could expedite vaccine discovery through studying the virus's capabilities, virulence, and genome using genetic databanks. That study merged discussions of deep learning–based drug screening for predicting the interaction between protein and ligands, and using imaging results linked to AI tools for detecting SARS-CoV-2 infections.

Oncology

Four studies described the utility of ML, computerized clinical decision systems, and deep learning in oncology [24,28,29,31]. Using computerized clinical decision support systems (DSS) significantly improves process outcomes in oncology [24]. A compelling example shows that initial decisions were modified in 31% of cases after consultation of clinical DSS, which consistently resulted in improved patient management. Furthermore, implementing clinical DSS led to an average cost

reduction of US $17,000 for lung cancer patients. A remarkable workload decrease reportedly occurs when these systems are implemented in oncology facilities, leading to improved patient management and adherence to guidelines [24].

One study evaluated ML techniques in a genomic study of head and neck cancers, and found a wide range of accuracy rates (56.7% to 99.4%) based on the use of genomic data in prognostic prediction. Lastly, two studies reported accuracy levels ranging from 68% to 99.6% when using deep learning algorithms in the automatic detection of pulmonary nodules in computerized tomography images.

Cardiovascular and Neurological Conditions

Six studies described the effect of big data analytics in cardiology [13,16,21,42] and neurology [43,44]. One review assessed the use of ML techniques for predicting cardiac arrest [42]. Different variables were used as predictors among individual studies, including electrocardiographic parameters, heart rate variability, echocardiography, and others. Supervised ML techniques were most frequently applied to predict cardiac arrest events, with clear evidence of regression techniques and SVM algorithms. The authors reported a mean AUC of 0.76 for risk score development and efficiency evaluation [42].

Similarly, two studies assessed the use of intelligent systems in diagnosing acute coronary syndrome and heart failure [13,21], demonstrating high accuracy levels using several methods such as SVM, feature selection, and neural networks. These studies also described useful clinical features for creating prediction and diagnostic models, such as patient clinical data, electrocardiogram characteristics, and cardiac biomarkers.

Scores to identify patients at higher risk to develop QT-interval prolongation have been developed, and predictive analytics incorporated into clinical decision support tools have been tested for their ability to alert physicians of individuals who are at risk of or have QT-interval prolongation [16].

Regarding stroke, two systematic reviews evaluated using ML models for predicting outcomes and diagnosing cerebral ischemic events [43,44]. Generally, ML models were most frequently associated with mortality prediction, functional outcomes, neurological deterioration, and quality of life. The diagnosis of ischemic stroke was associated with similar or better comparative accuracy for detecting large vessel occlusion compared with humans, depending on the AI algorithm employed [44]. RF algorithms had 68% sensitivity and over 80% specificity compared with humans. Analyses of convolutional neural network (CNN) algorithms were limited, but systems using CNNs reported performance metrics on average 8% to 10% greater than those of ML employing RF, with up to 85% mean sensitivity for automatic large vessel occlusion detection. However, AI algorithm performance metrics used different standards, precluding objective comparison. Core and perfusion studies from RAPID-computed tomography and magnetic resonance imaging had the highest metrics for AI accuracy, above 80%, with some datasets showing 100% sensitivity to predict favorable perfusion mismatch. The authors noted several errors of AI use in diagnosing stroke [44].

Miscellaneous Conditions

Several studies reported significant improvement in disease diagnosis and event prediction using big data analytics tools, including remarkable enhancement of sepsis prediction using ML techniques [26]. Another review provided moderate evidence that ML models can reach high performance standards in detecting health care–associated infections [33].

One review focused on the diagnostic accuracy of AI systems in analyzing radiographic images for pulmonary tuberculosis, mostly referring to development instead of clinical evaluation [27]. In studies assessing accuracy, the sensitivity ranged from 56% to 97%, specificity ranged from 36% to 95%, and the AUC ranged from 78% to 99%.

One review also assessed multiple sclerosis diagnosis. Among detection methodologies, rule-based and natural language processing methods were deemed to have superior diagnostic performance based of elevated accuracy and positive predictive value [41]. This study indicates that these methods have potential impacts for early recognition of the disease, increasing quality of life, and allowing prompt pharmacological and nonpharmacological intervention.

Asthma exacerbation events and predictive models for early detection were evaluated in one review, which reached a pooled diagnostic ability of 77% (95% CI 73%-80%) [17]. Among the included studies, most models for predicting asthma development had less than 80% accuracy. None of the 42 studies modeled the reincidence of exacerbation events, and overall accuracy performance was considered inadequate. However, the authors encouraged creating models using large datasets to increase prediction accuracy levels. Logistic regression and Cox proportional hazard regression appeared to be the most commonly used methodologies. Gastric tissue disease and the usability of deep learning techniques were evaluated in one study [36]. CNN was the most common model used for gastric problem classification or detection. Additionally, residual neural network and fully convolutional network were considered to be appropriate models for disease generation, classification, and segmentation.

Two reviews analyzed the use of big data analytics and AI in public health [22,30]. One listed the impact of continuous pharmacological exposure of pregnant women, emphasizing that AI could improve popular understanding of drug effects on pregnancy, mainly through: (i) reliable clinical information disclosure, (ii) adequate scientific research design, and (iii) implementation of DSS [30]. Another review assessing the use of big data in disaster preparedness evidenced that most existing methods are qualitative, covering the response phase of the disaster chain of events [11]. The utilized tools included data originating from geographic information systems, social media interfaces, and disaster prediction modeling studies.

Challenges and Opportunities

Two systematic reviews provided narrative evaluations of the challenges of big data analytics in health care [23,39]. Evidence from these two systematic reviews, and those from the other reviews, are summarized in Textbox 2.

Textbox 2. Current challenges to use big data tools for peoples' health, and future perspectives and opportunities.

Current Challenges

1. Data structure: issues with fragmented data and incompatible or heterogeneous data formats

2. Data security: problems with privacy, lack of transparency, integrity, and inherent data structure

3. Data standardization: concerns with limited interoperability, data obtention, mining, and sharing, along with language barriers

4. Inaccuracy: issues with inconsistencies, lack of precision, and data timeliness

5. Limited awareness of big data analytics capabilities among health managers and health care professionals

6. Lack of evidence of big data analytics on the impact on clinical outcomes for peoples' health

7. Lack of skills and training among professionals to collect, process, or extract data

8. Managerial issues: ownership and government dilemma, along with data management, organizational, and financial issues

9. Regulatory, political, and legal concerns

10. Expenses with data storage and transfer

Future Perspectives and Opportunities

1. To improve the decision-making process with real-time analytics

2. To improve patient-centric health care and to enhance personalized medicine

3. To support early detection of diseases and prognostic assessment by predicting epidemics and pandemics, improving disease monitoring, implementing and tracking health behaviors, predicting patients' vulnerabilities

4. To improve data quality, structure, and accessibility by enabling the improvement of rapid acquisition of large volumes and types of data, in a transparent way, and the improvement of data error detection

5. To enable potential health care cost reduction

6. To improve quality of care by improving efficient health outcomes, reducing the waste of resources, increasing productivity and performance, promoting risk reduction, and optimizing process management

7. To provide better forms to manage population health either through early detection of diseases or establishing ways to support health policy makers.

8. To enhance fraud detection

9. To enhance health-threat detection plans by governmental entities

10. To support the creation of new research hypotheses

Discussion

This overview is the first to assess the effects of big data analytics on the prioritized WHO indicators, which offers utility for noncommunicable diseases and the ongoing COVID-19 pandemic. Although the research question focused on the impact of big data analytics on people's health, studies assessing the impact on clinical outcomes are still scarce. Most of the reviews assessed performance values using big data tools and ML techniques, and demonstrated their applications in medical practice. Most of the reviews were associated with the GPW13 indicator "probability of dying from any cardiovascular disease, cancer, diabetes, chronic respiratory disease." This indicator outranks others because of the incidence, prevalence, premature mortality, and economic impact of these diseases [46]. Similarly, many reviews were related to "people requiring interventions against noncommunicable diseases." The included reviews in this study addressed many necessary health-related tasks; however, the quality of evidence was found to be low to moderate, and studies assessing the impact on clinical outcomes are notably scarce.

The low to moderate quality of evidence suggests that big data analytics has moderate to high accuracy for the (1) diagnosis and prediction of complications of DM, (2) diagnosis of mental diseases, (3) prediction of suicidal behaviors, and (4) diagnosis of chronic diseases. Most studies presented performance values, although no study assessed whether big data analytics or ML could improve the early detection of specific diseases.

Clinical research and clinical trials significantly contribute to understanding the patterns and characteristics of diseases, as well as for improving detection of acute or chronic pathologies and to guide the development of novel medical interventions [47]. However, experimental/theoretical investigations, mathematical approaches, and computer-based studies hinge on handling sample size limitations and performing data imputation [48,49]. Computer-driven analysis can easily handle missing data, examine variable mechanisms in complex systems, and employ essential tools for exploratory evaluations using voluminous input data. Big data analytics can execute an operation on/process data within microseconds after generation of the dataset, allowing for real-time follow up [50,51]. These studies and prospective applications could generate innovative knowledge and promote actionable insights; however, adapting, validating, and translating scientific data into practical medical protocols or evaluation studies is necessary.

Many systematic reviews reported simple or inappropriate evaluation measures for the task at hand. The most common metric used to evaluate the performance of a classification predictive model is accuracy, which is calculated as the proportion of correct predictions in the test set divided by the total number of predictions that were made on the test set. This metric is easy to use and to interpret, as a single number summarizes the model capability. However, accuracy values and error rate, which is simply the complement of accuracy, are not adequate for skewed or imbalanced classification tasks (ie, when the distribution of observations in the training dataset across the classes is not equal), because of the bias toward the majority class. When the distribution is slightly skewed, accuracy can still be a useful metric; however, when the distribution is severely skewed, accuracy becomes an unreliable measure of model performance.

For instance, in a binary classification task with a distribution of (95%, 5%) for the classes (eg, healthy vs sick), a "dumb classifier" that simply chooses the class "healthy" for all instances will have 95% of accuracy in this task, although the most important issue in this task would be correctly classifying the "sick" class. Precision (also called the positive predictive value), which captures the fraction of correctly classified instances among the instances predicted for a given class (eg, "sick"); recall or sensitivity, which captures the fraction of instances of a class (eg, "sick") that were correctly classified; and F-measure, the harmonic mean of precision and recall calculated per class of interest, are more robust metrics for several practical situations. The proper choice of an evaluation metric should be carefully determined, as these indices ought to be used by regulatory bodies for screening tests and not for diagnostic reasoning [52]. The most important issue is to choose the appropriate (most robust) performance metric given the particularities of each case.

Another pitfall identified among the included reviews was the lack of reporting the precise experimental protocols used for testing ML algorithms and the specific type of replication performed.

There is no formal tool for assessing quality and risk of bias in big data studies. This is an area that is ripe for development. In Textbox 3, we summarize our recommendations for systematic reviews on the application of big data and ML for people's health based on our experience, the findings of this systematic review, and inspired by Cunha et al [53].

High variability in the results was evident across different ML techniques and approaches among the 35 reviews, even for those assessing the same disease or condition. Indeed, designing big data analysis and ML experiments involves elevated model complexity and commonly requires testing of several modeling algorithms [54]. The diversity of big data tools and ML algorithms requires proper standardization of protocols and comparative approaches. Additionally, the process of tuning the hyperparameters of the algorithms is not uniformly reported. Important characteristics essential for replicability and external validation were not frequently available. Lastly, most of the studies provide little guidance to explain the results. Without knowing *how* and *why* the models achieve their results, applicability and trust of the models in real-world scenarios are severely compromised. Therefore, we urge the testing and assessment of supervised, unsupervised, and semisupervised methodologies, with explanation and interpretation to justify the results. Moreover, we encourage hyperparameter optimization to achieve adjusted improvement of models, enhance model generalizations for untrained data, and avoid overfitting to increase predictive accuracy.

Only two published systematic reviews evaluated the impact of big data analytics on the COVID-19 pandemic. Primary studies on COVID-19 are lacking, which indicates an opportunity to apply big data and ML to this and future epidemics/pandemics [35,37]. As of November 30, 2020, many published protocols were retrieved through a standard search on PROSPERO. The titles of these review protocols showed an intention to evaluate ML tools in diagnosis and prediction, the impact of telemedicine using ML techniques, and the use of AI-based disease surveillance [55].

Although DSS are an important application of big data analytics and may benefit patient care [56-58], only two reviews assessed such systems [16,24]. One focused on predictive analytics for identifying patients at risk of drug-induced QTc interval prolongation, discussing the efficacy of a DSS that has shown evidence of reduced prescriptions for QT interval–prolonging drugs. Similarly, one study exploring the impact of DSS on quality care in oncology showed that implementing these systems might positively impact physician-prescribing behaviors, health care costs, and clinician workload.

This overview of systematic reviews updates the available evidence from multiple primary studies intersecting computer science, engineering, medicine, and public health. We used a comprehensive search strategy (performed by an information specialist) with a predefined published protocol, precise inclusion criteria, rigorous data extraction, and quality assessment of retrieved records. We avoided reporting bias through the dual and blinded examination of systematic reviews and by having one review author standardizing the extracted data.

Textbox 3. Recommendations for systematic reviews on the application of big data and machine learning for people's health.

- *Choose an appropriate evaluation measure for the task and data characteristics, and justify your choice*

Different evaluation measures such as accuracy, area under the receiver operating characteristic curve, precision, recall, and F-measure capture different aspects of the task and are influenced by data characteristics such as skewness (ie, imbalance), sampling bias, etc. Choose your measures wisely and justify your choice based on the aforementioned aspects of the task and the data.

- *Ensure the employment of appropriate experimental protocols/design to guarantee generalization of the results*

Authors should use experimental protocols based on cross-validation or multiple training/validation/test splits of the employed datasets with more than one repetition of the experimental procedure. The objective of this criterion is to analyze whether the study assesses the capacity of generalization of each method compared in the experiments. The use of a single default split of the input dataset with only one training/test split does not fit this requirement. Repetitions are essential to demonstrate the generalization of the investigated methods for multiple training and test sets, and to avoid any suspicion of a "lucky" (single) partition that favors the authors' method.

- *Properly tune, and explicitly report the tuning process and values of the hyperparameters of all compared methods*

The effectiveness of big data solutions and machine-learning methods is highly affected by the choice of the parameters of these methods (ie, parameter tuning). The wrong or improper choice of parameters may make a highly effective method exhibit very poor behavior in a given task. Ideally, the parameters should be chosen for each specific task and dataset using a partition of the training set (ie, validation), which is different from the dataset used to train and to test the model. This procedure is known as cross-validation on the training set or nested cross-validation.

Even if the tuning of all methods is properly executed, this should be explicitly reported in the paper, with the exact values (or range of values) used for each parameter and the best choices used. When the tuning information is missing or absent, it is impossible to determine whether the methods have been implemented appropriately and if they have achieved their maximum potential in a given task. It is also impossible to assess whether the comparison is fair, as some methods may have been used at their maximum capacity and others not.

- *Pay attention to the appropriate statistical tests*

Authors should employ statistical significance tests to contrast the compared strategies in their experimental evaluation. Statistical tests are essential to assess whether the performance of the analyzed methods in the sample (ie, the considered datasets) is likely to reflect, with certain confidence, their actual performance in the whole population. As such, they are key to support any claim of superiority of a particular method over others. Without such tests, the relative performance observed in the sample cannot, by any means, be extrapolated to the population. The choice of the tests should also reflect the characteristics of the data (ie, determining whether the data follow a normal distribution).

- *Make the data and code freely available with proper documentation*

One of the issues that hampers reproducibility of studies, and therefore scientific progress, is the lack of original implementation (with proper documentation) of the methods and techniques, and the unavailability of the original data used to test the methods. Therefore, it is important to make all data, models, code, documentation, and other digital artifacts used in the research available for others to reuse. The artifacts made available must be sufficient to ensure that published results can be accurately reproduced.

- *Report other dimensions of the problem such as model costs (time) and potential for explainability*

Effectiveness of the solutions, as captured by accuracy-oriented measures, is not the only dimension that should be evaluated. Indeed, if the effectiveness of the studied models is similar and sufficient for a given health-related application, other dimensions such as time efficiency (or the costs) to train and deploy (test) the models are essential to evaluate the practical applicability of such solutions. Another dimension that may influence the decision for the practical use of a big data or a machine-learning method in a real practical situation is the ability to understand why the model has produced certain outputs (ie, explainability). Solutions such as those based on neural networks may be highly effective when presented with huge amounts of data, but their training and deployment costs as well as their opaqueness may not make them the best choice for a given health-related application.

However, limitations exist. The inferior quality scores based on the AMSTAR 2 tool might reflect incomplete reporting and lack of adherence to substandardized review methods. There is neither an established bias risk tool specifically for big data or ML studies nor any systematic way of presenting the findings of such studies. Furthermore, most studies provided a narrative description of results, requiring summarization. Nevertheless, all of the reviews were inspected by most authors, and the most relevant data were condensed in the text or in descriptive tables.

Big data analytics provide public health and health care with powerful instruments to gather and analyze large volumes of heterogeneous data. Although research in this field has been growing exponentially in the last decade, the overall quality of evidence is found to be low to moderate. High variability of results was observed across different ML techniques and approaches, even for the same disease or condition. The diversity of big data tools and ML algorithms require proper standardization of protocols and comparative approaches, and the process of tuning the hyperparameters of the algorithms is not uniformly reported. Important characteristics essential for replicability and external validation were not frequently available.

Additionally, the included reviews in this systematic review addressed different health-related tasks; however, studies assessing the impact on clinical outcomes remain scarce. Thus, evidence of applicability in daily medical practice is still needed. Further studies should focus on how big data analytics impact clinical outcomes and on creating proper methodological guidelines for reporting big data/ML studies, as well as using robust performance metrics to assess accuracy.

Acknowledgments

We highly appreciate the efforts provided by our experienced librarian Maria Björklund from Lund University, who kindly prepared the search strategy used in this research. In addition, we thank Anneliese Arno (University College of London and Covidence Platform) for providing guidance in performing this research through Covidence. We also thank Raisa Eda de Resende, Edson Amaro Júnior, and Kaíque Amâncio Alvim for helping the group with data extraction and double-checking the input data.

Authors' Contributions

IJBdN, MM, MG, NAM, and DNO designed the study. HA, IW, and IJBdN performed first- and second-stage screening, and extracted the presented data. MM solved any disagreements. HA, IW, and IBdN carried out the quality assessment. IJBdN, MM, MG, and DNO drafted the manuscript and its final version. DNO and NAM are staff members of the WHO. The authors alone are responsible for the views expressed in this article and they do not necessarily represent the decisions, policy, or views of the WHO.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy used in the research.
[DOCX File , 26 KB - jmir_v23i4e27275_app1.docx]

Multimedia Appendix 2
Quality assessment judgment using the AMSTAR 2 tool.
[DOCX File , 28 KB - jmir_v23i4e27275_app2.docx]

Multimedia Appendix 3
Main characteristics of included studies.
[DOCX File , 38 KB - jmir_v23i4e27275_app3.docx]

Multimedia Appendix 4
Results and limitations of included systematic reviews.
[DOCX File , 53 KB - jmir_v23i4e27275_app4.docx]

References

1. Ristevski B, Chen M. Big data analytics in medicine and healthcare. J Integr Bioinform 2018 May 10;15(3):20170030 [FREE Full text] [doi: 10.1515/jib-2017-0030] [Medline: 29746254]
2. Panahiazar M, Taslimitehrani V, Jadhav A, Pathak J. Empowering personalized medicine with big data and semantic web technology: promises, challenges, and use cases. Proc IEEE Int Conf Big Data 2014 Oct;2014:790-795 [FREE Full text] [doi: 10.1109/BigData.2014.7004307] [Medline: 25705726]
3. 13th General Programme of Work (GPW13) WHO Impact Framework. World Health Organization. 2019. URL: https://www.who.int/about/what-we-do/GPW13_WHO_Impact_Framework_Indicator_Metadata.pdf [accessed 2020-12-05]
4. The European Programme of Work, 2020–2025 – "United Action for Better Health". World Health Organization, regional office for Europe. 2020. URL: https://www.euro.who.int/en/health-topics/health-policy/european-programme-of-work/about-the-european-programme-of-work/european-programme-of-work-20202025-united-action-for-better-health-in-europe [accessed 2020-12-15]
5. Whitlock EP, Lin JS, Chou R, Shekelle P, Robinson KA. Using existing systematic reviews in complex systematic reviews. Ann Intern Med 2008 May 20;148(10):776-782. [doi: 10.7326/0003-4819-148-10-200805200-00010] [Medline: 18490690]
6. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
7. Higgins JPT, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, Cochrane Bias Methods Group, Cochrane Statistical Methods Group. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ 2011 Oct 18;343:d5928 [FREE Full text] [doi: 10.1136/bmj.d5928] [Medline: 22008217]
8. Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Stroup DF. Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. Quality of Reporting of Meta-analyses. Lancet 1999 Nov 27;354(9193):1896-1900. [doi: 10.1016/s0140-6736(99)04149-5] [Medline: 10584742]

9. Baştanlar Y, Ozuysal M. Introduction to machine learning. Methods Mol Biol 2014;1107:105-128. [doi: 10.1007/978-1-62703-748-8_7] [Medline: 24272434]
10. Shea BJ, Reeves BC, Wells G, Thuku M, Hamel C, Moran J, et al. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. BMJ 2017 Sep 21;358:j4008 [FREE Full text] [doi: 10.1136/bmj.j4008] [Medline: 28935701]
11. Freeman JD, Blacker B, Hatt G, Tan S, Ratcliff J, Woolf TB, et al. Use of big data and information and communications technology in disasters: an integrative review. Disaster Med Public Health Prep 2019 Apr;13(2):353-367. [doi: 10.1017/dmp.2018.73] [Medline: 30047353]
12. Librenza-Garcia D, Kotzian BJ, Yang J, Mwangi B, Cao B, Pereira Lima LN, et al. The impact of machine learning techniques in the study of bipolar disorder: A systematic review. Neurosci Biobehav Rev 2017 Sep;80:538-554. [doi: 10.1016/j.neubiorev.2017.07.004] [Medline: 28728937]
13. Sprockel J, Tejeda M, Yate J, Diaztagle J, González E. Intelligent systems tools in the diagnosis of acute coronary syndromes: A systemic review. Arch Cardiol Mex 2018;88(3):178-189. [doi: 10.1016/j.acmx.2017.03.002] [Medline: 28359602]
14. Shatte ABR, Hutchinson DM, Teague SJ. Machine learning in mental health: a scoping review of methods and applications. Psychol Med 2019 Jul;49(9):1426-1448. [doi: 10.1017/S0033291719000151] [Medline: 30744717]
15. Nielsen KB, Lautrup ML, Andersen JKH, Savarimuthu TR, Grauslund J. Deep learning-based algorithms in screening of diabetic retinopathy: a systematic review of diagnostic performance. Ophthalmol Retina 2019 Apr;3(4):294-304. [doi: 10.1016/j.oret.2018.10.014] [Medline: 31014679]
16. Tomaselli Muensterman E, Tisdale JE. Predictive analytics for identification of patients at risk for QT interval prolongation: a systematic review. Pharmacotherapy 2018 Aug;38(8):813-821. [doi: 10.1002/phar.2146] [Medline: 29882591]
17. Bridge J, Blakey JD, Bonnett L. A systematic review of methodology used in the development of prediction models for future asthma exacerbation. BMC Med Res Methodol 2020 Feb 05;20(1):22 [FREE Full text] [doi: 10.1186/s12874-020-0913-7] [Medline: 32024484]
18. El Idrissi T, Idri A, Bakkoury Z. Systematic map and review of predictive techniques in diabetes self-management. Int J Inf Manage 2019 Jun;46:263-277. [doi: 10.1016/j.ijinfomgt.2018.09.011]
19. Bernert RA, Hilberg AM, Melia R, Kim JP, Shah NH, Abnousi F. Artificial intelligence and suicide prevention: a systematic review of machine learning investigations. Int J Environ Res Public Health 2020 Aug 15;17(16):5929 [FREE Full text] [doi: 10.3390/ijerph17165929] [Medline: 32824149]
20. Woldaregay AZ, Årsand E, Walderhaug S, Albers D, Mamykina L, Botsis T, et al. Data-driven modeling and prediction of blood glucose dynamics: Machine learning applications in type 1 diabetes. Artif Intell Med 2019 Jul;98:109-134 [FREE Full text] [doi: 10.1016/j.artmed.2019.07.007] [Medline: 31383477]
21. Tripoliti EE, Papadopoulos TG, Karanasiou GS, Naka KK, Fotiadis DI. Heart failure: diagnosis, severity estimation and prediction of adverse events through machine learning techniques. Comput Struct Biotechnol J 2017;15:26-47 [FREE Full text] [doi: 10.1016/j.csbj.2016.11.001] [Medline: 27942354]
22. Yin Z, Sulieman LM, Malin BA. A systematic literature review of machine learning in online personal health data. J Am Med Inform Assoc 2019 Jun 01;26(6):561-576 [FREE Full text] [doi: 10.1093/jamia/ocz009] [Medline: 30908576]
23. Kruse CS, Goswamy R, Raval Y, Marawi S. Challenges and opportunities of big data in health care: a systematic review. JMIR Med Inform 2016 Nov 21;4(4):e38 [FREE Full text] [doi: 10.2196/medinform.5359] [Medline: 27872036]
24. Klarenbeek SE, Weekenstroo HH, Sedelaar JM, Fütterer JJ, Prokop M, Tummers M. The effect of higher level computerized clinical decision support systems on oncology care: a systematic review. Cancers (Basel) 2020 Apr 22;12(4):1032 [FREE Full text] [doi: 10.3390/cancers12041032] [Medline: 32331449]
25. Burke TA, Ammerman BA, Jacobucci R. The use of machine learning in the study of suicidal and non-suicidal self-injurious thoughts and behaviors: A systematic review. J Affect Disord 2019 Feb 15;245:869-884. [doi: 10.1016/j.jad.2018.11.073] [Medline: 30699872]
26. Fleuren LM, Klausch TLT, Zwager CL, Schoonmade LJ, Guo T, Roggeveen LF, et al. Machine learning for the prediction of sepsis: a systematic review and meta-analysis of diagnostic test accuracy. Intensive Care Med 2020 Mar;46(3):383-400 [FREE Full text] [doi: 10.1007/s00134-019-05872-y] [Medline: 31965266]
27. Harris M, Qi A, Jeagal L, Torabi N, Menzies D, Korobitsyn A, et al. A systematic review of the diagnostic accuracy of artificial intelligence-based computer programs to analyze chest x-rays for pulmonary tuberculosis. PLoS One 2019;14(9):e0221339 [FREE Full text] [doi: 10.1371/journal.pone.0221339] [Medline: 31479448]
28. Patil S, Habib Awan K, Arakeri G, Jayampath Seneviratne C, Muddur N, Malik S, et al. Machine learning and its potential applications to the genomic study of head and neck cancer-A systematic review. J Oral Pathol Med 2019 Oct;48(9):773-779. [doi: 10.1111/jop.12854] [Medline: 30908732]
29. Pehrson L, Nielsen M, Ammitzbøl Lauridsen C. Automatic pulmonary nodule detection applying deep learning or machine learning algorithms to the LIDC-IDRI database: a systematic review. Diagnostics (Basel) 2019 Mar 07;9(1):29 [FREE Full text] [doi: 10.3390/diagnostics9010029] [Medline: 30866425]
30. Davidson L, Boland MR. Enabling pregnant women and their physicians to make informed medication decisions using artificial intelligence. J Pharmacokinet Pharmacodyn 2020 Aug;47(4):305-318 [FREE Full text] [doi: 10.1007/s10928-020-09685-1] [Medline: 32279157]

31. Li D, Mikela Vilmun B, Frederik Carlsen J, Albrecht-Beste E, Ammitzbøl Lauridsen C, Bachmann Nielsen M, et al. The performance of deep learning algorithms on automatic pulmonary nodule detection and classification tested on different datasets that are not derived from LIDC-IDRI: a systematic review. Diagnostics (Basel) 2019 Nov 29;9(4):207 [FREE Full text] [doi: 10.3390/diagnostics9040207] [Medline: 31795409]
32. Chaki J, Thillai Ganesh S, Cidham S, Ananda Theertan S. Machine learning and artificial intelligence based diabetes mellitus detection and self-management: a systematic review. J King Saud Univ Comput Inf Sci 2020 Jul:In press [FREE Full text] [doi: 10.1016/j.jksuci.2020.06.013]
33. Scardoni A, Balzarini F, Signorelli C, Cabitza F, Odone A. Artificial intelligence-based tools to control healthcare associated infections: A systematic review of the literature. J Infect Public Health 2020 Aug;13(8):1061-1077 [FREE Full text] [doi: 10.1016/j.jiph.2020.06.006] [Medline: 32561275]
34. Luo G, Nkoy FL, Stone BL, Schmick D, Johnson MD. A systematic review of predictive models for asthma development in children. BMC Med Inform Decis Mak 2015 Nov 28;15:99 [FREE Full text] [doi: 10.1186/s12911-015-0224-9] [Medline: 26615519]
35. Kannan S, Subbaram K, Ali S, Kannan H. The role of artificial intelligence and machine learning techniques: race for COVID-19 vaccine. Arch Clin Infect Dis 2020 May 10;15(2):e103232. [doi: 10.5812/archcid.103232]
36. Gonçalves WGE, Dos Santos M, Lobato FMF, Ribeiro-Dos-Santos Â, de Araújo GS. Deep learning in gastric tissue diseases: a systematic review. BMJ Open Gastroenterol 2020 Mar 26;7(1):e000371 [FREE Full text] [doi: 10.1136/bmjgast-2019-000371] [Medline: 32337060]
37. Albahri AS, Hamid RA, Alwan JK, Al-Qays ZT, Zaidan AA, Zaidan BB, et al. Role of biological data mining and machine learning techniques in detecting and diagnosing the novel coronavirus (COVID-19): a systematic review. J Med Syst 2020 May 25;44(7):122 [FREE Full text] [doi: 10.1007/s10916-020-01582-x] [Medline: 32451808]
38. Kavakiotis I, Tsave O, Salifoglou A, Maglaveras N, Vlahavas I, Chouvarda I. Machine learning and data mining methods in diabetes research. Comput Struct Biotechnol J 2017;15:104-116 [FREE Full text] [doi: 10.1016/j.csbj.2016.12.005] [Medline: 28138367]
39. Galetsi P, Katsaliaki K, Kumar S. Values, challenges and future directions of big data analytics in healthcare: A systematic review. Soc Sci Med 2019 Nov;241:112533. [doi: 10.1016/j.socscimed.2019.112533] [Medline: 31585681]
40. Abhari S, Niakan Kalhori SR, Ebrahimi M, Hasannejadasl H, Garavand A. Artificial intelligence applications in type 2 diabetes mellitus care: focus on machine learning methods. Healthc Inform Res 2019 Oct;25(4):248-261 [FREE Full text] [doi: 10.4258/hir.2019.25.4.248] [Medline: 31777668]
41. Arani LA, Hosseini A, Asadi F, Masoud SA, Nazemi E. Intelligent computer systems for multiple sclerosis diagnosis: a systematic review of reasoning techniques and methods. Acta Inform Med 2018 Dec;26(4):258-264 [FREE Full text] [doi: 10.5455/aim.2018.26.258-264] [Medline: 30692710]
42. Layeghian Javan S, Sepehri MM, Aghajani H. Toward analyzing and synthesizing previous research in early prediction of cardiac arrest using machine learning based on a multi-layered integrative framework. J Biomed Inform 2018 Dec;88:70-89 [FREE Full text] [doi: 10.1016/j.jbi.2018.10.008] [Medline: 30389440]
43. Wang W, Kiik M, Peek N, Curcin V, Marshall IJ, Rudd AG, et al. A systematic review of machine learning models for predicting outcomes of stroke with structured data. PLoS One 2020;15(6):e0234722 [FREE Full text] [doi: 10.1371/journal.pone.0234722] [Medline: 32530947]
44. Murray NM, Unberath M, Hager GD, Hui FK. Artificial intelligence to diagnose ischemic stroke and identify large vessel occlusions: a systematic review. J Neurointerv Surg 2020 Feb;12(2):156-164. [doi: 10.1136/neurintsurg-2019-015135] [Medline: 31594798]
45. Alonso SG, de la Torre-Díez I, Hamrioui S, López-Coronado M, Barreno DC, Nozaleda LM, et al. Data mining algorithms and techniques in mental health: a systematic review. J Med Syst 2018 Jul 21;42(9):161. [doi: 10.1007/s10916-018-1018-2] [Medline: 30030644]
46. GBD 2016 Disease Injury Incidence Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 328 diseases and injuries for 195 countries, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet 2017 Sep 16;390(10100):1211-1259 [FREE Full text] [doi: 10.1016/S0140-6736(17)32154-2] [Medline: 28919117]
47. Nass SJ, Levit LA, Gostin LO. Rule I of M (US) C on HR and the P of HITHP. In: The value, importance, and oversight of health research. Washington, DC: National Academy of Sciences; 2009:334.
48. Sivarajah U, Kamal MM, Irani Z, Weerakkody V. Critical analysis of Big Data challenges and analytical methods. J Bus Res 2017 Jan;70:263-286. [doi: 10.1016/j.jbusres.2016.08.001]
49. Cassidy R, Singh NS, Schiratti P, Semwanga A, Binyaruka P, Sachingongu N, et al. Mathematical modelling for health systems research: a systematic review of system dynamics and agent-based models. BMC Health Serv Res 2019 Nov 19;19(1):845 [FREE Full text] [doi: 10.1186/s12913-019-4627-7] [Medline: 31739783]
50. Ngiam KY, Khor IW. Big data and machine learning algorithms for health-care delivery. Lancet Oncol 2019 May;20(5):e262-e273. [doi: 10.1016/S1470-2045(19)30149-4] [Medline: 31044724]

51. Lee S, Mohr NM, Street WN, Nadkarni P. Machine learning in relation to emergency medicine clinical and operational scenarios: an overview. West J Emerg Med 2019 Mar;20(2):219-227 [FREE Full text] [doi: 10.5811/westjem.2019.1.41244] [Medline: 30881539]
52. Llewelyn H. Sensitivity and specificity are not appropriate for diagnostic reasoning. BMJ 2017 Sep 06;358:j4071. [doi: 10.1136/bmj.j4071] [Medline: 28877897]
53. Cunha W, Mangaravite V, Gomes C, Canuto S, Resende E, Nascimento C, et al. On the cost-effectiveness of neural and non-neural approaches and representations for text classification: A comprehensive comparative study. Inf Process Manage 2021 May;58(3):102481. [doi: 10.1016/j.ipm.2020.102481]
54. Deo RC. Machine learning in medicine. Circulation 2015 Nov 17;132(20):1920-1930 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.115.001593] [Medline: 26572668]
55. National Institute for Health Research: PROSPERO - International Prospective Register of Systematic Reviews. University of York Centre for Reviews and Dissemination. URL: https://www.crd.york.ac.uk/prospero/ [accessed 2020-12-04]
56. Rumsfeld JS, Joynt KE, Maddox TM. Big data analytics to improve cardiovascular care: promise and challenges. Nat Rev Cardiol 2016 Jun;13(6):350-359. [doi: 10.1038/nrcardio.2016.42] [Medline: 27009423]
57. Tabary MY, Memariani A, Ebadati E. Chapter 3 - Developing a decision support system for big data analysis and cost allocation in national healthcare. In: Dey N, Ashour AS, Bhat C, Fong SJ, editors. Healthcare data analytics and management. Cambridge, MA: Academic Press; 2019:89-109.
58. Dagliati A, Tibollo V, Sacchi L, Malovini A, Limongelli I, Gabetta M, et al. Big data as a driver for clinical decision support systems: a learning health systems perspective. Front Digit Humanit 2018 May 1;5:8. [doi: 10.3389/fdigh.2018.00008]

Abbreviations

AI: artificial intelligence
AUC: area under the receiver operating characteristic curve
AMSTAR 2: A Measurement Tool to Assess Systematic Reviews 2
CNN: convolutional neural network
DM: diabetes mellitus
DSS: decision support system
EPW: European Programme of Work
GPW13: Thirteenth General Programme of Work
ML: machine learning
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-analyses
QUOROM: Quality of Reporting of Meta-analyses
RF: random forest
SVM: support vector machine
WHO: World Health Organization

Edited by R Kukafka, G Eysenbach; submitted 19.01.21; peer-reviewed by Y Mejova, A Benis; comments to author 09.02.21; revised version received 19.02.21; accepted 24.03.21; published 13.04.21.

Please cite as:
Borges do Nascimento IJ, Marcolino MS, Abdulazeem HM, Weerasekara I, Azzopardi-Muscat N, Gonçalves MA, Novillo-Ortiz D Impact of Big Data Analytics on People's Health: Overview of Systematic Reviews and Recommendations for Future Studies J Med Internet Res 2021;23(4):e27275
URL: https://www.jmir.org/2021/4/e27275
doi:10.2196/27275
PMID:33847586

Review

eHealth Interventions to Address Sexual Health, Substance Use, and Mental Health Among Men Who Have Sex With Men: Systematic Review and Synthesis of Process Evaluations

Rebecca Meiksin[1], BA, MPH; G J Melendez-Torres[2], BSc, MPhil, MPH, DPhil; Jane Falconer[1], BSc, MA; T Charles Witzel[1], BA, MSc, PhD; Peter Weatherburn[1], BA, MSc; Chris Bonell[1], BA, MA, MSc, PhD

[1]London School of Hygiene & Tropical Medicine, London, United Kingdom

[2]Peninsula Technology Assessment Group, College of Medicine and Health, University of Exeter, Exeter, United Kingdom

Corresponding Author:
Rebecca Meiksin, BA, MPH
London School of Hygiene & Tropical Medicine
15-17 Tavistock Place
London, WC1H 9SH
United Kingdom
Phone: 44 (0)20 7927 2893
Email: rebecca.meiksin@lshtm.ac.uk

Abstract

Background: Men who have sex with men (MSM) face disproportionate risks concerning HIV and other sexually transmitted infections, substance use, and mental health. These outcomes constitute an interacting syndemic among MSM; interventions addressing all 3 together could have multiplicative effects. eHealth interventions can be accessed privately, and evidence from general populations suggests these can effectively address all 3 health outcomes. However, it is unclear how useable, accessible, or acceptable eHealth interventions are for MSM and what factors affect this.

Objective: We undertook a systematic review of eHealth interventions addressing sexual risk, substance use, and common mental illnesses among MSM and synthesized evidence from process evaluations.

Methods: We searched 19 databases, 3 trials registers, OpenGrey, and Google, and supplemented this by reference checks and requests to experts. Eligible reports were those that discussed eHealth interventions offering ongoing support to MSM aiming to prevent sexual risk, substance use, anxiety or depression; and assessed how intervention delivery or receipt varied with characteristics of interventions, providers, participants, or context. Reviewers screened citations on titles, abstracts, and then full text. Reviewers assessed quality of eligible studies, and extracted data on intervention, study characteristics, and process evaluation findings. The analysis used thematic synthesis.

Results: A total of 12 reports, addressing 10 studies of 8 interventions, were eligible for process synthesis. Most addressed sexual risk alone or with other outcomes. Studies were assessed as medium and high reliability (reflecting the trustworthiness of overall findings) but tended to lack depth and breadth in terms of the process issues explored. Intervention acceptability was enhanced by ease of use; privacy protection; use of diverse media; opportunities for self-reflection and to gain knowledge and skills; and content that was clear, interactive, tailored, reflective of MSM's experiences, and affirming of sexual-minority identity. Technical issues and interventions that were too long detracted from acceptability. Some evidence suggested that acceptability varied by race or ethnicity and educational level; findings on variation by socioeconomic status were mixed. No studies explored how intervention delivery or receipt varied by provider characteristics.

Conclusions: Findings suggest that eHealth interventions targeting sexual risk, substance use, and mental health are acceptable for MSM across sociodemographic groups. We identified the factors shaping MSM's receipt of such interventions, highlighting the importance of tailored content reflecting MSM's experiences and of language affirming sexual-minority identities. Intervention developers can draw on these findings to increase the usability and acceptability of integrated eHealth interventions to address the syndemic of sexual risk, substance use, and mental ill health among MSM. Evaluators of these interventions can draw on our findings to plan evaluations that explore the factors shaping usability and acceptability.

(J Med Internet Res 2021;23(4):e22477) doi:10.2196/22477

KEYWORDS

eHealth; digital health; men who have sex with men; sexual health; HIV; STI; substance use; mental health; systematic review; process evaluation

Introduction

Men who have sex with men (MSM) face disproportionate risks in relation to use of tobacco, alcohol, and legal and illegal drugs (henceforth termed *substance use*), mental ill health, and HIV and other sexually transmitted infections [1-12]. These outcomes constitute a syndemic, whereby they interact [2,13] to increase overall risks of substance use, mental ill health, and sexual risk among MSM across age groups and ethnicities [13-16]. This clustering and interaction of adverse outcomes suggests that interventions which address substance use, mental ill health, and sexual risk together could have multiplicative effects. eHealth interventions, delivered via electronic media and devices, offer a means to access prevention and treatment programs privately and anonymously particularly for MSM, who generally report high use of such media and devices [17]. Systematic reviews for general or mixed populations suggest that eHealth interventions can be effective in reducing alcohol use [18] and addressing common mental health issues [19-25], and emerging evidence indicates potential effects on drug use and sexual risk [26-29]. The few reviews assessing eHealth interventions among MSM suggest they can achieve short-term behavior change for sexually transmitted infections/HIV prevention [28,30], increase HIV testing [28,31], and improve treatment adherence among HIV-positive MSM [31,32]. To our knowledge, no systematic reviews have assessed the effectiveness of eHealth interventions to reduce substance use or improve mental health among MSM.

In addition to their effectiveness, it is important to examine what factors affect the usability and acceptability of eHealth interventions addressing these various outcomes among MSM. This should inform the development of interventions that can feasibly and acceptably address all 3 outcomes together [33]. Designing eHealth interventions to address MSM's needs and affirm sexual-minority identities is likely to be important [34]. Product assessments suggest that eHealth interventions to reduce depression and anxiety among the general population rarely address the needs of gay and lesbian users [34]. However, there have been no systematic reviews to date conducted on the acceptability and usability of eHealth interventions addressing sexual health, substance use, or mental health risks among MSM.

Toward this end, we undertook a systematic review of eHealth interventions addressing these 3 outcomes and targeting this population. We included interventions addressing these outcomes together or separately, and aimed to synthesize evidence of effectiveness, describe intervention theories of change, and synthesize evidence from economic and process evaluations. This paper reports on the synthesis of process evaluations examining what factors related to interventions, providers, participants, or contexts (ie, environmental or structural factors) promote or impede delivery or receipt of these interventions.

Methods

Inclusion Criteria

Reports eligible for inclusion in the overall review reported on eHealth interventions that were delivered by mobile phone apps, the internet, or other electronic communication technology; offered ongoing support to populations consisting entirely or principally of gay, bisexual, and other men (including cisgender and transgender men) who have sex with men; and aimed to prevent HIV/sexually transmitted infections, sexual risk behavior, alcohol, tobacco or drug use, anxiety, or depression. We excluded interventions that offered one-off (rather than ongoing) support or that involved human providers (eg, in a chat room). Reports eligible for the process evaluation synthesis reported on characteristics of interventions, providers, participants, or context affecting delivery or receipt of eligible interventions. We included published and grey literature and set no restrictions by location or language.

Search Strategy and Screening for Eligibility

Terms used in our search strategy covered 2 concepts joined by the Boolean operator "and": MSM and eHealth. We searched 19 databases containing health and social science literature (from October 2018 to November 2018 and updated on April 2020). Our complete search strategy for the original OvidSP MEDLINE database is included in Multimedia Appendix 1 and the search strategies for all databases are available at the London School of Hygiene & Tropical Medicine's Data Repository [35]. We conducted additional searches by searching 3 clinical trials registers, the OpenGrey database, and Google (limited to first 100 results), and by completing reference checks of included reports and requests from experts.

Citations were uploaded to EndNote (Clarivate Analytics), deduplicated, and then uploaded to EPPI-Reviewer (version 4.0, EPPI-Centre) for screening. CB and JF independently screened titles and abstracts in batches of the same 50 references, resolving disagreements by discussion. After reaching an agreement rate of at least 95%, they single-screened all remaining citations. Screening of full texts followed a comparable process.

Data Extraction and Assessment of Quality

For process evaluations, CB and RM used an adapted version of an existing tool [36] to independently extract data reporting empirically on how processes of delivery or receipt varied with characteristics of interventions, providers, participants, or contexts. They also extracted data on basic study details, methods, and intervention descriptions. They then independently assessed the quality of process evaluation reports using standard Critical Appraisal Skills Program and EPPI-Centre tools [37]. These addressed the rigor of sampling, data collection and data analysis; the extent to which study findings were grounded in the data; whether the study privileged the perspectives of participants (eg, by including open-ended questions or otherwise

allowing space for participants to set out their own views); and the breadth and depth of findings (ie, the extent to which the study explored a broad range of process issues or provided in-depth insights into participant perspectives). Drawing on these criteria, each reviewer then assigned weights (high, medium, or low) to rate the reliability or trustworthiness of the findings, and the usefulness of the findings for shedding light on the research question (ie, the extent to which they shed light on how processes of intervention delivery and receipt varied with characteristics of interventions, providers, participants, or contexts). Reliability reflected the trustworthiness of the overall findings (ie, the extent to which the methods employed were rigorous and could minimize bias and error in the findings). CB and RM met to compare their assessments, resolving all differences through discussion.

Data Analysis

Using thematic synthesis methods [38-40], we explored themes concerning the characteristics of interventions, participants, and context acting as potential barriers and facilitators of delivery and receipt, and which themes applied across or only within the domains of sexual health, substance use, and mental health interventions. Synthesis followed a meta-ethnographic approach [41]. We undertook line-by-line coding of reports examining "first-order constructs" (directly quoted qualitative data) and second-order constructs (author interpretations). In the case of findings from quantitative study components, we coded author interpretations, first checking as part of quality assessment whether these aligned with quantitative data presented (ie, the extent to which study findings were grounded in the data, as noted above). Coding developed third-order constructs by drawing connections between these data. We did not exclude studies based on quality assessment, but rather gave less interpretive weight to conclusions that drew only on poorer-quality reports.

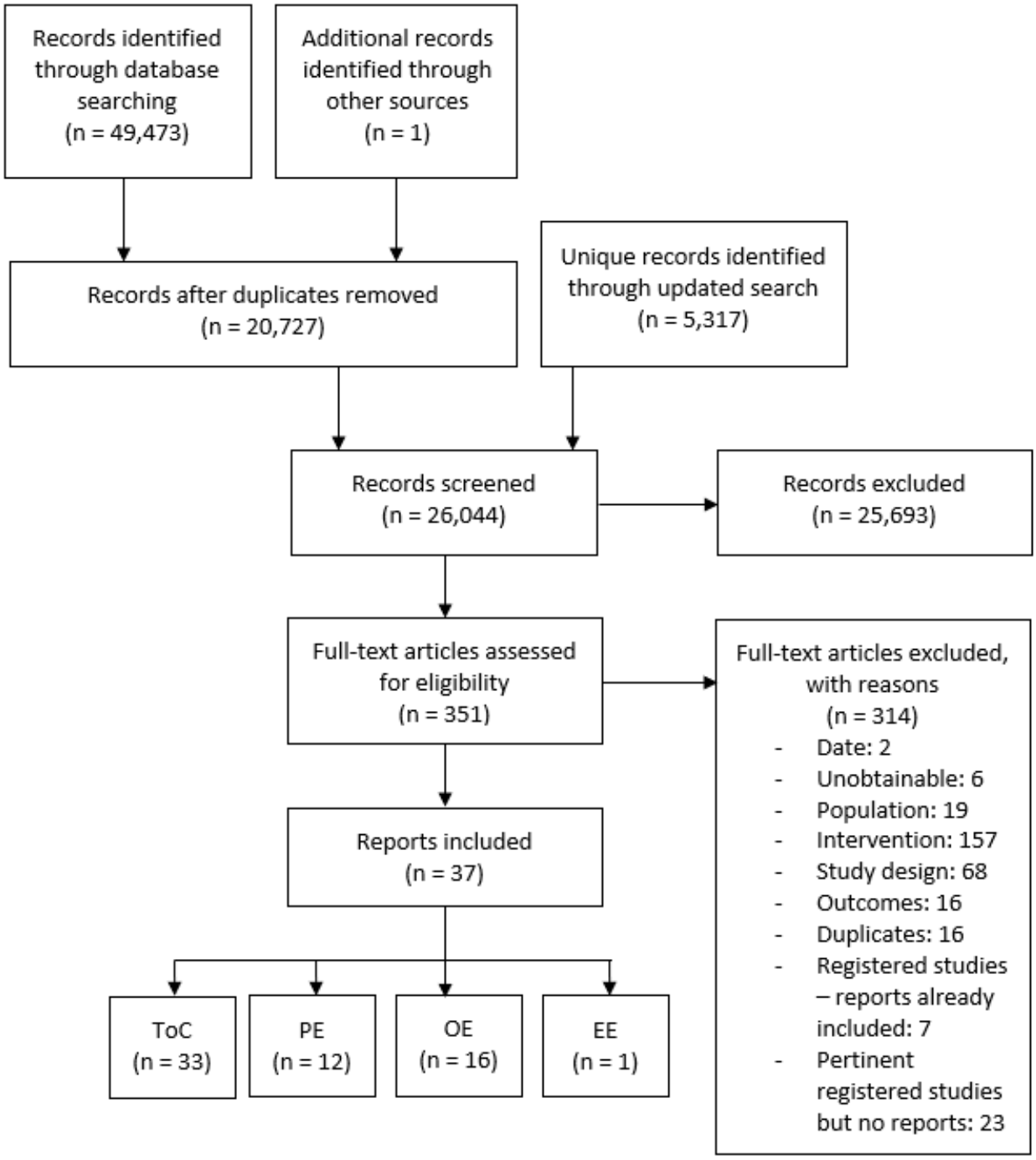
First, CB and RM prepared tables describing the quality of each evaluation, intervention details, study site and population, and pertinent findings. Second, the reviewers independently piloted coding of 2 high-quality studies. Coding began with in vivo codes which closely reflected the words used in the findings sections. The reviewers then grouped and organized codes, applying axial codes that reflected higher-order, cross-cutting themes. They then met to compare and contrast their coding, developing an overall set of codes. Next, they each went on to independently code the remaining reports, drawing on the agreed set of codes and developing new in vivo and axial codes as new themes emerged. At the end of this process, they met to compare their sets of codes. They identified commonalities, differences of emphasis, and contradictions to develop an overall analysis which drew on the strengths of the 2 sets of codes and which resolved any contradictions or inconsistencies.

Results

Results of the Search for Overall Review

Our search retrieved 26,044 unique results, including 1 identified via reference checking (see Figure 1). After title and abstract screening, 6 full texts were unobtainable and 345 reports were screened on full text. Of these, 37 reports were eligible for inclusion in the overall review. These reported on 28 unique studies and 23 unique interventions [42-78]. Reports were published between 2006 and 2020, with most published in 2015 or later.

Figure 1. Searches and screening. EE: economic evaluation; OE: outcome evaluation; PE: process evaluation; ToC: theory of change.



Reports Included in Process Evaluation Synthesis

Twelve reports were eligible for inclusion in the process evaluation synthesis (see Multimedia Appendix 2 for details of each intervention and Multimedia Appendix 3 for study characteristics) [42,49,52,54,59,63,64,67,68,75,76,78]. These reported on 11 studies which assessed 8 unique interventions, and 2 interventions were assessed in 2 different studies [54,59,67,78]. Included process evaluation reports presented findings on how intervention receipt (but not delivery) varied by characteristics of the intervention [42,49,52,59,63,64,67,68,75,76,78] participants [42,49,59,63,64,75,76,78] and context [54,75,78] but not providers. Additionally, 3 interventions addressed sexual health alone [54,68,75,78], 2 addressed mental health alone [42,63,64], 1 addressed sexual health and substance use [49,59,67], and 2 addressed all 3 outcomes of interest for this review [52,76] (see Multimedia Appendix 2). Moreover, 4 interventions targeted sexual minority youth or young adults [49,52,59,63,64,67,68], 2 targeted MSM more generally [42,75], 1 targeted rural MSM [54,78], and 1 targeted people living with HIV [76]. In terms of delivery mode, 5 were delivered via the internet [42,49,52,54,59,67,68,78], 2 via smartphone apps [75,76], and 1 via computer CD-ROM [63,64]. Process evaluations for 7 of the included interventions took place in the United States

[42,49,52,54,59,67,68,75,76,78], and 1 took place in New Zealand [63,64,76].

Quality Assessment

Multimedia Appendix 3 shows the results of our quality assessments. In total, 11 of the 12 included reports were assessed as reporting findings that were grounded in the data presented. Overall quality varied, with most reports assessed as medium or high quality. In terms of the reliability or trustworthiness of their overall findings, 4 reports were assessed as medium quality [42,68,75,76] and 8 as high quality [49,52,54,59,63,64,67,78]. In terms of their overall usefulness for addressing our research questions, 4 were assessed as low quality [52,54,68,78], 3 as medium quality [63,64,67], and 5 as high quality [42,49,59,75,76]. Only 1 report was assessed as high quality in both reliability/trustworthiness and usefulness [49,59], and all were assessed as medium or high quality in at least one of these 2 domains [42,49,52,54,59,63,64,67,68,75,76,78].

Themes Emerging From Synthesis of Process Evaluation Reports

Multimedia Appendix 4 shows the relationship between primary, secondary, and tertiary codes developed through our analysis and synthesis of process data.

Intervention Characteristics Affecting Intervention Receipt

Nearly all process evaluations explored ways in which intervention characteristics affected receipt, although the included reports tended to lack breadth in the areas explored and in-depth exploration of the findings that they did report [42,49,52,59,63,64,67,68,75,76,78]. Nonetheless, several subthemes emerged in our analysis.

Ease of Use

Across health domains, acceptability was enhanced when interventions were easy to use and free of technical problems. Few technical problems were reported. For example, from studies assessed as medium reliability, 10% or fewer Smartphone Self-Monitoring users reported technical difficulties [76] and participants reported that the HealthMindr app was easy to use [75]. However, from studies of medium [42,76] and high [49,54,59,63,67,78] reliability, when participants did encounter technical issues, such as freezing [59] or incompatibility with mobile devices [42,49], this eroded acceptability. In a 2007 study of an intervention targeting rural MSM, features such as sound, animation or graphics could cause the intervention to load too slowly for participants with slower internet speeds, which authors suggested might derail participation [54].

From studies of medium and high reliability, accompanying materials outside of the electronic environment (such as printable materials [42] or a notebook [63,64]) were reported to potentially enhance acceptability, but participants disliked exercises that required using materials they might not have readily at hand [42].

Intervention Content

Clear and Comprehensive Content

From studies of medium reliability across health domains, it was apparent that intervention content which involved clear and comprehensive information facilitated acceptability. For example, Queer Sex Ed participants appreciated that this intervention provided comprehensive information on a range of sexual health and relationship topics rather than focusing narrowly on sexually transmitted infections [68]. In studies of other interventions, acceptability was reportedly enhanced where content was clear and up to date [42], while content participants found confusing detracted from acceptability [76].

Engaging Intervention Content

Fun [68] and enjoyable [42] content increased acceptability, and the use of different types of content arose as a common theme influencing the acceptability. For example, in studies of medium [42] and high [49,59,63,67] reliability, participants tended to give positive feedback on the use of diverse contents [42,49,59] including animations, videos, graphics, and games [67] as well as on the interventions' visual appearances [43,63]. In a high-reliability study of Rainbow SPARX, users were particularly positive about the computer game format and the intervention's "look and feel" [63] as expressed by one user aged 13 years: "I liked, like, how it looked really shiny on my computer, and it looked like a completely different world" [13]. Rainbow SPARX participants also liked particular characters who appeared in the game [63], a theme echoed in a high-reliability study of the Keep it Up! intervention where participants reported liking the scenarios and examples presented [67]. Factors detracting from acceptability included content that participants found boring, repetitive [42,76], too easy [63], too difficult or draining [42], "not soothing" [42], "cheesy" [49,59], or generally unenjoyable [42]; and videos that users judged as too long or that featured low-quality sound and dialogue [49].

Language and Tone

Language and tone emerged as an important aspect of acceptability across interventions addressing all 3 health domains and in studies of medium [42,68,75] and high [49,59,63,64,67] reliability. Participants liked what authors described as a "frank, candid, and sex-positive tone" [59], colloquial language [67], and what one participant described as an "up-beat manner" [67]. For example, Queer Sex Ed users appreciated that the intervention did not rely on "scare tactics" and that its content was easy to understand without making them feel "talked down to" [68]. A Keep it Up! user echoed this sentiment, describing the intervention as "realistic and not condescending or out of touch" [49].

There were also some challenges in getting the language right for MSM-specific interventions. Some users of the Rainbow SPARX and Online Mindfulness-Based Cognitive Therapy interventions suggested that the sexuality-related terminology could be improved [63] and voiced concerns about the intervention's approach to sexual minorities and a feeling of "anti-gay sentiment"[42]. Avellar [42] suggests content might have been overly clinical and miscommunicated the aim of improving overall well-being, although it was not clear whether

participant concerns stemmed primarily from intervention content or from content about participating in a research study.

Interaction and Personalization

Participants in studies of medium [68] and high [49,63,67] reliability valued interactive aspects of interventions spanning all 3 health outcomes. Studies of medium [75,76] and high [59] reliability found that individual-level tailoring based on participant assessments could enhance acceptability. For example, 81% of HealthMindr users found recommendations based on their responses useful or very useful [75] and Smartphone Self-Monitoring users recommended adding what the authors summarized as "more in-depth questions to better reflect their experiences" [76].

Privacy and Intrusiveness

In studies of medium reliability, privacy and intrusiveness emerged as important themes influencing acceptability across 2 interventions, which between them addressed all 3 health outcomes [75,76]. Some Smartphone Self-Monitoring users felt the intervention's use of daily surveys on substance use, sexual behaviors, and medication adherence, and 4-times daily surveys on physical and mental health-related quality of life were too long or too frequent [76]. Users expressed concerns about privacy regarding questions assessing sexual behavior including experiences with individual partners [76]. The vast majority of HealthMindr app users (86%) reported feeling confident in the app's security, including its password features and the fact that the app's name and icon did not suggest it was focused on HIV prevention [75]. At least one participant in the Smartphone Self-Monitoring intervention was uncomfortable with geolocation tagging of phone survey responses, although the authors noted that participants were instructed on how to disable this feature [76].

Pacing and Structuring

The pacing and structuring of content influenced acceptability across health domains. In studies of medium [42] and high [49,64,67] reliability, a modular as opposed to single-session approach could reportedly help users absorb content [67] although they tended to like setting their own pace [64], and one suggested they would have preferred to complete all modules in one sitting [49]. Requiring a full week between sessions of the Online Mindfulness-Based Cognitive Therapy was reported as too long, detracting from acceptability [42].

Users liked when intervention content progressed in a cumulative way [42]. Module order and how far the participant had progressed could also affect acceptability. Findings from a high-reliability study of the 3-module Hope Project (addressing knowledge, motivation, and behavior), which randomized the order in which modules were delivered, suggested that participants were more likely to find the knowledge module interesting when they encountered it last rather than first [78]. Assessing level of interest after each module, this study also found that among those completing all modules, participants were more likely to report finding them very interesting after completing all 3 compared to only the first.

Program length arose as a common theme affecting the acceptability of some modular interventions. Users of the 8-session Online Mindfulness-Based Cognitive Therapy [42], the 7-module Keep it Up! intervention [49,59,67], and 5 five-module Queer Sex Ed intervention [68] suggested that these programs were too long or too time-consuming.

Content Designed To Be Relevant to Participants' Lives and Experiences

Participants valued that interventions were designed for people like them. From studies of high reliability, it was apparent that participants valued when interventions presented realistic scenarios and examples and addressed issues relevant to their own lives [49,59,63,67]. A Keep it Up! user appreciated that the intervention "was geared towards gay men and it understood how we operate and how dating works in the contemporary moment" [49].

Users of the Rainbow SPARX and Queer Sex Ed interventions liked that these programs were "[lesbian, gay, bisexual, and transgender] LGBT–specific" [68], designed for young people [64], and included "'rainbow content" tailored to this group [63]. Some felt there was room to go further [63,68], for example by removing content on female sexual anatomy for MSM users and adding more trans-specific content [68].

Online Mindfulness-Based Cognitive Therapy users had mixed views on how effectively this intervention was tailored for people like them [42]. Some reported appreciating that the program was designed for men who were attracted to men, while others felt the intervention "did not have much value in the context of their lives" [42]. Some Rainbow SPARX users reported that tailoring could be further enhanced by including more sexuality-specific content [63].

Perceived Usefulness of the Intervention

Gaining Knowledge and Skills

In studies of medium [42] and high [49,52,59,63,64,68,78] reliability, participants frequently cited perceived usefulness as positive in terms of the knowledge and skills the interventions aimed to impact. For example, Queer Sex Ed users liked that the intervention aimed to support communication and closeness with their partners, helping, as one participant described, to "open up doors to healthy communication" [68].

Opportunities for Self-Monitoring and Self-Reflection

Findings from the evaluation of the Smartphone Self-Monitoring intervention (targeting sexual health, substance use, and mental health outcomes) suggest that some participants valued its daily, mobile-based self-monitoring in contrast to the comparison group's biweekly, web-based approach. One user described the benefits this way [76]:

> *Helps me keep a "log", like therapy—but can do it every day instead of waiting for a week to see your therapist…Nice to do it throughout the day, multiple times a day, on a daily basis. Life happens daily—not weekly like when you see a therapist.*

Participants in 3 interventions, which between them addressed all 3 health domains, highlighted the opportunities for introspection and self-reflection that the interventions presented

[42,49,59,67,76]. For instance, a Smartphone Self-Monitoring user described the following [76]:

> *I started changing my behavior once I started taking the surveys—I have been thinking about it for a while but the surveys make me concentrate on certain areas of my life that I wasn't focusing on.*

A few also reported that engaging in self-monitoring across multiple domains enhanced their awareness of the relationships between their substance use, sexual behaviors, and other triggers for drug use [76]. A Keep it Up! user described how observing the characters in the intervention helped him to reflect on his own behaviors [67]:

> *I was able to see mistakes that I make in the actions of the characters. I wasn't completely aware of my behavior until I judged a character's behavior and then compared the same behavior to my own.*

Opportunity for Self-Expression

Participants in the Smartphone Self-Monitoring intervention, which addressed all 3 health outcomes, valued the opportunity for self-expression that the intervention offered, as described by this participant: "I feel free to vent to the phone about things that I can't talk to my partner about—I can really express how I feel" [76].

Participant Characteristics Affecting Intervention Engagement and Receipt

Evaluations of 4 interventions (2 targeting sexual health alone [75,78], 1 targeting mental health alone [42], and 1 targeting sexual health and substance use [49]) quantitatively explored the relationship between participant characteristics and intervention engagement.

A medium-reliability study of the HealthMindr mobile phone app found no differences in the time spent on the app by participant location in different cities in the United States, age, ethnicity, or knowledge of local HIV testing [75], while a high-reliability study of the Keep it Up! intervention, targeting young ethnically and racially diverse MSM, found that among Black users, those with graduate degrees spent more time on the intervention than those with high school–level or lower levels of education [49]. A study of medium reliability found no significant variation in retention for a modular mental health intervention by age, socioeconomic status, ethnicity, internalized homonegativity, or experience of homophobic bullying [42]. A study of high reliability found no differences in participants completing 1 versus all 3 modules of the Hope Project (an extension of the WRAPP intervention, targeting rural MSM) by age, ethnicity, marital status, sexual orientation, education, or student status, but did find higher completion among higher-earning participants [78].

Madkins et al [49] conducted a high-reliability, extensive exploration of the relationship between participant characteristics and receipt of the Keep it Up! intervention, which was developed with the engagement of diverse young MSM and designed for young MSM of all racial groups [67]. Researchers found several differences in the acceptability of the Keep it Up! intervention by race/ethnicity, education level, age, and city in the United States [49]. Black, Latino, and other non-White users reported higher acceptability in a range of domains than did White users, and Latino users rated content more highly compared to other non-White users. In the overall sample, users with high school–level education or lower rated the intervention more highly than those with higher education. Exploring the interaction of race/ethnicity and education level, the study found that higher levels of education were associated with lower acceptability among White users, while no such differences were found among Black, Latino, or other non-White users. Older users and those in Atlanta compared to New York tended to rate modules more highly.

Exploring intervention receipt qualitatively, a high-reliability study found that for Rainbow SPARX, a computer game intervention for sexual minority youth aged 13-19 years, some older users reported that some aspects were too easy and the program "babied" them [63]. Acknowledging the challenge of designing a program appropriate for a range of young people, one participant aged 19 years made the following remark [63]:

> *I thought some things were a little easy…Like overall it wasn't difficult to figure out what you needed to do. Those little puzzles were quite easy to do. I guess it would be hard to make them more difficult though because you would have to be careful that everyone could actually get it."*

Qualitative research with participants of Rainbow SPARX and Smartphone Self-Monitoring found that these interventions could play a role in complementing the external mental health support participants were receiving [64,76].

Contextual Factors Affecting Intervention Engagement

Few studies explored how the context for using the intervention was associated with the experience of its use. Those that did focused on internet speed in the high-reliability 2007 [54] and 2010 [78] studies of 2 iterations of the WRAPP sexual health intervention, which targeted rural MSM in the United States. Bowen, et al [54] found that users with dial-up compared to high-speed connections were more likely to report taking too long to load program graphics, while Williams et al [78] found no differences in participants completing 1 versus all 3 modules by type of internet connection.

Discussion

Summary of Findings

One-third of reports included in the overall review included process evaluation data. All but one process evaluation took place in the United States. Most interventions targeted a single health domain of interest for this review (sexual health, substance use, or mental health), with the majority focused on sexual health. However, 2 aimed to address aspects of all 3 [52,76]. Some interventions employed personal tailoring, an approach that has been associated with effective eHealth behavior change interventions [79,80].

Process evaluations rarely explored how intervention receipt varied between contexts. We found no eligible reports examining what factors affected intervention delivery as opposed to receipt.

This seems to reflect the emerging state of process evaluations in eHealth literature, with other reviews of eHealth interventions reporting a similar pattern [81-84]. There was some suggestion that slower internet speed could reduce acceptability of a multimedia intervention among rural MSM in the United States, who are less likely than nonrural residents to have high-speed internet at home [85].

In terms of intervention characteristics, as with use of eHealth interventions among general populations [83], participants appreciated when interventions were easy to use and free of technical problems, while incompatibility with mobile platforms detracted from acceptability and could impede participation. Privacy also emerged as an important aspect of acceptability, suggesting that detailed partner-level questions on sexual behavior could feel intrusive and that features protecting app access and obscuring the purpose of apps (for sensitive health domains) promote acceptability. The importance of privacy is also supported by existing evidence on behavior change interventions for MSM [86] and general populations [83].

Participants liked content that was interactive and aesthetically pleasing, and they enjoyed the use of diverse media such as animations, videos, and graphics. However, among rural MSM these media could also reduce loading times for users with slower internet connectivity. Although modular approaches could support users to absorb program content cumulatively, interventions that were too long detracted from acceptability, with some users preferring that less or no time be required between sessions. The ideal number and length of modules is likely dependent on a variety of participant, intervention, and contextual factors.

Individual tailoring based on participant characteristics and risk profiles increased acceptability, highlighting this as a particularly promising approach and aligning with other studies of eHealth behavioral interventions [79,84,87]. Participants valued when interventions presented scenarios and other content that reflected their experiences as MSM, an approach that stands in contrast to most existing eHealth interventions targeting mental health and HIV prevention [34,88]. Where interventions targeted sexual minority groups more broadly, some suggested further tailoring based on the sexual and gender identities of its users. The language and tone of intervention content emerged as an important factor shaping acceptability for MSM, who appreciated the use of colloquial, direct, “up-beat” [67], and sex-positive language. Our findings also highlight the importance of paying careful attention to language and framing to ensure that these affirm sexual-minority identities. That these concerns arose in interventions designed explicitly for sexual minority users, including one adapted for sexual minority young people using participatory approaches [63], suggests this is an important area to explore during the pilot phase of intervention development.

As with studies of eHealth interventions for general populations [81,83], perceived usefulness was key to acceptability. Participants liked gaining new knowledge and skills from eHealth interventions and developing an awareness of the relationship between sexual behaviors and substance use.

Although reviews of eHealth interventions for general populations report higher use and engagement among participants with higher levels of education [81,83,84], our findings suggest that in the context of generally high use of electronic devices among MSM [17], the targeting of intervention content might be a more important determinant of the relationship between education level and receipt of eHealth interventions than their electronic mode of delivery [49]. Similarly, our findings on the higher acceptability of the Keep it Up! intervention among Black, Latino, and other non-White users compared to White users suggest that eHealth interventions can be developed to enhance inclusive acceptability among racially and ethnically diverse users [49]. There was otherwise little evidence of engagement varying by sociodemographic factors, although findings on socioeconomic status were mixed [42,78]. Qualitative data suggest eHealth interventions can play a role in complementing external mental health support among MSM [63,76] and that interventions targeting all adolescents might struggle to pitch content appropriately for those across this age range [63].

Limitations

Our process evaluation synthesis was limited by the size and quality of eligible reports. Most were assessed as medium- or high-quality in terms of their reliability and usefulness. However, studies often lacked depth and breadth of analysis, and only around half were judged to privilege MSM’s perspectives.

The vast majority of interventions targeted MSM only and all were evaluated principally among MSM, although 3 were assessed among samples that included cisgender women [63,64,68,76]. Author narratives and quantitative data did not always disaggregate MSM from other participants, presenting the possibility that specific findings from these 3 studies might reflect data from other groups. The process evaluation of Smartphone Self-Monitoring was the sole study contributing to findings on intervention benefits of self-monitoring and self-expression [76]. Although the intervention targeted people of all genders and sexual identities living with HIV, more than 80% of study participants identified as male and more than 80% identified as gay or bisexual [76]. In 2 studies, just under half of participants identified as female [63,64,68], but all themes informed by these studies also drew on other studies. The make-up of participants in these 3 studies is therefore unlikely to affect the validity of the themes to which they contributed. Studies of relevant interventions among broader sexual and gender minority populations might add further insight but could not be included, as we could not be certain which findings reflected experiences of or relevant to MSM.

Implications for Research and Practice

eHealth interventions offer an avenue for MSM to access behavior change interventions privately, anonymously, and at times they find convenient. This synthesis identified several factors shaping MSM’s receipt of eHealth interventions addressing substance use, mental ill health, and sexual risk. Its findings suggest such interventions are acceptable for MSM across sociodemographic groups, although evidence in this area is limited and mixed. Different content for younger and older

adolescents might be warranted. Variation in engagement and acceptability by participant characteristics should be explored in future research, and new interventions should be rigorously piloted to refine aspects affecting usability and acceptability [30,81].

Our review has identified several intervention characteristics affecting acceptability that existing research suggests are applicable to eHealth interventions for MSM and non-MSM populations alike. These include aspects of usability, length, aesthetics, multimedia use, and tailoring to participants' personal and risk characteristics [79,81,83,84,86,87]. Other factors should be considered carefully in designing interventions for MSM, including ensuring that language and tone are affirming of sexual minority identity and that content reflects the reality and experiences of MSM. These findings can inform the development of integrated eHealth interventions to address the syndemic of substance use, mental ill health, and sexual risk among MSM and guide research questions for pilot and process evaluation studies. Going forward, process evaluations should explore a broader range of individual, intervention, and contextual factors that might affect implementation, and they should collect more in-depth—ideally qualitative—data privileging the perspectives of intended beneficiaries. Outcome evaluations of such eHealth interventions should conduct linked process evaluations wherever possible, which would shed further light on factors affecting how they are delivered and received [89].

Our findings regarding the value that participants place on interventions that address the reality of their lives and the interrelationships between the different domains of health suggest that eHealth interventions simultaneously addressing sexual health, substance use, and mental health might be particularly acceptable. Our review of theories of change [90] suggests that interventions addressing these different outcomes may aim to exert impacts via common mechanisms of action, further adding to the potential for eHealth interventions targeting multiple outcomes together. Our next analyses will assess the potential effectiveness of eHealth interventions on these outcomes.

Acknowledgments

This work was funded by the National Institute for Health Research Public Health Research Programme (NIHR PHR; PHR 17/44/48). This report presents independent research commissioned by NIHR PHR. The views and opinions expressed by authors in this publication are those of the authors and do not necessarily reflect those of the National Health Service, the NIHR, the Medical Research Council, the Central Commissioning Facility, The NIHR Evaluation, Trials and Studies Coordinating Centre, the PHR Programme, or the Department of Health. The funder had no role in study design, in the collection, analysis or interpretation of data, or in writing this manuscript.

Authors' Contributions

CB conceptualized and led the design of the study. AM, GJMT, JF, PW, RM, and TCM contributed to the development of the study's methods. RM and CB developed the intervention typology and quality-assessed and synthesized process evaluation studies. RM led the drafting of the manuscript, with significant input from AM, CB, GJMT, JF, PW, and TCW.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search terms and strategy for the MEDLINE database.
[DOCX File , 22 KB - jmir_v23i4e22477_app1.docx]

Multimedia Appendix 2
Descriptions of interventions included in process synthesis.
[DOCX File , 44 KB - jmir_v23i4e22477_app2.docx]

Multimedia Appendix 3
Characteristics and quality of appraisal of process evaluations.
[DOCX File , 91 KB - jmir_v23i4e22477_app3.docx]

Multimedia Appendix 4
Coding structure for process evaluation synthesis.
[DOCX File , 22 KB - jmir_v23i4e22477_app4.docx]

References

1. Knight DA, Jarrett D. Preventive health care for men who have sex with men. Am Fam Physician 2015 Jun 15;91(12):844-851 [FREE Full text] [Medline: 26131944]
2. Stall R, Friedman M, Catania J. Interacting epidemics and gay men's health: a theory of syndemic production among urban gay men. In: Wolitski R, Stall R, Valdiserri R, editors. Unequal Opportunity: Health Disparities Affecting Gay and Bisexual Men in the United States. New York, NY: Oxford University Press; 2008:251-274.
3. Beyrer C, Baral SD, van Griensven F, Goodreau SM, Chariyalertsak S, Wirtz AL, et al. Global epidemiology of HIV infection in men who have sex with men. Lancet 2012 Jul 28;380(9839):367-377 [FREE Full text] [doi: 10.1016/S0140-6736(12)60821-6] [Medline: 22819660]
4. Aghaizu A, Brown AE, Nardone A, Gill ON, Delpech VC, and contributors. HIV in the United Kingdom 2013 report. Public Health England. London; 2013. URL: https://www.unaids.org/sites/default/files/country/documents/GBR_narrative_report_2014.pdf [accessed 2021-04-09]
5. Guasp A. Gay and bisexual men's health survey (2013). Stonewall. London: Stonewall; 2015. URL: https://www.stonewall.org.uk/resources/gay-and-bisexual-men%E2%80%99s-health-survey-2013 [accessed 2021-04-09]
6. Buffin J, Roy A, Williams H, Winter A. Part of the picture: lesbian, gay and bisexual people's alcohol and drug use in England (2009-2011). The Lesbian & Gay Foundation; University of Central Lancashire. England: The Lesbian & Gay Foundation; University of Central Lancashire; 2012. URL: https://s3-eu-west-1.amazonaws.com/lgbt-website-media/Files/87413329-79fc-4390-853e-a27f3ef5cf5c/Part%2520of%2520the%2520Picture%25207.pdf [accessed 2021-04-09]
7. Lee J, Griffin GK, Melvin C. Tobacco use among sexual minorities in the USA, 1987 to May 2007: a systematic review. Tob Control 2009 Aug;18(4):275-282. [doi: 10.1136/tc.2008.028241] [Medline: 19208668]
8. Vosburgh HW, Mansergh G, Sullivan PS, Purcell DW. A review of the literature on event-level substance use and sexual risk behavior among men who have sex with men. AIDS Behav 2012 Aug;16(6):1394-1410. [doi: 10.1007/s10461-011-0131-8] [Medline: 22323004]
9. Hickson F, Bonell C, Weatherburn P, Reid D. Illicit drug use among men who have sex with men in England and Wales. Addiction Research & Theory 2010 Jan 08;18(1):14-22. [doi: 10.3109/16066350902770433]
10. Glass R, Hope VD, Tanner C, Desai M. 'Slamming' among men who have sex with men accessing general drug services, in response to Schmidt, AJ et al., 2016, Illicit drug use among gay and bisexual men in 44 cities: Findings from the European MSM Internet Survey (EMIS). Int J Drug Policy 2017 Nov;49:24-25. [doi: 10.1016/j.drugpo.2017.07.028] [Medline: 28873358]
11. King M, Semlyen J, Tai SS, Killaspy H, Osborn D, Popelyuk D, et al. A systematic review of mental disorder, suicide, and deliberate self harm in lesbian, gay and bisexual people. BMC Psychiatry 2008 Aug 18;8:70 [FREE Full text] [doi: 10.1186/1471-244X-8-70] [Medline: 18706118]
12. McFall S. Understanding Society: Findings 2012. In: Institute for Social and Economic Research, University of Essex. Colchester: Institute for Social and Economic Research, University of Essex; 2012.
13. Friedman M, Stall R, Silvestre AJ, Wei C, Shoptaw S, Herrick A, et al. Effects of syndemics on HIV viral load and medication adherence in the multicentre AIDS cohort study. AIDS 2015 Jun 01;29(9):1087-1096 [FREE Full text] [doi: 10.1097/QAD.0000000000000657] [Medline: 25870981]
14. Halkitis PN, Kapadia F, Bub KL, Barton S, Moreira AD, Stults CB. A longitudinal investigation of syndemic conditions among young gay, bisexual, and other MSM: the P18 Cohort Study. AIDS Behav 2015 Jun;19(6):970-980 [FREE Full text] [doi: 10.1007/s10461-014-0892-y] [Medline: 25192900]
15. Halkitis PN, Kupprat SA, Hampton MB, Perez-Figueroa R, Kingdon M, Eddy JA, et al. Evidence for a syndemic in aging HIV-positive gay, bisexual, and other MSM: implications for a holistic approach to prevention and healthcare. Nat Resour Model 2012 Nov 01;36(2):365-386 [FREE Full text] [doi: 10.1111/napa.12009] [Medline: 24347817]
16. Jie W, Ciyong L, Xueqing D, Hui W, Lingyao H. A syndemic of psychosocial problems places the MSM (men who have sex with men) population at greater risk of HIV infection. PLoS One 2012;7(3):e32312 [FREE Full text] [doi: 10.1371/journal.pone.0032312] [Medline: 22479319]
17. Melendez-Torres GJ, Nye E, Bonell C. Internet sex-seeking is inconsistently linked with sexual risk in men who have sex with men: systematic review of within-subjects comparisons. Sex. Health 2015;12(3):183. [doi: 10.1071/sh14183]
18. Riper H, Blankers M, Hadiwijaya H, Cunningham J, Clarke S, Wiers R, et al. Effectiveness of guided and unguided low-intensity internet interventions for adult alcohol misuse: a meta-analysis. PLoS One 2014;9(6):e99912 [FREE Full text] [doi: 10.1371/journal.pone.0099912] [Medline: 24937483]
19. Andersson G, Cuijpers P. Internet-based and other computerized psychological treatments for adult depression: a meta-analysis. Cogn Behav Ther 2009;38(4):196-205. [doi: 10.1080/16506070903318960] [Medline: 20183695]
20. Andrews G, Cuijpers P, Craske MG, McEvoy P, Titov N. Computer Therapy for the Anxiety and Depressive Disorders Is Effective, Acceptable and Practical Health Care: A Meta-Analysis. PLoS ONE 2010 Oct 13;5(10):e13196. [doi: 10.1371/journal.pone.0013196]
21. Arnberg FK, Linton SJ, Hultcrantz M, Heintz E, Jonsson U. Internet-delivered psychological treatments for mood and anxiety disorders: a systematic review of their efficacy, safety, and cost-effectiveness. PLoS One 2014;9(5):e98118 [FREE Full text] [doi: 10.1371/journal.pone.0098118] [Medline: 24844847]

22. Păsărelu CR, Andersson G, Bergman Nordgren L, Dobrean A. Internet-delivered transdiagnostic and tailored cognitive behavioral therapy for anxiety and depression: a systematic review and meta-analysis of randomized controlled trials. Cognitive Behaviour Therapy 2016 Oct 07;46(1):1-28. [doi: 10.1080/16506073.2016.1231219]
23. Spijkerman M, Pots W, Bohlmeijer E. Effectiveness of online mindfulness-based interventions in improving mental health: A review and meta-analysis of randomised controlled trials. Clin Psychol Rev 2016 Apr;45:102-114 [FREE Full text] [doi: 10.1016/j.cpr.2016.03.009] [Medline: 27111302]
24. Christensen H, Batterham P, Calear A. Online interventions for anxiety disorders. Curr Opin Psychiatry 2014 Jan;27(1):7-13. [doi: 10.1097/YCO.0000000000000019] [Medline: 24257123]
25. Kaltenthaler E, Parry G, Beverley C, Ferriter M. Computerised cognitive-behavioural therapy for depression: systematic review. Br J Psychiatry 2008 Sep;193(3):181-184. [doi: 10.1192/bjp.bp.106.025981] [Medline: 18757972]
26. Gabarron E, Wynn R. Use of social media for sexual health promotion: a scoping review. Glob Health Action 2016;9:32193 [FREE Full text] [doi: 10.3402/gha.v9.32193] [Medline: 27649758]
27. L'Engle KL, Mangone ER, Parcesepe AM, Agarwal S, Ippoliti NB. Mobile phone interventions for adolescent sexual and reproductive health: a systematic review. Pediatrics 2016 Sep;138(3):e20160884-e20160884 [FREE Full text] [doi: 10.1542/peds.2016-0884] [Medline: 27553221]
28. Schnall R, Travers J, Rojas M, Carballo-Diéguez A. eHealth interventions for HIV prevention in high-risk men who have sex with men: a systematic review. J Med Internet Res 2014 May 26;16(5):e134 [FREE Full text] [doi: 10.2196/jmir.3393] [Medline: 24862459]
29. Noar S, Black H, Pierce L. Efficacy of computer technology-based HIV prevention interventions: a meta-analysis. AIDS 2009 Jan 02;23(1):107-115. [doi: 10.1097/QAD.0b013e32831c5500] [Medline: 19050392]
30. Nguyen LH, Tran BX, Rocha LEC, Nguyen HLT, Yang C, Latkin CA, et al. A systematic review of eHealth interventions addressing HIV/STI prevention among men who have sex with men. AIDS Behav 2019 Sep;23(9):2253-2272 [FREE Full text] [doi: 10.1007/s10461-019-02626-1] [Medline: 31401741]
31. Purnomo J, Coote K, Mao L, Fan L, Gold J, Ahmad R, et al. Using eHealth to engage and retain priority populations in the HIV treatment and care cascade in the Asia-Pacific region: a systematic review of literature. BMC Infect Dis 2018 Feb 17;18(1):82 [FREE Full text] [doi: 10.1186/s12879-018-2972-5] [Medline: 29454322]
32. Muessig K, LeGrand S, Horvath K, Bauermeister J, Hightow-Weidman L. Recent mHealth interventions to support medication adherence among HIV-positive men who have sex with men. Curr Opin HIV AIDS 2017;12(5):432-441. [doi: 10.1097/coh.0000000000000401]
33. Glasgow RE, Vogt TM, Boles SM. Evaluating the public health impact of health promotion interventions: the RE-AIM framework. Am J Public Health 1999 Sep;89(9):1322-1327. [doi: 10.2105/ajph.89.9.1322] [Medline: 10474547]
34. Rozbroj T, Lyons A, Pitts M, Mitchell A, Christensen H. Assessing the applicability of e-therapies for depression, anxiety, and other mood disorders among lesbians and gay men: analysis of 24 web- and mobile phone-based self-help interventions. J Med Internet Res 2014 Jul 03;16(7):e166 [FREE Full text] [doi: 10.2196/jmir.3529] [Medline: 24996000]
35. Falconer J. Search strategies for: how can e-health interventions reduce the 'syndemic' of HIV/STIs and sexual risk, substance use and mental ill health among men who have sex with men? [Data Collection]. London School of Hygiene & Tropical Medicine. London, United Kingdom; 2020. URL: https://datacompass.lshtm.ac.uk/id/eprint/1751/ [accessed 2021-04-01]
36. Egan M, Bambra C, Petticrew M, Whitehead M. Reviewing evidence on complex social interventions: appraising implementation in systematic reviews of the health effects of organisational-level workplace interventions. J Epidemiol Community Health 2009 Jan;63(1):4-11 [FREE Full text] [doi: 10.1136/jech.2007.071233] [Medline: 18718981]
37. Shepherd J, Harden A, Rees R, Brunton G, Garcia J, Oliver S, et al. Young people and healthy eating: a systematic review of research on barriers and facilitators. Health Educ Res 2006 Apr;21(2):239-257. [doi: 10.1093/her/cyh060] [Medline: 16251223]
38. Arai L. It might work in Oklahoma but will it work in Oakhampton? Context and implementation in the effectiveness literature on domestic smoke detectors. Injury Prevention 2005 Jun 01;11(3):148-151. [doi: 10.1136/ip.2004.007336]
39. Noyes J, Popay J, Garner P. What can qualitative research contribute to a Cochrane systematic review of DOT for promoting adherence to tuberculosis treatment? 2005 Presented at: Qualitative Research and Systematic Reviews workshop; 2005 June 28-29; Oxford, UK.
40. Thomas J, Harden A. Methods for the thematic synthesis of qualitative research in systematic reviews. BMC Med Res Methodol 2008 Jul 10;8:45 [FREE Full text] [doi: 10.1186/1471-2288-8-45] [Medline: 18616818]
41. Green J, Torogood N. Qualitative Methods for Health Research (4th edition). London: Sage; 2018.
42. Avellar T. The feasibility and acceptability of an online mindfulness-based cognitive therapy intervention for same-sex attracted men [Master's thesis]. University of California Santa Barbara. 2016. URL: https://www.alexandria.ucsb.edu/lib/ark:/48907/f3ng4qs1 [accessed 2021-03-31]
43. Cheng W, Xu H, Tang W, Zhong F, Meng G, Han Z, et al. Online HIV prevention intervention on condomless sex among men who have sex with men: a web-based randomized controlled trial. BMC Infect Dis 2019 Jul 19;19(1):644 [FREE Full text] [doi: 10.1186/s12879-019-4251-5] [Medline: 31324233]

44. Chiou P, Liao P, Liu C, Hsu Y. Effects of mobile health on HIV risk reduction for men who have sex with men. AIDS Care 2020 Mar;32(3):316-324. [doi: 10.1080/09540121.2019.1668531] [Medline: 31558040]
45. Coulter RW, Sang JM, Louth-Marquez W, Henderson ER, Espelage D, Hunter SC, et al. Pilot testing the feasibility of a game intervention aimed at improving help seeking and coping among sexual and gender minority youth: protocol for a randomized controlled trial. JMIR Res Protoc 2019 Feb 15;8(2):e12164 [FREE Full text] [doi: 10.2196/12164] [Medline: 30767903]
46. Hirshfield S, Downing MJ, Chiasson MA, Yoon IS, Houang ST, Teran RA, et al. Evaluation of Sex Positive! A video eHealth intervention for men living with HIV. AIDS Behav 2019 Nov;23(11):3103-3118 [FREE Full text] [doi: 10.1007/s10461-019-02498-5] [Medline: 31011912]
47. Kuhns LM, Garofalo R, Hidalgo M, Hirshfield S, Pearson C, Bruce J, et al. A randomized controlled efficacy trial of an mHealth HIV prevention intervention for sexual minority young men: MyPEEPS mobile study protocol. BMC Public Health 2020 Jan 15;20(1):65 [FREE Full text] [doi: 10.1186/s12889-020-8180-4] [Medline: 31941475]
48. Jones J, Dominguez K, Stephenson R, Stekler JD, Castel AD, Mena LA, et al. A theoretically based mobile app to increase pre-exposure prophylaxis uptake among men who have sex with men: protocol for a randomized controlled trial. JMIR Res Protoc 2020 Feb 21;9(2):e16231 [FREE Full text] [doi: 10.2196/16231] [Medline: 32130178]
49. Madkins K, Moskowitz D, Moran K, Dellucci T, Mustanski B. Measuring aacceptability and engagement of the Keep It Up! internet-based HIV prevention randomized controlled trial for young men who have sex with men. AIDS Educ Prev 2019 Aug;31(4):287-305 [FREE Full text] [doi: 10.1521/aeap.2019.31.4.287] [Medline: 31361516]
50. Reback CJ, Fletcher JB, Leibowitz AA. Cost effectiveness of text messages to reduce methamphetamine use and HIV sexual risk behaviors among men who have sex with men. J Subst Abuse Treat 2019 May;100:59-63 [FREE Full text] [doi: 10.1016/j.jsat.2019.02.006] [Medline: 30898329]
51. Tan RKJ, Koh WL, Le D, Tan A, Tyler A, Tan C, et al. Effect of a web drama video series on HIV and other sexually transmitted infection testing among gay, bisexual and queer men: study protocol for a community-based, pragmatic randomised controlled trial in Singapore: the People Like Us (PLU) Evaluation Study. BMJ Open 2020 Apr 07;10(4):e033855 [FREE Full text] [doi: 10.1136/bmjopen-2019-033855] [Medline: 32269026]
52. Bauermeister JA, Tingler RC, Demers M, Connochie D, Gillard G, Shaver J, et al. Acceptability and preliminary efficacy of an online HIV prevention intervention for single young men who have sex with men seeking partners online: the mydex project. AIDS Behav 2019 Nov;23(11):3064-3077 [FREE Full text] [doi: 10.1007/s10461-019-02426-7] [Medline: 30762190]
53. Bauermeister JA, Tingler RC, Demers M, Harper GW. Development of a tailored HIV prevention intervention for single young men who have sex with men who meet partners online: protocol for the mydex project. JMIR Res Protoc 2017 Jul 19;6(7):e141 [FREE Full text] [doi: 10.2196/resprot.7965] [Medline: 28724513]
54. Bowen AM, Horvath K, Williams ML. A randomized control trial of Internet-delivered HIV prevention targeting rural MSM. Health Educ Res 2007 Feb;22(1):120-127 [FREE Full text] [doi: 10.1093/her/cyl057] [Medline: 16849391]
55. Bowen AM, Williams ML, Daniel CM, Clayton S. Internet based HIV prevention research targeting rural MSM: feasibility, acceptability, and preliminary efficacy. J Behav Med 2008 Dec;31(6):463-477 [FREE Full text] [doi: 10.1007/s10865-008-9171-6] [Medline: 18770021]
56. Carpenter KM, Stoner SA, Mikko AN, Dhanak LP, Parsons JT. Efficacy of a web-based intervention to reduce sexual risk in men who have sex with men. AIDS Behav 2010 Jun;14(3):549-557 [FREE Full text] [doi: 10.1007/s10461-009-9578-2] [Medline: 19499321]
57. Christensen JL, Miller LC, Appleby PR, Corsbie-Massay C, Godoy CG, Marsella SC, et al. Reducing shame in a game that predicts HIV risk reduction for young adult men who have sex with men: a randomized trial delivered nationally over the web. Journal of the International AIDS Society 2013 Nov 13;16:18716. [doi: 10.7448/ias.16.3.18716]
58. Davidovich U, de Wit J, Stroebe W. Using the internet to reduce risk of HIV-infection in steady relationships: a randomized controlled trial of a tailored intervention for gay men. Utrecht University Repository. 2006. URL: https://dspace.library.uu.nl/bitstream/1874/9298/7/c6.pdf [accessed 2021-04-12]
59. Greene GJ, Madkins K, Andrews K, Dispenza J, Mustanski B. Implementation and evaluation of the Keep It Up! online HIV prevention intervention in a community-based setting. AIDS Educ Prev 2016 Jun;28(3):231-245. [doi: 10.1521/aeap.2016.28.3.231] [Medline: 27244191]
60. Hirshfield S, Downing MJ, Parsons JT, Grov C, Gordon RJ, Houang ST, et al. Developing a video-based eHealth intervention for HIV-positive gay, bisexual, and other men who have sex with men: study protocol for a randomized controlled trial. JMIR Res Protoc 2016 Jun 17;5(2):e125 [FREE Full text] [doi: 10.2196/resprot.5554] [Medline: 27315764]
61. Kok G, Harterink P, Vriens P, Zwart O, Hospers H. The gay cruise: developing a theory- and evidence-based Internet HIV-prevention intervention. Sex Res Soc Policy 2006 Jun;3(2):52-67. [doi: 10.1525/srsp.2006.3.2.52]
62. Linnemayr S, MacCarthy S, Kim A, Giguere R, Carballo-Dieguez A, Barreras JL. Behavioral economics-based incentives supported by mobile technology on HIV knowledge and testing frequency among Latino/a men who have sex with men and transgender women: protocol for a randomized pilot study to test intervention feasibility and acceptability. Trials 2018 Oct 05;19(1):540 [FREE Full text] [doi: 10.1186/s13063-018-2867-1] [Medline: 30290851]

63. Lucassen MF, Hatcher S, Fleming TM, Stasiak K, Shepherd MJ, Merry SN. A qualitative study of sexual minority young people's experiences of computerised therapy for depression. Australas Psychiatry 2015 Jun;23(3):268-273. [doi: 10.1177/1039856215579542] [Medline: 25881962]
64. Lucassen MF, Merry SN, Hatcher S, Frampton CM. Rainbow SPARX: a novel approach to addressing depression in sexual minority youth. Cognitive and Behavioral Practice 2015 May;22(2):203-216. [doi: 10.1016/j.cbpra.2013.12.008]
65. Milam J, Morris S, Jain S, Sun X, Dubé MP, Daar ES, CCTG 592 Team. Controlled trial of an internet-based risk reduction intervention in HIV+ men who have sex with men. 2014 Presented at: The 2014 Conference on Retroviruses and Opportunistic Infections; 2014 Mar 3-6; Boston, MA, USA.
66. Milam J, Morris S, Jain S, Sun X, Dubé MP, Daar ES, CCTG 592 Team. Randomized controlled trial of an internet application to reduce HIV transmission behavior among HIV infected men who have sex with men. AIDS Behav 2016 Jun;20(6):1173-1181 [FREE Full text] [doi: 10.1007/s10461-015-1215-7] [Medline: 26487300]
67. Mustanski B, Garofalo R, Monahan C, Gratzer B, Andrews R. Feasibility, acceptability, and preliminary efficacy of an online HIV prevention program for diverse young men who have sex with men: the keep it up! intervention. AIDS Behav 2013 Nov;17(9):2999-3012 [FREE Full text] [doi: 10.1007/s10461-013-0507-z] [Medline: 23673793]
68. Mustanski B, Greene GJ, Ryan D, Whitton SW. Feasibility, acceptability, and initial efficacy of an online sexual health promotion program for LGBT youth: the Queer Sex Ed intervention. J Sex Res 2015;52(2):220-230. [doi: 10.1080/00224499.2013.867924] [Medline: 24588408]
69. Mustanski B, Madkins K, Greene GJ, Parsons JT, Johnson BA, Sullivan P, et al. Internet-based HIV prevention with at-home sexually transmitted infection testing for young men having sex with men: study protocol of a randomized controlled trial of Keep It Up! 2.0. JMIR Res Protoc 2017 Jan 07;6(1):e1 [FREE Full text] [doi: 10.2196/resprot.5740] [Medline: 28062389]
70. Mustanski B, Parsons JT, Sullivan PS, Madkins K, Rosenberg E, Swann G. Biomedical and behavioral outcomes of Keep It Up!: an eHealth HIV prevention program RCT. Am J Prev Med 2018 Aug;55(2):151-158 [FREE Full text] [doi: 10.1016/j.amepre.2018.04.026] [Medline: 29937115]
71. Reback C, Fletcher J, Swendeman D. Theory-based text messages reduce methamphetamine use and HIV sexual risk behaviors among MSM. 2017 Presented at: The 23rd Scientific Conference of the Society on Neuroimmune Pharmacology; 2017 Mar 1; Philadelphia, PA, USA.
72. Reback CJ, Fletcher JB, Swendeman DA, Metzner M. Theory-based text-messaging to reduce methamphetamine use and HIV sexual risk behaviors among men who have sex with men: automated unidirectional delivery outperforms bidirectional peer interactive delivery. AIDS Behav 2019 Jan;23(1):37-47 [FREE Full text] [doi: 10.1007/s10461-018-2225-z] [Medline: 30006792]
73. Rosser B, Oakes J, Konstan J, Hooper S, Horvath KJ, Danilenko GP, et al. Reducing HIV risk behavior of men who have sex with men through persuasive computing: results of the Men's INTernet Study-II. AIDS 2010 Aug 24;24(13):2099-2107 [FREE Full text] [doi: 10.1097/QAD.0b013e32833c4ac7] [Medline: 20601853]
74. Schonnesson LN, Bowen AM, Williams ML. Project SMART: preliminary results from a test of the efficacy of a Swedish internet-based HIV risk-reduction intervention for men who have sex with men. Arch Sex Behav 2016 Aug;45(6):1501-1511. [doi: 10.1007/s10508-015-0608-z] [Medline: 26545912]
75. Sullivan PS, Driggers R, Stekler JD, Siegler A, Goldenberg T, McDougal SJ, et al. Usability and acceptability of a mobile comprehensive HIV prevention app for men who have sex with men: a pilot study. JMIR Mhealth Uhealth 2017 Mar 09;5(3):e26 [FREE Full text] [doi: 10.2196/mhealth.7199] [Medline: 28279949]
76. Swendeman D, Ramanathan C, Baetscher L, Medich M, Scheffler A, Comulada WS, et al. Smartphone self-monitoring to support self-management among people living with HIV: perceived benefits and theory of change from a mixed-methods randomized pilot study. J Acquir Immune Defic Syndr 2015 May 01;69 Suppl 1:S80-S91 [FREE Full text] [doi: 10.1097/QAI.0000000000000570] [Medline: 25867783]
77. Wilkerson JM, Danilenko GP, Smolenski DJ, Myer BB, Rosser BRS. The role of critical self-reflection of assumptions in an online HIV intervention for men who have sex with men. AIDS Educ Prev 2011 Feb;23(1):13-24 [FREE Full text] [doi: 10.1521/aeap.2011.23.1.13] [Medline: 21341957]
78. Williams M, Bowen A, Ei S. An evaluation of the experiences of rural MSM who accessed an online HIV/AIDS health promotion intervention. Health Promot Pract 2010 Jul;11(4):474-482 [FREE Full text] [doi: 10.1177/1524839908324783] [Medline: 19116419]
79. Morrison LG, Yardley L, Powell J, Michie S. What design features are used in effective e-health interventions? A review using techniques from Critical Interpretive Synthesis. Telemed J E Health 2012 Mar;18(2):137-144. [doi: 10.1089/tmj.2011.0062] [Medline: 22381060]
80. Noar S, Black H, Pierce L. Efficacy of computer technology-based HIV prevention interventions: a meta-analysis. AIDS 2009 Jan 02;23(1):107-115. [doi: 10.1097/QAD.0b013e32831c5500] [Medline: 19050392]
81. Chib A, Lin SH. Theoretical advancements in mHealth: a systematic review of mobile apps. J Health Commun 2018;23(10-11):909-955. [doi: 10.1080/10810730.2018.1544676] [Medline: 30449261]
82. Christie HL, Bartels SL, Boots LM, Tange HJ, Verhey FR, de Vugt ME. A systematic review on the implementation of eHealth interventions for informal caregivers of people with dementia. Internet Interventions 2018 Sep;13:51-59. [doi: 10.1016/j.invent.2018.07.002]

83. Simblett S, Greer B, Matcham F, Curtis H, Polhemus A, Ferrão J, et al. Barriers to and facilitators of engagement with remote measurement technology for managing health: systematic review and content analysis of findings. J Med Internet Res 2018 Jul 12;20(7):e10480 [FREE Full text] [doi: 10.2196/10480] [Medline: 30001997]
84. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Behav. Med. Pract. Policy Res 2016 Dec 13;7(2):254-267. [doi: 10.1007/s13142-016-0453-1]
85. Internet/broadband fact sheet. Pew Research Center. Internet & Technology Web site. 2019. URL: https://www.pewresearch.org/internet/fact-sheet/internet-broadband/ [accessed 2020-06-21]
86. Hergenrather KC, Emmanuel D, Durant S, Rhodes SD. Enhancing HIV prevention among young men who have sex with men: a systematic review of HIV behavioral interventions for young gay and bisexual men. AIDS Educ Prev 2016 Jun;28(3):252-271. [doi: 10.1521/aeap.2016.28.3.252] [Medline: 27244193]
87. Gorini A, Mazzocco K, Triberti S, Sebri V, Savioni L, Pravettoni G. A P5 approach to m-Health: design suggestions for advanced mobile health technology. Front Psychol 2018;9:2066 [FREE Full text] [doi: 10.3389/fpsyg.2018.02066] [Medline: 30429810]
88. Sullivan PS, Jones J, Kishore N, Stephenson R. The roles of technology in primary HIV prevention for men who have sex with men. Curr HIV/AIDS Rep 2015 Dec;12(4):481-488. [doi: 10.1007/s11904-015-0293-5] [Medline: 26519083]
89. Moore GF, Audrey S, Barker M, Bond L, Bonell C, Hardeman W, et al. Process evaluation of complex interventions: Medical Research Council guidance. BMJ 2015 Mar 19;350:h1258 [FREE Full text] [doi: 10.1136/bmj.h1258] [Medline: 25791983]
90. Meiksin R, Melendez-Torres GJ, Falconer J, Witzel TC, Weatherburn P, Bonell C. Theories of change for e-health interventions targeting HIV/STIs and sexual risk, substance use and mental ill health amongst men who have sex with men: systematic review and synthesis. Syst Rev 2021 Jan 11;10(1):21 [FREE Full text] [doi: 10.1186/s13643-020-01523-2] [Medline: 33423693]

Abbreviations

LGBT: lesbian, gay, bisexual, and transgender
MSM: men who have sex with men
NIHR PHR: National Institute for Health Research Public Health Research Programme

Edited by R Kukafka; submitted 13.07.20; peer-reviewed by K Burns, D Pförringer; comments to author 03.09.20; revised version received 05.11.20; accepted 04.03.21; published 23.04.21.

Please cite as:
Meiksin R, Melendez-Torres GJ, Falconer J, Witzel TC, Weatherburn P, Bonell C
eHealth Interventions to Address Sexual Health, Substance Use, and Mental Health Among Men Who Have Sex With Men: Systematic Review and Synthesis of Process Evaluations
J Med Internet Res 2021;23(4):e22477
URL: https://www.jmir.org/2021/4/e22477
doi:10.2196/22477
PMID:33890855

Review

Clinical Effectiveness of Different Technologies for Diabetes in Pregnancy: Systematic Literature Review

Claudia Eberle[1], MD, Prof Dr; Maxine Loehnert[1], MSc; Stefanie Stichling[1], MSc

Medicine with Specialization in Internal Medicine and General Medicine, Hochschule Fulda - University of Applied Sciences, Fulda, Germany

Corresponding Author:
Claudia Eberle, MD, Prof Dr
Medicine with Specialization in Internal Medicine and General Medicine
Hochschule Fulda - University of Applied Sciences
Leipziger Street 123
Fulda, 36037
Germany
Phone: 49 661 9640 ext 6328
Email: claudia.eberle@hs-fulda.de

Abstract

Background: Hyperglycemia in pregnancy occurs worldwide and is closely associated with health issues in women and their offspring, such as pregnancy and birth complications, respectively, as well as comorbidities, such as metabolic and cardiovascular diseases. To optimize the management of diabetic pregnancies, sustainable strategies are urgently needed. Investigation of constantly evolving technologies for diabetes that help to manage pregnancy and health is required.

Objective: We aimed to conduct a systematic review to assess the clinical effectiveness of technologies for diabetes in pregnancy.

Methods: Relevant databases including MEDLINE (PubMed), Cochrane Library, Embase, CINAHL, and Web of Science Core Collection were searched in September 2020 for clinical studies (2008-2020). Findings were organized by type of diabetes, type of technology, and outcomes (glycemic control, pregnancy- and birth-related outcomes, and neonatal outcomes). Study quality was assessed using Effective Public Health Practice Project criteria.

Results: We identified 15 randomized controlled trials, 3 randomized crossover trials, 2 cohort studies, and 2 controlled clinical trials. Overall, 9 studies focused on type 1 diabetes, 0 studies focused on gestational diabetes, and 3 studies focused on both type 1 diabetes and type 2 diabetes. We found that 9 studies were strong quality, 11 were moderate quality, and 2 were weak quality. Technologies for diabetes seemed to have particularly positive effects on glycemic control in all types of diabetes, shown by some strong and moderate quality studies. Positive trends in pregnancy-related, birth-related, and neonatal outcomes were observed.

Conclusions: Technologies have the potential to effectively improve the management of diabetes during pregnancy. Further research on the clinical effectiveness of these technologies is needed, especially in pregnant women with type 2 diabetes.

(J Med Internet Res 2021;23(4):e24982) doi:10.2196/24982

KEYWORDS

diabetes technologies; diabetes management; pregnancy; digital health; eHealth; systematic review

Introduction

Hyperglycemia in pregnancy occurs worldwide and is closely associated with short- and long-term health issues, such as pregnancy complications in women and birth complications in offspring, and comorbidities, such as type 2 diabetes, as well as metabolic and cardiovascular diseases [1]. Approximately 20.4 million live births (16%) suffered from some form of hyperglycemia in pregnancy in 2019, worldwide [1]—84% were defined as gestational diabetes, 7.9% were diagnosed prior to pregnancy, and 8.5% were type 1 or type 2 diabetes first observed in pregnancy [1].

In 2019, 1,566,993.4 live births in North America and the Caribbean; 2,001,816.8 live births in Europe; and 6,594,159.4 live births in southeast Asia were affected by hyperglycemia in pregnancy [1]. Interestingly, the prevalence of gestational diabetes is rising quickly and was estimated to be 20.8% in North America and the Caribbean, 16.3% in Europe, and 27.0% in southeast Asia in 2019 [1]. The number of unreported cases is assumed to be high.

Hyperglycemia in pregnancy is associated with adverse pregnancy- and birth-related outcomes for mothers and offspring, such as increased risk of preeclampsia, cesarean deliveries, macrosomia, and shoulder dystocia [2-6], but hyperglycemia in pregnancy is also associated with adverse long-term consequences in mothers and children, such as subsequent increased risk of type 2 diabetes, obesity, metabolic syndrome, cardiovascular diseases, or even depression [2,4,5,7,8]. Based on the hypotheses of perinatal programming (fetal programming, developmental programming, transgenerational programming), intrauterine exposure to hyperglycemia is clearly associated with obesity, glucose, intolerance, type 2 diabetes, insulin resistance, metabolic syndrome, high blood pressure, and cardiovascular diseases in postnatal life of the offspring [5,9,10].

Sustainable strategies are urgently needed to effectively improve the management of diabetes in pregnancy with seamlessly integrated technological support for hyperglycemia in pregnancy. Technologies for diabetes, including hardware, devices, and software, are constantly evolving and can help to manage the condition, improving lives and health [11]. *Care 4.0* is a new paradigm to develop digital health and care services by focusing on trusted integrated networks of organizations, people, and technologies [12].

Recently, we demonstrated first insights in clinical effectiveness and successful approaches of mobile health (mHealth) apps as well as telemetric implementations to improve diabetic management in pregnancy [13,14]. Diabetes-specific mHealth apps and telemetric approaches improved clinical outcomes in the management of gestational diabetes effectively.

As a matter of course, there are many more technologies that have proven to be effective in diabetes management in general, for example, continuous glucose monitoring systems [15,16], continuous subcutaneous insulin infusions [17,18], and sensor augmented insulin pumps or closed-loop systems [19,20].

A previous review [21] examined the capability of diabetes technologies in the treatment of diabetic pregnancies; however, to further optimize management of diabetes in pregnancy, up-to-date analyses of technology for diabetes in pregnancies are needed. Therefore, we conducted a systematic review to assess the clinical effectiveness of different technologies for diabetes in pregnancies by analyzing glycemic control, pregnancy- and birth related outcomes, and neonatal outcomes.

Methods

Data Sources and Searches

A systematic review was carried out using PRISMA guidelines [22]. MEDLINE (PubMed), Cochrane Library, Embase, CINAHL, and Web of Science Core Collection were systematically searched for studies published from January 2008 until September 2020 (Multimedia Appendix 1). The following keywords were used as title and abstract search terms and selected from Medical Subject Headings and Embase subject headings databases: “diabetes mellitus,” “pregnancy,” “gestational diabetes mellitus,” “mobile application,” “insulin pump,” “continuous glucose monitoring,” “CGM,” “flash glucose monitoring,” “FGM,” “insulin infusion system” (Multimedia Appendix 2). Using a priori criteria for eligibility, 2 reviewers screened and selected the studies independently. We screened titles and abstracts, assessed full texts, and manually searched reference lists and Google Scholar.

Eligibility Criteria

We included peer-reviewed studies, published in German and English, that investigated the clinical effectiveness of diabetes technologies in pregnancies affected by hyperglycemia, including pre- and postconception. Studies with pregnant woman with a diagnosis of gestational diabetes or with preexisting type 1 diabetes or type 2 diabetes and that clearly reported the type of diabetes investigated were included. We included studies examining gestational diabetes patients only, type 1 diabetes only, type 2 diabetes only, and studies with mixed pregestational diabetes (type 1 diabetes and type 2 diabetes) groups. We included prospective controlled trials (clinical and observational trials). The following types of technologies were included: continuous glucose monitoring systems, continuous subcutaneous insulin infusions, sensor augmented insulin pumps, closed-loop systems, and mHealth apps. We involved studies examining glycemic control, pregnancy- and birth-related, and neonatal outcomes. Poster, comments, letter, study protocols, notes, and proceedings papers were excluded.

Data Extraction and Synthesis

We extracted the most common glycemic control (eg, hemoglobin A_{1c} [HbA_{1c}], fasting blood glucose, hyper- or hypoglycemia, insulin dose), pregnancy- and birth-related (eg, preterm birth, gestational weight gain), and neonatal (eg, birthweight, macrosomia) outcomes as well as author, year, study design, type of diabetes, type of technology, sample size, and main results. Findings were organized by type of diabetes, type of technology, and outcomes.

Quality Assessment

We applied Effective Public Health Practice Project (EPHPP), a validated tool that is suitable for intervention studies such as randomized controlled trials and controlled clinical trials. EPHPP is composed of criteria regarding selection bias, study design, confounders, blinding, data collection methods, withdrawals, and dropout. Study quality is rated as strong (SQ), moderate (MQ), or weak (WQ) [23].

Results

Study Selection and Characteristics

The search yielded 974 records, of which 754 remained after removing duplicates (3 records were identified through reference lists and Google Scholar). Of those, 693 were excluded based on titles and abstracts with our inclusion and exclusion criteria, and the full texts of 61 remaining publications were screened. Finally, 22 studies were included in this systematic review: 15 randomized controlled trials, 3 randomized crossover trials, 2 cohort studies, and 2 controlled clinical trials. Overall, 9 studies focused on type 1 diabetes, 10 studies focused on gestational diabetes, and 3 studies focused on type 1 diabetes and type 2 diabetes (mixed populations). We rated 9 studies as strong

quality, 11 studies as moderate quality, and 2 studies as weak quality (Multimedia Appendix 3). Table 1 gives an overview of the technologies, outcomes, and number of study participants..

Table 1. Overview of the technologies, outcomes, and study participants.

Population, technology, and outcome	Studies, n	Intervention group, n	Control group, n
Type 1 diabetes			
Continuous glucose monitoring systems			
HbA_{1c}[a]	3	199	236
Maternal hypoglycemia	2	172	177
Insulin dose and insulin requirements	3	199	236
Cesarean and vaginal deliveries	3	139	174
Preterm birth[b]	3	139	174
Maternal weight gain	2	112	115
Neonatal birthweight	2	127	159
Neonatal macrosomia	2	112	113
Large for gestational age	2	127	113
Continuous subcutaneous insulin infusion			
HbA_{1c}	4	218	485
Insulin dose	2	145	292
Maternal hypoglycemia	2	160	237
Maternal weight gain	2	223	273
Cesarean delivery	3	271	353
Preterm birth	3	271	353
Maternal hypertension and preeclampsia	2	159	196
Neonatal birthweight	2	158	195
Macrosomia (≥4000 g)	2	158	195
Large for gestational age	2	158	195
Small for gestational age	2	158	195
Admission to a higher level of neonatal care	2	158	195
Neonatal hypoglycemia	2	158	195
Closed-loop systems			
Hypoglycemic events, insulin dose	2	16	16
Gestational diabetes			
Continuous glucose monitoring systems			
HbA_{1c}	3	96	99
Fasting blood glucose	2	85	87
Preterm birth	3	96	99
Cesarean and vaginal deliveries	3	87	92
Maternal weight gain	2	76	80
Neonatal birthweight	4	147	154
Macrosomia (≥4000 g)	4	147	154
Large for gestational age	3	87	92
Small for gestational age	3	87	92
Admission to a higher level of neonatal care	3	96	99
Neonatal hypoglycemia	3	136	142
mHealth apps			

Population, technology, and outcome	Studies, n	Intervention group, n	Control group, n
HbA_{1c}	2	164	161
Fasting blood glucose	2	69	62
Off-target fasting blood glucose measurement	2	124	120
Patient compliance[c]	2	124	120
Pregnancy-induced hypertension or preeclampsia	3	218	212
Preterm birth	2	158	152
Vaginal and cesarean deliveries	5	390	393
Neonatal birthweight	3	218	211
Macrosomia	3	233	231
Neonatal hypoglycemia	4	277	263
Admission to a higher level of neonatal care	4	330	330
Type 1 and type 2 diabetes			
Continuous glucose monitoring systems			
HbA_{1c}	2	117	108
Plasma glucose	1	79	75
Preeclampsia	2	117	108
Cesarean delivery	2	117	108
Preterm birth	2	117	108
Neonatal birthweight	2	117	108
Large for gestational age	2	117	108
Neonatal hypoglycemia	2	117	108
Continuous subcutaneous insulin infusion			
HbA_{1c}	1	24	18

[a]HbA_{1c}: hemoglobin A_{1c}.

[b]*Preterm delivery* was defined as <37 weeks.

[c]*Patient compliance* was defined as ratio between actual and instructed blood glucose measurements×100.

Type 1 Diabetes: Overview

Of 9 publications analyzing 379 women in intervention (IG) and 763 in control groups (CG), 3 studies examined continuous glucose monitoring system, 4 studies examined continuous subcutaneous insulin infusion, and 2 studies examined closed-loop systems.

Type 1 Diabetes: Continuous Glucose Monitoring Systems

In general, 3 randomized controlled trials [24-26], analyzing 199 women in intervention groups and 236 women in control groups, were identified.

Glycemic Control Outcomes

HbA_{1c} values declined in all interventions. Feig et al [25] (MQ) observed significantly lower HbA_{1c} values in the intervention group than the control group (mean difference −0.19%; 95% CI −0.34 to −0.03; *P*=.0207). Cordua et al [24] (SQ) found positive but nonsignificant effects (IG: 6.0% vs CG: 6.2%; *P*=.23), and Petrovski et al [26] (SQ) reported significantly improved HbA_{1c} values (*P*<.05) in both groups (real-time continuous glucose monitoring system and intermittent continuous glucose monitoring system).

Petrovski et al [26] (SQ) found significantly less severe events of maternal hypoglycemia (real-time continuous glucose monitoring system group: *P*<.05), whereas Feig et al [25] (MQ) found time spent hypoglycemic was comparable in both groups (3% vs 4%; *P*=.10).

The intervention groups tended to have better target insulin dose or requirement values than those of the control groups (*P*=.14 [25], MQ; *P*=.08 [24], SQ). Petrovski et al [23] (SQ) did not find significant differences between constant and intermittent continuous glucose monitoring system (*P*>.05).

Pregnancy- and Birth-Related Outcomes

Cesarean delivery rates were clearly lower in the intervention groups (MQ, *P*=.18 [25]; SQ, *P*<.05 [23]). Cordua et al [24] (SQ) found significantly more vaginal deliveries in the intervention groups (IG: 74%; CG: 54%; *P*=.08).

All studies found slightly fewer preterm births in the interventions groups (Feig et al [25] <37 weeks: *P*=.57; <34 weeks: *P*=.19; Petrovski et al [23]: *P*>.05, Cordua et al [24]: *P*=.84).

Maternal gestational weight gains were lower in the continuous glucose monitoring system group (Feig et al [25] +8.9 kg vs +9.7 kg, *P*=.09) and the intermittent continuous glucose monitoring system group (Petrovski et al [26] +12.9 kg vs 13.4 kg, *P*>.05).

Neonatal Outcomes

Feig et al [25] (MQ) displayed a slightly lower mean birthweight in the continuous glucose monitoring system group (*P*=.37), whereas Cordua et al [24] (SQ) found a lower mean birthweight in their control group (*P*=.19).

Macrosomia rates were lower in intervention (Feig et al [25], MQ, IG: 23%; CG: 27%; *P*=.62) and real-time continuous glucose monitoring system (Petrovski et al [26], SQ, *P*>.05) groups.

Feig et al [25] found rates for large for gestational age were significantly lower in the intervention groups (*P*=.021), whereas Cordua et al [24] found no significant differences (*P*=.08).

Type 1 Diabetes: Continuous Subcutaneous Insulin Infusion

In general, 2 cohort studies [27,28], 1 randomized controlled trials [29], 1 randomized crossover trial [30] analyzing 148 intervention and 495 control participants were identified.

Glycemic Control Outcomes

Overall, 3 studies found positive effects on HbA_{1c} levels in favor of continuous subcutaneous insulin infusion. Jotic et al [28] (SQ) found a significant lower mean HbA_{1c} level in their intervention group than that in the control group at follow-up (6.5% intervention vs 6.8% control, *P*=.002). Cyganek et al [27] (SQ) also reported significant improvements in HbA_{1c} within all their treatment groups but did not find significant differences between groups.

Gutaj et al [30] (SQ) found significant lower insulin doses in the intervention group (0.54 U/kg/day intervention vs 0.63 U/kg/day control, *P*=.02), whereas Cyganek et al [27] (SQ) showed significantly lower daily insulin doses within each treatment group but no significant differences between groups.

Jotic et al [28] (SQ) reported that in early pregnancy, the majority of women on continuous subcutaneous insulin infusion had significantly fewer instances of hypoglycemia (*P*<.01), but Cyganek et al [27] (SQ) found fewer instances of severe hypoglycemia in their control group (*P*=.04).

Pregnancy- and Birth-Related Outcomes

Cyganek et al [27] (SQ) found less weight gain (+13.0 kg) in women in the control group receiving multiple daily injections than in the other groups (+14.7 kg for those only continuous subcutaneous insulin infusion; +15.2 kg for those moving from multiple daily injections to continuous subcutaneous insulin infusion; *P*=.005), and Feig at al [29] (MQ) did not find significant differences between groups (IG: mean +13.5 kg, SD 5.4 kg; CG: mean+13.5 kg, SD 0.7 kg; *P*=.48).

The studies could not find any clearly positive effects regarding cesarean deliveries in favor of continuous subcutaneous insulin infusion. Cyganek et al [27] (SQ) reported that the proportion of cesarean deliveries was larger in women using insulin pumps (no *P* value reported) because of clinical factors such as age, duration of diabetes, and microvascular complications. Feig et al [29] (MQ) (IG: n=81 vs CG: n=74, *P*=.32) and Jotic et al [28] (SQ) (IG: n=12 vs CG: n=13, *P*=.20) found no significant differences between groups.

There were no significant differences in preterm births between groups (*P*=.30 [27]; pump: n=39, 35%, vs multiple daily injections: n=35, 30%, *P*=.10 [29]; continuous subcutaneous insulin infusion n=13, 27.7%, vs multiple daily injections n=13, 16.3%, *P*=.17 [28]).

Feig et al [29] (MQ) found significantly fewer women with hypertension in the control group (IG: 14.4% vs CG: 5.2%; *P*=.025), whereas Jotic et al [28] (SQ) reported a lower, but not significantly so, rate of women with hypertension in their intervention group (*P*=.09). Feig et al [26] (MQ) also found fewer women with preeclampsia in the control group (*P*=.31), whereas Jotic et al [28] (MQ) noted fewer women with preeclampsia in the intervention group (IG: n=1, CG: n=3, *P* value not available).

Neonatal Outcomes

Neonatal birthweight was marginally higher in intervention groups (Feig et al [29], MQ, *P*=.91; Jotic et al [28], SQ, *P*=.98).

Jotic et al [28] (SQ) found fewer instances of macrosomia in the intervention group (*P*=.46), while Feig et al [29] (MQ) found slightly more instances in the intervention group (*P*=.89).

Jotic et al [28] reported fewer instances of newborns being large for gestational age in the intervention group (*P*=.59 [28]), whereas Feig et al found slightly more newborns who were large for gestational age in the interventions group (*P*=.55).

Jotic et al [28] (SQ) found no instances of newborns being small for gestational age in the intervention and 5 (4%) instances of newborns being small for gestational age in the control group (*P* value not available), whereas Feig et al [29] (MQ) observed 3 and 1 instances of newborns being small for gestational age in the intervention and control groups, respectively (*P* value not available).

Feig et al [29] (MQ) and Jotic et al [28] (SQ) found differences in admissions to higher level neonatal care in favor of the control groups (IG: 44.5% vs CG: 29.6%, *P*=.02 [29]; and *P*=.28 [28]).

Jotic et al [28] (SQ) found slightly fewer instances of neonatal hypoglycemia in the intervention group (IG: n=5, CG: n=6, *P*=.54), whereas Feig et al [29] (MQ) reported a significant difference in favor of the control group (IG: 31.8% vs CG: 19.1%, *P*=.05).

Type 1 Diabetes: Closed-Loop Systems

Closed-loop systems were investigated in 2 randomized crossover trials including a total of 32 women comparing

closed-loop systems with sensor augmented insulin pumps [29,31].

Stewart et al [31] found a significant difference between groups in percentage of time that blood glucose levels were between 63 and 140 mg/dL for overnight evaluation (*P*=.002) as well as day and night evaluation (*P*<.001) in favor of the closed-loop group [31], while Stewart et al [29] (MQ) found no significant difference (closed-loop: 62.3% vs sensor augmented insulin pumps: 60.1%, 95% CI 24.1% to 8.3%, *P*=.47).

Stewart et al [32] (MQ) reported significant fewer hypoglycemic events during the closed-loop treatment (n=8 vs n=12.5, *P*=.04) [32], whereas Stewart et al [31] (MQ) found no significant differences in overnight analysis (closed-loop: median 3.0; sensor augmented insulin pumps: median 2.5, *P*=.68) or day and night analysis (closed-loop: median 11.0, sensor augmented insulin pumps: median 12.0, *P*=.19).

Neither Stewart et al [32] (MQ) (closed-loop 43.7 vs sensor augmented insulin pumps 41.5 units/day, *P*=.56) nor Stewart et al [31] (MQ) (closed-loop 58.2 vs sensor augmented insulin pumps 59.8 units/day, *P*=.56) found significant differences between treatments regarding insulin dose [31,32].

Gestational Diabetes: Overview

In general, 4 studies focused on continuous glucose monitoring system, and 6 studies focused on mHealth apps (overall: IG: n=555; CG: n=609).

Gestational Diabetes: Continuous Glucose Monitoring Systems

We identified 4 randomized controlled trials (IG: n=147, CG: n=154). Whereas 3 studies used self-monitoring blood glucose as control [33-35], 1 study used blinded continuous glucose monitoring system which did not display the glucose readings to the participant [36].

Glycemic Control Outcomes

All studies reported clearly lower HbA_{1c} values in their intervention groups. Paramasivam et al [34] (MQ) found a significant difference in favor of the intervention group (mean 5.2%, SD 0.4% vs mean 5.6%, SD 0.6%; *P*=.006). Alfadhli et al [33] (*P*=.168) (MQ) and Lane et al [36] (*P*=.3) (SQ) found lower mean HbA_{1c} values in their intervention groups (*P*>.05).

Paramasivam et al [34] (MQ) (*P*=.101) and Alfadhli et al [33] (*P*=.092) (MQ) found lower fasting blood glucose values in the intervention groups.

Pregnancy- and Birth-Related Outcomes

Alfadhli et al [33] (MQ) (IG=16.3%, CG=9.5%, *P*=.373), Paramasivam et al [34] (MQ) (IG: n=3, 12%; CG: n=1, 4%; *P*=.609) and Lane et al [36] (SQ) (IG: n=1, 9.1%; CG: n=2, 16.7%; *P*>.999) found no significant differences in the occurrence of preterm births.

Paramasivam et al [34] (MQ) reported clearly more vaginal deliveries in the intervention group (*P*=.258) [34], while Wei et al [35] (MQ) (*P*=.370) and Lane et al [36] (SQ) (*P*>.999) noted slightly more cesarean deliveries in the control groups.

Both studies showed positive effects on maternal weight gain. Wei et al [35] (MQ) reported that the intervention participants had significantly more appropriate and less excessive weight gain (*P*=.039) and Paramasivam et al [34] (MQ) mentioned slightly less weight gain in the intervention group (*P*=.917).

Neonatal Outcomes

Wei et al [35] (*P*=.084), Alfadhli et al [33] (MQ) (*P*=.130) and Paramasivam et al [34] (MQ) (*P*=.311) reported a lower, yet not significant, mean birthweight of newborns in the intervention groups. Lane et al [36] (SQ) reported a higher mean birthweight of newborns in the intervention group (*P*=.4).

Wei at al [35] (MQ) (*P*=.410) and Alfadhli et al [33] (MQ) (*P*=.488) reported fewer instances of macrosomia in the intervention groups, while Lane et al [36] (SQ) noted more instances of macrosomia in the intervention group (*P*=.2), and Paramasivam et al (MQ) [34] found none.

Wei et al [35] (MQ) (*P*=.071) and Paramasivam et al [34] (MQ) (*P*=.490) found clearly fewer instances of newborns being large for gestational age in their intervention groups and, in contrast, Lane et al [36] (SQ) found more instances of newborns being large for gestational age in the intervention group (*P*=.2).

Lane et al [36] (SQ) (*P*>.999) and Wei et al [35] (MQ) (*P*>.999) reported very slightly fewer instances of newborns being small for gestational age in the intervention groups. Paramasivam et al [34] reported none.

Alfadhli et al [33] (MQ) (*P*=.653) reported slightly more admissions, Lane et al [36] (*P*>.999) noted slightly fewer, and Paramasivam et al [34] (MQ) (both groups n=1, 4%) observed the same number of admissions (*P*>.999) to a higher level of neonatal care.

Wei at al [35] (MQ) (*P*=.410) and Paramasivam et al [34] (MQ) (*P*>.999) reported slightly fewer instances of neonatal hypoglycemia among intervention participants whereas Alfadhli et al [33] (MQ) (*P*=.758) found slightly fewer instances of neonatal hypoglycemia in the control group.

Gestational Diabetes: mHealth Apps

We identified 5 randomized controlled trials and 1 controlled clinical trial investigating a total of 408 women in the intervention and 405 women in the control groups.

Glycemic Control Outcomes

Overall, the intervention groups showed lower HbA_{1c} values than the control groups. Guo et al [37] (SQ) displayed a significant difference in favor of the intervention group (–1.3% intervention vs –0.6% control, *P*<.001), while Mackillop et al [38] (WQ) recognized a slight increase in both groups (IG: 0.02% per 28 days; CG: 0.03% per 28 days; 95% CI –0.05 to 0.03).

Both Yang et al [39] (MQ) (*P*<.001) and Bromuri et al [40] (MQ) (*P*<.001) found clear, significant differences in fasting blood glucose levels favoring the intervention groups.

Miremberg et al [41] (SQ) (*P*<.001) as well as Guo et al [37] (SQ) (*P*<.001) found significant differences in off-target fasting

blood glucose measurement favoring the intervention groups [37,41].

Guo et al [37] (SQ) (*P*<.001) and Miremberg et al [41] (SQ) (*P*<.001) reported significant differences in support of the intervention groups in patient compliance (defined as ratio between actual and instructed blood glucose measurements×100).

Pregnancy- and Birth-Related Outcomes

All intervention groups showed lower rates of pregnancy-induced hypertension or preeclampsia, but no significant differences were found by Mackillop et al [38] (WQ) (*P*=.22), Miremberg et al [41] (SQ) (*P*>.99), or Yang et al [39] MQ) (*P*=.347).

Overall, there were fewer preterm births in the intervention groups, but neither Yang et al [39] (MQ) (*P*=.248) nor Mackillop et al [38] (WQ) (*P*>.05) found significant differences.

Positive intervention effects—more vaginal deliveries and less cesarean sections—were observed in almost all studies [37-39,41,42]; 2 studies showed statistical significance (Borgen et al [42], WQ, *P*=.03; Mackillop et al [38], WQ, *P*=.005).

Neonatal Outcomes

The results show a clear trend toward lower birthweight in the newborns of the intervention groups, but no significant differences were found by Mackillop et al [38] (WQ) (*P*=.18), Miremberg et al [41] (SQ) (*P*=.878), or Yang et al [39] (MQ) (*P*=.988).

Fewer instances of macrosomia occurred in the intervention groups than in the control groups, but no significant differences were found by Guo et al [37] (SQ) (*P*=.295), Yang et al (MQ) [39] (*P*=.542) and Borgen et al [42] (WQ) (*P*=.69).

In general, most studies showed positive intervention effects. Guo et al [37] (SQ)(*P*=.185), Mackillop et al (WQ) [38] (*P*=.42) and Yang et al (MQ)[39] (*P* value not available) observed slightly fewer instances of neonatal hypoglycemia for intervention participants.

The results show a trend toward infants of mothers in the interventions groups being transferred less often to neonatal intensive care units (Borgen et al [42], WQ, *P*=.38; Mackillop et al [38], WQ, *P*=.08; Miremberg et al [41], SQ, *P*>.99; Yang et al [39], MQ, *P*=.657).

Type 1 and Type 2 Diabetes: Overview

In total, 3 studies analyzing 141 women in the intervention and 126 women in the control groups were included [43-45]. Two studies observed continuous glucose monitoring system and one insulin pumps.

Type 1 and Type 2 Diabetes: Continuous Glucose Monitoring Systems

Overall, 2 randomized controlled trials analyzed 117 intervention and 108 control participants. Secher et al [45] included 123 women with type 1 diabetes and 31 women with type 2 diabetes, while Murphy et al [44] included 45 women with type 1 diabetes and 25 type 2 diabetes.

Glycemic Control Outcomes

Murphy et al [44] (MQ) noted that HbA_{1c} levels were consistently and significantly lower in the intervention group (gestation 32-36 weeks: *P*=.007), while Secher et al [45] (MQ) reported comparable HbA_{1c} values (IG: 6.1%, 95% CI 5.1%-7.8%; CG: 6.1%, 95% CI 4.8%-8.2%; *P*=.39).

Median self-monitored plasma glucose levels at 33 weeks were comparable between groups (IG: 6.2 mmol/L, 4.7-7.9; CG: 6.2 mmol/L, 4.9-7.9) [45].

Pregnancy- and Birth-Related Outcomes

Secher et al (MQ) and Murphy et al (MQ) observed very slightly more instances of preeclampsia in the treatment groups (*P*=.83 [45] and *P*=.5 [44], respectively).

Secher et al [45] (MQ) observed a lower rate of cesarean delivery (*P*=.30) and Murphy et al [44] (MQ) reported a lower emergency cesarean delivery rate (*P*=.3) in the intervention group.

Murphy et al [44] (MQ) observed slightly fewer preterm births (*P*=.8), whereas Secher et al [45] (MQ) reported slightly more preterm births (*P*=.47) in the treatment group.

Neonatal Outcomes

Murphy et al [44] (MQ) found significantly lower median birthweight percentile in the intervention group (69 vs 93, *P*=.02), whereas Secher et al [45] (MQ) found the birthweight was slightly higher in the intervention group (3510 g vs 3436 g, *P*=.80).

Murphy et al [44] (MQ) reported that fewer newborns who were extremely large for gestational age (≥97.7th percentile) were born to intervention participants (14% vs 30%, *P*=.1]), whereas Secher et al [45] (MQ) reported slightly more newborns who were extremely large for gestational were born to participants in the treatment group (*P*=.19 [45]).

Both studies reported clearly fewer instances of neonatal hypoglycemia in the intervention groups (36% vs 40%, *P*=.62 [45]; and 8% vs 17%, *P*=.5 [44])

Type 1 and Type 2 Diabetes: Continuous Subcutaneous Insulin Infusion

One study [43] examined continuous subcutaneous insulin infusion compared to conventional insulin therapy (intervention n=24 vs control n=18). Kernaghan et al [43] (SQ) included n=1 women with type 2 diabetes.

The intervention group showed lower mean HbA_{1c} levels in the first (*P*=.41) and second (*P*=.27) trimester. The birthweight *z* score was slightly lower in the control group (*P*=.86). The mean estimated fetal weight *z* score was lower in the intervention group (*P*=.16).

Discussion

Principal Results

In general, technologies seemed to have particularly positive effects on glycemic control in all types of diabetes, as shown

by studies of strong and moderate quality. For type 1 diabetes, there were no studies evaluating the effectiveness of mHealth apps and, for gestational diabetes, there were no studies that evaluated the effectiveness of continuous subcutaneous insulin infusion or closed-loop systems. It is particularly useful for women with pregestational diabetes to use mHealth apps, as they can become familiar with them before pregnancy. Furthermore, there is a lack of research focusing on the effectiveness of technologies for pregnant women with type 2 diabetes.

Type 1 Diabetes

Continuous Glucose Monitoring Systems

Overall, glycemic control improved with continuous glucose monitoring systems in strong and moderate quality studies, especially HbA_{1c}. Positive trends were observed for hypoglycemia and insulin dose. Intermittent continuous glucose monitoring systems seemed to be preferable to real-time continuous glucose monitoring systems, with fewer cesarean deliveries and more vaginal deliveries; however, the findings were not statistically significant. Regarding maternal weight gain and preterm births, improvements seem to be possible through the use of continuous glucose monitoring system, especially intermittent continuous glucose monitoring systems. Birthweight, large for gestational age, macrosomia, and neonatal hypoglycemia rates showed trends for improvement with continuous glucose monitoring system, as some moderate and strong quality studies showed nonsignificant, positive effects.

Continuous Subcutaneous Insulin Infusion

Overall, continuous subcutaneous insulin infusion improved glycemic control, especially HbA_{1c} and insulin dose, as most strong and moderate quality studies showed positive intervention effects. In the case of maternal hypoglycemia, one study [28] found significantly fewer instances in the intervention group, but another study [27] found fewer instances in the control group.

No clear trend regarding the effectiveness of continuous subcutaneous insulin infusion can be derived regarding pregnancy- and birth-related outcomes such as maternal weight gain, preterm births, cesarean delivery, maternal hypertension, or preeclampsia. There were no clear trends in relation to neonatal outcomes such as birthweight, large for gestational age, small for gestational age, macrosomia, hypoglycemia, and admission to high-level neonatal care because too little data were available.

Closed-Loop Systems

A trend toward improved glycemic control was shown in moderate quality studies, but the sample size was very small (n=32 patients).

Gestational Diabetes

Continuous Glucose Monitoring Systems

In general, continuous glucose monitoring system improved glycemic control, particularly HbA_{1c} and fasting blood glucose levels in strong and moderate quality studies; however, there was no obvious trend regarding 2-hour plasma blood glucose level and insulin dose. There were also no clear trends regarding preterm births or cesarean and vaginal deliveries. Both studies examining maternal weight gain showed positive but nonsignificant intervention effects.

Moreover, continuous glucose monitoring system improved birthweight, macrosomia, large for gestational age, small for gestational age, and neonatal hypoglycemia measures compared to those in control groups, though not significantly. Overall, there was no clear trend with respect to admission to higher level of neonatal care.

mHealth Apps

Gestational diabetes–specific mHealth apps displayed positive, significant effects in glycemic control outcomes such as HbA_{1c} level, fasting blood glucose level, off-target blood glucose measurement, and patient adherence (actual vs instructed blood glucose measurements) in strong and moderate quality studies. mHealth apps improved pregnancy-induced hypertension or preeclampsia and preterm birth outcomes, but not significantly. In addition, the treatments showed positive effects regarding vaginal deliveries and cesarean deliveries. Gestational diabetes–specific mHealth apps indicated noticeable positive, but not significant effects, on neonatal outcomes such as birthweight, macrosomia, hypoglycemia, and rate of admission to a higher level of care.

Type 1 and Type 2 Diabetes

Continuous Glucose Monitoring Systems

The sample sizes of the studies were very small. One moderate-quality study [45] could not find any clear differences in HbA_{1c} values between intervention and control group, while another moderate-quality study [44] showed significant positive intervention effects. Cesarean deliveries were required less often. There was no obvious trend regarding preeclampsia and preterm births. While no clear effects were found regarding birthweight and being large for gestational age, continuous glucose monitoring system showed an improvement in terms of neonatal hypoglycemia outcomes, but not significantly.

Continuous Subcutaneous Insulin Infusion

One strong-quality study [43] indicated positive trends in terms of HbA_{1c}, birthweight *z* score, and estimated fetal weight *z* score.

Strengths and Limitations

To our knowledge, this is the first systematic review of different technologies for diabetes that differentiated between various types of diabetes. Hence, our review opens up new perspectives on the topic hyperglycemia in pregnancy. However, the body of research included in this review may be limited depending on the type of diabetes and technology. Further research, with larger sample sizes and that takes into account women in the pre-conception phase, is needed, especially. Furthermore, we included only German and English papers, and we performed a qualitative analysis.

Comparison With Prior Work

Our results are in line with those from other reviews. In their systematic review, Feig et al [46] also reported that continuous

glucose monitoring system improved HbA_{1c} values in patients with type 1 diabetes. Furthermore, they showed lower rates of large for gestational age, higher time in range, and fewer adverse neonatal outcomes [46].

In their meta-analysis, Rys et al [47] investigated HbA_{1c} values in pregnant women with type 1 diabetes using continuous subcutaneous insulin infusion or multiple daily injections. They reported lower HbA_{1c} in the first trimester for continuous subcutaneous insulin infusion users (weighted mean difference: −0.45%; 95% CI −0.62 to −0.27).

Research on gestational diabetes–specific mHealth apps is very limited. Skar et al [48] reported gestational diabetes–specific mHealth apps to be effective in increasing the confidence of women with gestational diabetes in their self-management and their motivation for behavior changes. In addition, Chen et al [49] concluded that gestational diabetes specific mHealth apps can provide time- and cost-efficient personalized interventions to improve gestational diabetes management and clinical outcomes. Yu et al [50] reported a clear superiority of continuous glucose monitoring system compared to self-monitoring blood glucose in detecting hypo- and hyperglycemic episodes.

Conclusions

Technologies for diabetes seem to have a particularly positive effect on glycemic control in all types of diabetes. In pregnant women with type 1 diabetes, continuous glucose monitoring system as well as closed-loop systems seem to improve glycemic control. In women with gestational diabetes, the use of continuous glucose monitoring system systems has been shown, by this review, to improve glycemic control. mHealth apps can also improve glycemic control as well as certain pregnancy and birth related and neonatal outcomes in women with gestational diabetes.

Furthermore, this review showed that there is a lack of research on the clinical effectiveness of technologies for pregnant women with type 2 diabetes. In addition, sample sizes were, in many cases, rather small. Further research is needed to gain more firm evidence on the clinical effectiveness of diabetes technologies in pregnant women.

Acknowledgments

This manuscript was created in the context of a project funded by the German Research Foundation (*Deutsche Forschungsgemeinschaft*, EB 440/4-1); therefore, we would like to thank the *Deutsche Forschungsgemeinschaft*.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategies.
[DOCX File , 20 KB - jmir_v23i4e24982_app1.docx]

Multimedia Appendix 2
PRISMA flow diagram.
[DOCX File , 43 KB - jmir_v23i4e24982_app2.docx]

Multimedia Appendix 3
Quality assessment using Effective Public Health Practice Project criteria.
[DOCX File , 27 KB - jmir_v23i4e24982_app3.docx]

References

1. Diabetes atlas 9th edition. International Diabetes Federation. 2019. URL: https://www.diabetesatlas.org/en/ [accessed 2021-04-19]
2. Chiefari E, Arcidiacono B, Foti D, Brunetti A. Gestational diabetes mellitus: an updated overview. J Endocrinol Invest 2017 Sep;40(9):899-909. [doi: 10.1007/s40618-016-0607-5] [Medline: 28283913]
3. Ferrara A. Increasing prevalence of gestational diabetes mellitus: a public health perspective. Diabetes Care 2007 Jul;30 Suppl 2:S141-S146. [doi: 10.2337/dc07-s206] [Medline: 17596462]
4. Eberle C. Fetale Programmierung des Diabetes mellitus Typ 2. MMW - Fortschritte der Medizin 2016 Aug 13;152(S3):76-82. [doi: 10.1007/bf03373173]
5. Eberle C, Ament C. Diabetic and metabolic programming: mechanisms altering the intrauterine milieu. ISRN Pediatr 2012;2012:975685 [FREE Full text] [doi: 10.5402/2012/975685] [Medline: 23213562]
6. Eberle C, Ament C. A combined in vivo and in silico model shows specific predictors of individual trans-generational diabetic programming. J Dev Orig Health Dis 2020 Aug 18:1-8. [doi: 10.1017/S2040174420000471] [Medline: 32808917]

7. Bellamy L, Casas J, Hingorani AD, Williams D. Type 2 diabetes mellitus after gestational diabetes: a systematic review and meta-analysis. Lancet 2009 May 23;373(9677):1773-1779. [doi: 10.1016/S0140-6736(09)60731-5] [Medline: 19465232]
8. Kozhimannil KB, Pereira MA, Harlow BL. Association between diabetes and perinatal depression among low-income mothers. JAMA 2009 Feb 25;301(8):842-847. [doi: 10.1001/jama.2009.201] [Medline: 19244191]
9. West NA, Crume TL, Maligie MA, Dabelea D. Cardiovascular risk factors in children exposed to maternal diabetes in utero. Diabetologia 2011 Mar;54(3):504-507 [FREE Full text] [doi: 10.1007/s00125-010-2008-1] [Medline: 21153896]
10. Tam WH, Ma RCW, Yang X, Li AM, Ko GTC, Kong APS, et al. Glucose intolerance and cardiometabolic risk in adolescents exposed to maternal gestational diabetes: a 15-year follow-up study. Diabetes Care 2010 Jun;33(6):1382-1384 [FREE Full text] [doi: 10.2337/dc09-2343] [Medline: 20215448]
11. American Diabetes Association. Diabetes technology. Diabetes Care 2020 Jan;43(Suppl 1):S77-S88. [doi: 10.2337/dc20-S007] [Medline: 31862750]
12. Chute C, French T. Introducing Care 4.0: an integrated care paradigm built on industry 4.0 capabilities. Int J Environ Res Public Health 2019 Jun 25;16(12):1-17 [FREE Full text] [doi: 10.3390/ijerph16122247] [Medline: 31242687]
13. Eberle C, Ament C. Diabetologie und Stoffwechsel 2017 May 5;12(S 01):S1-S84. [doi: 10.1055/s-0037-1601701]
14. Eberle C. Einsatz neuer Technologien in Diagnostik und Therapie: Diabetes-Smartphone-Apps. In: Stupin J, Schäfer-Graf U, Hummel M, editors. Diabetes in der Schwangerschaft 1. Auflage. Berlin: De Gruyter; 2020.
15. Acosta GJ, Brown S, Zand AM, Kansara A, Sadhu A. USA vs Europe: who is leading the diabetes tech race? Curr Diab Rep 2019 Nov 16;19(11):128. [doi: 10.1007/s11892-019-1237-z] [Medline: 31734741]
16. Aleppo G, Ruedy KJ, Riddlesworth TD, Kruger DF, Peters AL, Hirsch I, REPLACE-BG Study Group. REPLACE-BG: a randomized trial comparing continuous glucose monitoring with and without routine blood glucose monitoring in adults with well-controlled type 1 diabetes. Diabetes Care 2017 Apr;40(4):538-545 [FREE Full text] [doi: 10.2337/dc16-2482] [Medline: 28209654]
17. McAdams BH, Rizvi AA. An overview of insulin pumps and glucose sensors for the generalist. J Clin Med 2016 Jan 04;5(1):5 [FREE Full text] [doi: 10.3390/jcm5010005] [Medline: 26742082]
18. Sora ND, Shashpal F, Bond EA, Jenkins AJ. Insulin pumps: review of technological advancement in diabetes management. Am J Med Sci 2019 Nov;358(5):326-331. [doi: 10.1016/j.amjms.2019.08.008] [Medline: 31655714]
19. Ly TT, Nicholas JA, Retterath A, Lim EM, Davis EA, Jones TW. Effect of sensor-augmented insulin pump therapy and automated insulin suspension vs standard insulin pump therapy on hypoglycemia in patients with type 1 diabetes: a randomized clinical trial. JAMA 2013 Sep 25;310(12):1240-1247. [doi: 10.1001/jama.2013.277818] [Medline: 24065010]
20. Elleri D, Dunger DB, Hovorka R. Closed-loop insulin delivery for treatment of type 1 diabetes. BMC Med 2011 Nov 09;9:120 [FREE Full text] [doi: 10.1186/1741-7015-9-120] [Medline: 22071283]
21. Ming W, Mackillop LH, Farmer AJ, Loerup L, Bartlett K, Levy JC, et al. Telemedicine technologies for diabetes in pregnancy: a systematic review and meta-analysis. J Med Internet Res 2016 Nov 09;18(11):e290 [FREE Full text] [doi: 10.2196/jmir.6556] [Medline: 27829574]
22. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. BMJ 2009;339:b2535 [FREE Full text] [Medline: 19622551]
23. Armijo-Olivo S, Stiles CR, Hagen NA, Biondo PD, Cummings GG. Assessment of study quality for systematic reviews: a comparison of the Cochrane Collaboration Risk of Bias Tool and the Effective Public Health Practice Project Quality Assessment Tool: methodological research. J Eval Clin Pract 2012 Feb;18(1):12-18. [doi: 10.1111/j.1365-2753.2010.01516.x] [Medline: 20698919]
24. Cordua S, Secher AL, Ringholm L, Damm P, Mathiesen ER. Real-time continuous glucose monitoring during labour and delivery in women with Type 1 diabetes - observations from a randomized controlled trial. Diabet Med 2013 Nov;30(11):1374-1381. [doi: 10.1111/dme.12246] [Medline: 23758126]
25. Feig DS, Donovan LE, Corcoy R, Murphy KE, Amiel SA, Hunt KF, CONCEPTT Collaborative Group. Continuous glucose monitoring in pregnant women with type 1 diabetes (CONCEPTT): a multicentre international randomised controlled trial. Lancet 2017 Nov 25;390(10110):2347-2359 [FREE Full text] [doi: 10.1016/S0140-6736(17)32400-5] [Medline: 28923465]
26. Petrovski G, Dimitrovski C, Bogoev M, Milenkovic T, Ahmeti I, Bitovska I. Is there a difference in pregnancy and glycemic outcome in patients with type 1 diabetes on insulin pump with constant or intermittent glucose monitoring? a pilot study. Diabetes Technol Ther 2011 Nov;13(11):1109-1113. [doi: 10.1089/dia.2011.0081] [Medline: 21751889]
27. Cyganek K, Hebda-Szydlo A, Katra B, Skupien J, Klupa T, Janas I, et al. Glycemic control and selected pregnancy outcomes in type 1 diabetes women on continuous subcutaneous insulin infusion and multiple daily injections: the significance of pregnancy planning. Diabetes Technol Ther 2010 Jan;12(1):41-47. [doi: 10.1089/dia.2009.0081] [Medline: 20082584]
28. Jotic A, Milicic T, Lalic K, Lukic L, Macesic M, Stanarcic Gajovic J, et al. Evaluation of glycaemic control, glucose variability and hypoglycaemia on long-term continuous subcutaneous infusion vs. multiple daily injections: observational study in pregnancies with pre-existing type 1 diabetes. Diabetes Ther 2020 Feb 15;11(4):845-858. [doi: 10.1007/s13300-020-00780-7]
29. Feig DS, Corcoy R, Donovan LE, Murphy KE, Barrett JFR, Sanchez JJ, CONCEPTT Collaborative Group. Pumps or multiple daily injections in pregnancy involving type 1 diabetes: a prespecified analysis of the CONCEPTT randomized trial. Diabetes Care 2018 Dec;41(12):2471-2479. [doi: 10.2337/dc18-1437] [Medline: 30327362]

30. Gutaj P, Zawiejska A, Brązert J, Wender-Ożegowska E. Association between preconceptional treatment with insulin pumps and improved metabolic status in early pregnancy in women with type 1 diabetes. Pol Arch Med Wewn 2015;125(5):329-336 [FREE Full text] [doi: 10.20452/pamw.2830] [Medline: 25827639]
31. Stewart ZA, Wilinska ME, Hartnell S, Temple RC, Rayman G, Stanley KP, et al. Closed-loop insulin delivery during pregnancy in women with type 1 diabetes. N Engl J Med 2016 Aug 18;375(7):644-654. [doi: 10.1056/NEJMoa1602494] [Medline: 27532830]
32. Stewart ZA, Wilinska ME, Hartnell S, O'Neil LK, Rayman G, Scott EM, et al. Day-and-night closed-loop insulin delivery in a broad population of pregnant women with type 1 diabetes: a randomized controlled crossover trial. Diabetes Care 2018 Jul;41(7):1391-1399. [doi: 10.2337/dc17-2534] [Medline: 29535135]
33. Alfadhli E, Osman E, Basri T. Use of a real time continuous glucose monitoring system as an educational tool for patients with gestational diabetes. Diabetol Metab Syndr 2016;8:48 [FREE Full text] [doi: 10.1186/s13098-016-0161-5] [Medline: 27468313]
34. Paramasivam SS, Chinna K, Singh AKK, Ratnasingam J, Ibrahim L, Lim LL, et al. Continuous glucose monitoring results in lower HbA in Malaysian women with insulin-treated gestational diabetes: a randomized controlled trial. Diabet Med 2018 Aug;35(8):1118-1129. [doi: 10.1111/dme.13649] [Medline: 29663517]
35. Wei Q, Sun Z, Yang Y, Yu H, Ding H, Wang S. Effect of a CGMS and SMBG on maternal and neonatal outcomes in gestational diabetes mellitus: a randomized controlled trial. Sci Rep 2016;6:19920 [FREE Full text] [doi: 10.1038/srep19920] [Medline: 26814139]
36. Lane AS, Mlynarczyk MA, de Veciana M, Green LM, Baraki DI, Abuhamad AZ. Real-time continuous glucose monitoring in gestational diabetes: a randomized controlled trial. Am J Perinatol 2019 Jul 28;36(9):891-897. [doi: 10.1055/s-0039-1678733] [Medline: 30818406]
37. Guo H, Zhang Y, Li P, Zhou P, Chen LM, Li SY. Evaluating the effects of mobile health intervention on weight management, glycemic control and pregnancy outcomes in patients with gestational diabetes mellitus. J Endocrinol Invest 2019 Jun;42(6):709-714. [doi: 10.1007/s40618-018-0975-0] [Medline: 30406378]
38. Mackillop L, Hirst JE, Bartlett KJ, Birks JS, Clifton L, Farmer AJ, et al. Comparing the efficacy of a mobile phone-based blood glucose management system with standard clinic care in women with gestational diabetes: randomized controlled trial. JMIR Mhealth Uhealth 2018 Mar 20;6(3):e71 [FREE Full text] [doi: 10.2196/mhealth.9512] [Medline: 29559428]
39. Yang P, Lo W, He ZL, Xiao XM. Medical nutrition treatment of women with gestational diabetes mellitus by a telemedicine system based on smartphones. J Obstet Gynaecol Res 2018 May 23;44(7):1228-1234. [doi: 10.1111/jog.13669]
40. Bromuri S, Puricel S, Schumann R, Krampf J, Ruiz J, Schumacher M. An expert personal health system to monitor patients affected by gestational diabetes mellitus: a feasibility study. AIS 2016 Mar 15;8(2):219-237. [doi: 10.3233/AIS-160365]
41. Miremberg H, Ben-Ari T, Betzer T, Raphaeli H, Gasnier R, Barda G, et al. The impact of a daily smartphone-based feedback system among women with gestational diabetes on compliance, glycemic control, satisfaction, and pregnancy outcome: a randomized controlled trial. Am J Obstet Gynecol 2018 Apr;218(4):453.e1-453.e7. [doi: 10.1016/j.ajog.2018.01.044] [Medline: 29425836]
42. Borgen I, Småstuen MC, Jacobsen AF, Garnweidner-Holme LM, Fayyad S, Noll J, et al. Effect of the Pregnant+ smartphone application in women with gestational diabetes mellitus: a randomised controlled trial in Norway. BMJ Open 2019 Nov 11;9(11):e030884 [FREE Full text] [doi: 10.1136/bmjopen-2019-030884] [Medline: 31719080]
43. Kernaghan D, Farrell T, Hammond P, Owen P. Fetal growth in women managed with insulin pump therapy compared to conventional insulin. Eur J Obstet Gynecol Reprod Biol 2008 Mar;137(1):47-49. [doi: 10.1016/j.ejogrb.2007.05.006] [Medline: 18006209]
44. Murphy HR, Rayman G, Lewis K, Kelly S, Johal B, Duffield K, et al. Effectiveness of continuous glucose monitoring in pregnant women with diabetes: randomised clinical trial. BMJ 2008 Sep 25;337:a1680 [FREE Full text] [doi: 10.1136/bmj.a1680] [Medline: 18818254]
45. Secher AL, Ringholm L, Andersen HU, Damm P, Mathiesen ER. The effect of real-time continuous glucose monitoring in pregnant women with diabetes: a randomized controlled trial. Diabetes Care 2013 Jul;36(7):1877-1883 [FREE Full text] [doi: 10.2337/dc12-2360] [Medline: 23349548]
46. Feig DS, Murphy HR. Continuous glucose monitoring in pregnant women with Type 1 diabetes: benefits for mothers, using pumps or pens, and their babies. Diabet Med 2018 Apr;35(4):430-435. [doi: 10.1111/dme.13585] [Medline: 29352491]
47. Rys P, Ludwig-Slomczynska A, Cyganek K, Malecki M. Continuous subcutaneous insulin infusion vs multiple daily injections in pregnant women with type 1 diabetes mellitus: a systematic review and meta-analysis of randomised controlled trials and observational studies. Eur J Endocrinol 2018;178(5):545-563. [doi: 10.1530/eje-17-0804]
48. Skar JB, Garnweidner-Holme LM, Lukasse M, Terragni L. Women's experiences with using a smartphone app (the Pregnant+ app) to manage gestational diabetes mellitus in a randomised controlled trial. Midwifery 2018 Mar;58:102-108. [doi: 10.1016/j.midw.2017.12.021] [Medline: 29329023]
49. Chen Q, Carbone ET. Functionality, implementation, impact, and the role of health literacy in mobile phone apps for gestational diabetes: scoping review. JMIR Diabetes 2017 Oct 04;2(2):e25 [FREE Full text] [doi: 10.2196/diabetes.8045] [Medline: 30291088]

50. Yu Q, Aris I, Tan KH, Li LH. Application and utility of continuous glucose monitoring in pregnancy: a systematic review. Front Endocrinol (Lausanne) 2019 Oct 11;10:697.

Abbreviations

CG: control group
EPHPP: Effective Public Health Practice Project
HbA_{1c}: hemoglobin A_{1c}
IG: intervention group
MQ: moderate quality
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
SQ: strong quality
WQ: weak quality

Edited by R Kukafka; submitted 13.10.20; peer-reviewed by R Halkes, R De Carvalho, M Lozano-Lozano; comments to author 29.11.20; revised version received 08.12.20; accepted 16.03.21; published 28.04.21.

Please cite as:
Eberle C, Loehnert M, Stichling S
Clinical Effectiveness of Different Technologies for Diabetes in Pregnancy: Systematic Literature Review
J Med Internet Res 2021;23(4):e24982
URL: https://www.jmir.org/2021/4/e24982
doi:10.2196/24982
PMID:33908894

Review

eHealth for Addressing Balance Disorders in the Elderly: Systematic Review

Andréa G Martins Gaspar[1,2*], MD; Luís Velez Lapão[2*], PhD

[1]Hospital Beatriz Ângelo, Lisbon, Portugal

[2]Global Health and Tropical Medicine, Instituto de Higiene e Medicina Tropical, Universidade NOVA de Lisboa, Lisbon, Portugal

[*]all authors contributed equally

Corresponding Author:
Andréa G Martins Gaspar, MD
Global Health and Tropical Medicine
Instituto de Higiene e Medicina Tropical
Universidade NOVA de Lisboa
Rua da Junqueira, 100
Lisbon, 1349-008
Portugal
Phone: 351 213 652 600
Email: andreamartinsbr@hotmail.com

Abstract

Background: The population is aging on a global scale, triggering vulnerability for chronic multimorbidity, balance disorders, and falls. Falls with injuries are the main cause of accidental death in the elderly population, representing a relevant public health problem. Balance disorder is a major risk factor for falling and represents one of the most frequent reasons for health care demand. The use of information and communication technologies to support distance healthcare (eHealth) represents an opportunity to improve the access and quality of health care services for the elderly. In recent years, several studies have addressed the potential of eHealth devices to assess the balance and risk of falling of elderly people. Remote rehabilitation has also been explored. However, the clinical applicability of these digital solutions for elderly people with balance disorders remains to be studied.

Objective: The aim of this review was to guide the clinical applicability of eHealth devices in providing the screening, assessment, and treatment of elderly people with balance disorders, but without neurological disease.

Methods: A systematic review was performed in accordance with the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analysis) statement. Data were obtained through searching the PubMed, Google Scholar, Embase, and SciELO databases. Only randomized controlled trials (RCTs) or quasiexperimental studies (QESs) published between January 2015 and December 2019 were included. The quality of the evidence to respond to the research question was assessed using Joanna Briggs Institute (JBI) Critical Appraisal for RCTs and the JBI Critical Appraisal Checklist for QESs. RCTs were assessed using the Cochrane risk of bias tool. We provide a narrative synthesis of the main outcomes from the included studies.

Results: Among 1030 unduplicated articles retrieved, 21 articles were included in this review. Twelve studies explored different technology devices to obtain data about balance and risk of falling. Nine studies focused on different types of balance exercise training. A wide range of clinical tests, functional scales, classifications of faller participants, sensor-based tasks, intervention protocols, and follow-up times were used. Only one study described the clinical conditions of the participants. Instrumental tests of the inner ear were neither used as the gold-standard test nor performed in pre and postrehabilitation assessments.

Conclusions: eHealth has potential for providing additional health care to elderly people with balance disorder and risk of falling. In the included literature, the heterogeneity of populations under study, methodologies, eHealth devices, and time of follow-up did not allow for clear comparison to guide proper clinical applicability. This suggests that more rigorous studies are needed.

(J Med Internet Res 2021;23(4):e22215) doi:10.2196/22215

KEYWORDS

balance disorders; falls; elderly; eHealth; telemedicine

Introduction

Background

Aging and Balance Disorders

The improvement of health conditions and the increase in life expectancy have led to an aging global population, although this is not always accompanied by an increase in healthy life years [1-4].

Aging is associated with functional deterioration, including in the peripheral sensory structures, thereby affecting vision, hearing, and balance [5,6]. Additionally, elderly individuals are more likely to suffer from multiple chronic conditions, which often leads to frailty with risk of falls [1-4]. Falls in elderly people represent a serious public health problem as the main cause of accidental death in this population. The risk of falling increases with age [1,7-9]. Each year, approximately one in every three elderly people experiences a serious fall. Moreover, falling can lead to deterioration of the quality of life, anxiety, depression, restriction in daily activities, decreased mobility, social isolation, increased consumption of medications, and increased dependence on medical services and informal caregivers [1,2].

Several causes of falls in the elderly population have been identified, including age, environmental factors (eg, wet paths), inappropriate clothing and shoes, incorrect behavior (eg, climbing chairs), excessive alcohol consumption, inadequate use of medications, deteriorating chronic illness, and balance disorders [1,6,10].

Various clinical conditions are associated with balance disorders in elderly people, including age-related decline in balance function (prebyvestibulopathy); medications; and cardiovascular, metabolic, musculoskeletal, neurologic, and otologic diseases [5,6].

Although dizziness and vertigo are recognized as significant factors increasing the risk of falling and are common symptoms among the elderly, epidemiological studies have revealed large variability in the prevalence of balance disorders in this population [11-13]. It is estimated that at least 30% of individuals above 60 years old suffer from vertigo and dizziness, increasing to 50% for those above 85 years old [13]. According to the 2008 National Health Interview Survey, 33 million US adults had balance disorders, 26% of whom were elderly people (above 65 years) [14]. Approximately 20% of elderly people in the United States have a balance disorder event annually [15]. In fact, dizziness is a common complaint among the elderly population and is a strong predictor of falling events with a negative impact on quality of life [16]. Poor balance is frequently associated with falling [17,18]. In particular, asymmetrical vestibular function may often contribute to falls and fractures in elderly people [19-21].

Balance disorders and consequent falls have progressively represented a burden of disease, accompanied by high costs and pressure on the social services and health care systems related to medical care. This includes repeated consultations, excessive use of diagnostic imaging, and emergency care [22-24]. For example, the first national study in the context of dizziness and vertigo in the Emergency Services of United States of America for 2011 revealed that 25.7% of patient complaints of dizziness and vertigo were associated with balance disorders. The cost was estimated at about US $768 per episode, translating to an annual national cost of US $757 million. In the same context, cardiovascular diseases (linked to 16.5% of these episodes) represented a cost of approximately US $1489 per episode for an annual cost of US $941 million. By comparison, cerebrovascular diseases only accounted for 3.1% of these episodes, but with a cost per episode of approximately US $1059 or an annual cost of US $127 million. With the progressive aging of the population, worsening of this situation is expected in the future [25]. Indeed, vertigo is already contributing to the increasing trend of health care costs, which is linked to the aging of the population [23,24].

In this scenario of global aging, the use of digital solutions has been encouraged. Moreover, the additional pressure of the current COVID-19 pandemic has motivated the broader use of eHealth technologies [26].

Digital Health Care and the Elderly

The aging trend represents a relevant challenge to both patients and their families, and to the sustainability of health care systems globally. This is linked to the goal of global health policies for achieving a more active and healthy aging society with autonomy and independence [27,28]. The provision of new health care models, including eHealth services, has been encouraged to tackle access inequities, optimize health outcomes, and ensure autonomy and social support for elderly people. The use of eHealth seems to decrease costs associated with both institutionalization and unnecessary hospital visits [27-29].

eHealth consists of the use of information and communication technologies (ICTs) to support a health care communication channel at a distance, allowing for more efficient delivery of care services with optimized resource allocation. eHealth often contributes to improving the quality of health care services, including faster access to health information, promotion of the globalization of health care, and better health outcomes [30]. The World Health Organization has also recommended eHealth to promote universal health coverage, envisaging higher health care services availability with fewer resources and larger patient interaction. To date, eHealth has been used in the management of many conditions from health literacy promotion to teleconsultations [31]. The remote access systems can actively monitor elderly people in a real-life environment, leveraging the fact that there is an increasing interest and engagement of the elderly with technology. Moreover, eHealth technologies can enhance medical-patient interactions and mitigate many care access inequities. However, digital training of elderly people and caregivers is essential [4,32-34].

Assessment and Rehabilitation for the Elderly with Balance Disorders

There are several clinical tests and functional scales, including the Timed Up and Go Test (TUGT), Unipedal Standing Test, and Berg Balance Scale, that allow for assessments of balance,

gait, and risk of falling [5,35]. The use of sensors can improve the data quality of these tests and scales [36,37]. Additionally, functional tests of the inner ear, such as videonystagmography or the Video Head Impulse Test, are essential to identify and measure balance disorder cases, including an age-related decline in balance function (prebyvestibulopathy) [6].

Personalized balance training is a relevant option for the treatment of elderly people with balance disorders and risk of falling [5]. This training consists of an exercise-based program to address an individual's specific balance disorder, with goals of increasing postural stability, improving activities of daily living, and decreasing symptoms. Balance training should be focused on the functional deficiencies identified. Therefore, a prior medical evaluation is necessary to identify the clinical conditions related to poor balance as mentioned previously [38-41]. Moreover, these clinical conditions can affect the outcomes. For example, intervention success is more difficult when the patient has a disorder of both inner ears or has limited mobility due to an osteoarticular disease [38-41]. Exercises delivered through video games can be a promising intervention to achieve greater access and adherence among elderly people [42,43].

Several reviews have addressed the potential of digital solutions to improve the clinical observation and evaluation of balance disorders, and to promote the remote balance rehabilitation of elderly people [36,37,42-47]. However, most of these reviews included studies using a younger population as a preliminary assessment [42,44-47], and the majority did not describe the clinical conditions of the participants that might interfere with the outcomes, especially in the context of balance rehabilitation. Additionally, the clinical applicability of these devices was not assessed [36,37,42-47].

Therefore, there is a gap in this field in terms of evaluating the overall applicability of digital solutions according to the clinical conditions of elderly people with balance disorders and without neurological disease.

Objectives

The aim of this review was to evaluate and guide the clinical applicability of eHealth devices in the screening, assessment, and treatment of elderly people with balance disorders but without neurological disease.

Methods

Design

This systematic review was performed in accordance with the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analysis) statement [48] with the following steps: development of research questions, development of a search strategy with eligibility criteria, data selection, and qualitative analysis.

The protocol for this systematic review was registered in the International Prospective Register of Systematic Reviews (PROSPERO; CRD42019120774) and the complete protocol is available on the National Institute for Health Research program website.

This review focused on answering the following specific research questions, according to the PICO (Population, patient, or problem; Intervention; Control, Comparison, or Comparator; Outcome) strategy [49] (Textbox 1): (1) What are the main contributions of eHealth to elderly people with balance disorders with risk of falling? and (2) Is there any evidence that eHealth improves the quality of health care services in this context? If not, what are the reasons?

Textbox 1. Description of the PICO components.

- P (Population, Patient, Problem): Elderly people (over 60 years old) with balance disorders and risk of falling; studies with elderly people with functional limitation by neurological disease were excluded
- I (Intervention): eHealth devices for remote health education, screening, assessment, monitoring, or rehabilitation of elderly people with balance disorders with risk of falling
- C (Control, Comparison, Comparator): No intervention, paper booklet information, clinical evaluation, conservative balance training
- O (Outcome): Clinical applicability, increased fall prevention literacy, early identification and evaluation of balance deficits and risk of falling, improved balance and gait performance, reduced rate of falling, increased rehabilitation adherence, increased independence in daily activities

Definition of Concepts and Keywords used in the Search Strategy

In this study, we defined elderly people as those over 60 years of age [50]. Knudson [51] defined balance as a "person's ability to control their body position relative to some base of support." According to Agrawal et al [6], vertigo and dizziness are defined as "sensation of self-motion when no self-motion is occurring or the sensation of distorted self-motion during an otherwise normal head movement" and "sensation of disturbed or impaired spatial orientation without a false or distorted sense of motion," respectively. Falls refer to "inadvertently landing on the ground, floor or other lower level" [10]. Gait is defined as "the pattern of movement of the body during locomotion" [52].

Telemedicine is defined according to the World Health Organization Group Consultation on Health Telematics [53] as "delivery of health care services using ICT for the exchange of valid information for diagnosis, treatment and prevention of disease and injuries, research and evaluation, and for the continuing education of health care providers." eHealth is defined according to Eysenbach [54] as:

> *an emerging field in the intersection of medical informatics, public health and business, referring to health services and information delivered or enhanced through the Internet and related technologies. In a broader sense, the term characterizes not only a technical development, but also a state-of-mind, a*

> *way of thinking, an attitude, and a commitment for networked, global thinking, to improve health care locally, regionally, and worldwide by using information and communication technology.*

Teleconsultation is defined as "synchronous or asynchronous consultation using ICT to omit geographical and functional distance" [55]. Finally, a sensor is defined as a "device that responds to a physical input of interest with a recordable functionally related output that is usually electrical or optical" [56].

Search Strategy

Articles were retrieved through searching the PubMed, Google Scholar, Embase, and SciELO databases. The search algorithm included multiple group combinations, as shown in Table 1.

Table 1. Search strategy.

Concept	Keywords
Elderly people	("elderly" OR "older" OR "aged")
Balance	("Balance" OR "balance disorder" OR "balance problem" OR "vertigo" OR "dizziness) and/or ("falls" OR "fall detection" OR "fall prevention") and/or ("gait")
Telemedicine	("Telemedicine" OR "eHealth" OR "teleconsultation" OR "technology" OR "sensor")

Selection Criteria

The inclusion criteria were randomized controlled trials (RCTs) or quasiexperimental studies (QESs) published in English between January 2015 and December 2019, studies related to use of eHealth in the context of balance and falls, and the sample was restricted to an elderly population (60 years old and above).

The exclusion criteria were: (1) review articles, brief reports, protocols, proof-of-concepts, pilot studies, conference papers, and letters to the editor; (2) studies including elderly people with a reported functional limitation due to a neurological disease; and (3) articles without an age sample reference or with participants aged below 60 years.

Screening Process and Data Extraction

First, both authors screened the papers independently, looking at titles, abstracts, and methods, and agreed about their inclusion or exclusion according to the eligibility criteria. Second, the potentially relevant papers were retrieved for full-text evaluation against the eligibility criteria. Any articles that were deemed to be questionable in the first stage were included for further evaluation in the second stage. The selection of papers was performed by checking the extracted data and risk of bias.

Outcome Measures

The main outcomes included population characteristics, balance disorder, identification of faller participants, eHealth platform and services, health benefits, and fall prevention literacy.

Risk of Bias Assessment

The quality of the evidence to respond to the research questions was independently assessed using the Joanna Briggs Institute (JBI) Critical Appraisal for Experimental Studies and JBI Critical Appraisal Checklist for Quasi-Experimental Studies tools [57]. The two researchers discussed the results of the quality appraisal, reaching a consensus in case of any divergence. The included RCTs were assessed using the Cochrane risk of bias tool [58] to evaluate the risk of internal bias for a series of domains: selection bias, performance bias, detection bias, attrition bias, and reporting bias. Disagreements were solved by consensus between the two researchers.

Data Analysis

We provide a narrative synthesis of the main outcomes from the included studies. First, the articles were categorized according to the study design. Second, the articles were categorized based on the focus of eHealth services (screening/assessment and treatment/rehabilitation) for comparison of clinical use and applicability according to digital devices.

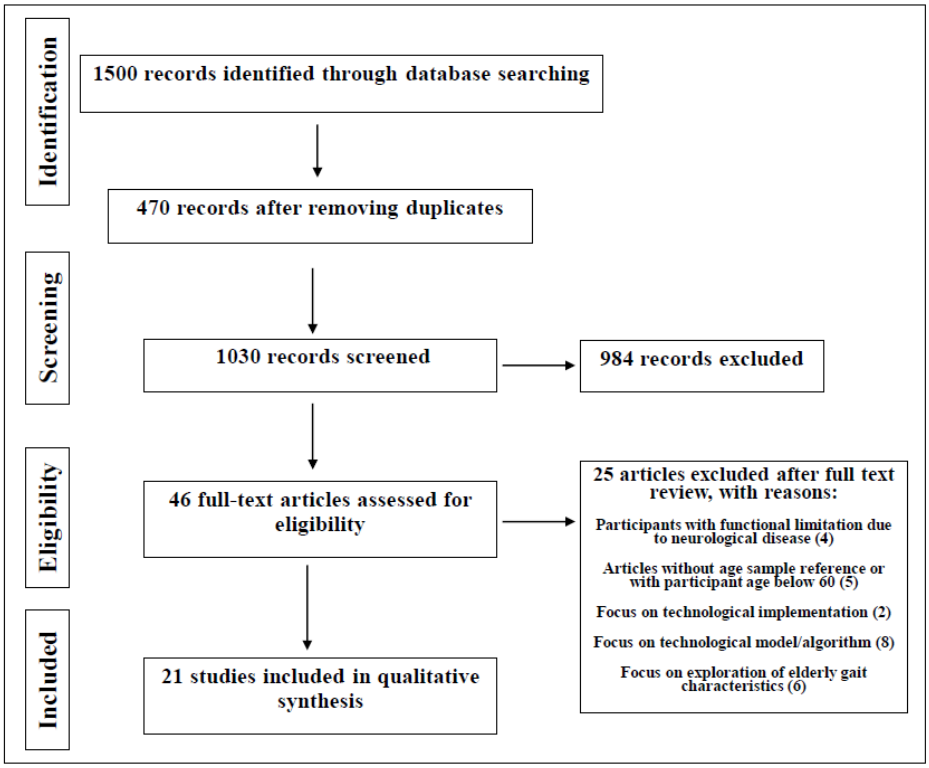
Results

Search Results

A total of 1030 unduplicated articles were identified, 984 of which were excluded after title and abstract screening. Among the 46 full-text publications assessed for eligibility, 25 articles were excluded owing to functional limitations due to neurological disease (n=4), age of participants (n=5), focus on technological implementation (n=2) or model/algorithm (n=8), and specific descriptions of elderly gait parameters (n=6).

Twenty-one articles [59-79] were ultimately included in the review (Figure 1).

Figure 1. Flow of selection for studies included and excluded in the review.



Study Design

RCT Design

Seven studies were RCTs [59-65], including one cross-over trial without a washout term [64] and one multicenter study [60].

Not all authors clearly described the randomization process [62] and the allocation concealment [62,64] (Tables 2 and 3).

The inclusion criteria were mentioned in all articles. However, the clinical conditions of the participants were only described in one study [60]. The function of the inner ear was never mentioned. Thus, the expected similarity between the control and intervention groups was not clear. This is relevant because various clinical conditions (eg, cardiovascular, metabolic, inner ear disease, medication) can interfere with the outcomes of balance rehabilitation [38-41]. Therefore, the lack of information about clinical conditions of the participants, including the lack of data about function of the inner ear, was considered as "other bias" and was a common weakness of all included RCTs (Table 3 and Figure 2). This approach led to a worse classification of the quality of these studies.

Additionally, the blinding of participants, personnel, and outcome assessment were unclear in some of these studies.

In the control and intervention groups of all RCTs, a few dropouts for medical and personal reasons were mentioned. However, this was not considered to be sufficiently relevant to have an impact on the results. Only two papers reported intention-to-treat analysis [60,63].

All outcomes were measured in a reliable manner and were considered to have been properly analyzed.

Table 2. Methodological quality of randomized controlled trials based on the Joanna Briggs Institute Critical Appraisal Checklist.

Study	Q1[a]	Q2[b]	Q3[c]	Q4[d]	Q5[e]	Q 6[f]	Q7[g]	Q8[h]	Q9[i]	Q10[j]	Q11[k]	Q12[l]	Q13[m]
Eggenberger et al [59]	Y[n]	Y	U[o]	Y	N[p]	U	Y	Y	N	Y	Y	Y	Y
Gschwind et al [60]	Y	Y	U	Y	Y	U	Y	Y	Y	Y	Y	Y	Y
Gschwind et al [61]	Y	Y	U	U	U	Y	Y	Y	N	Y	Y	Y	Y
Lim et al [62]	U	U	U	U	U	U	Y	Y	N	Y	Y	Y	Y
Oesch et al [63]	Y	Y	U	N	N	Y	Y	Y	Y	Y	Y	Y	Y
Ozaki et al [64]	Y	U	U	U	U	U	Y	Y	N	Y	Y	Y	Y
Hong et al [65]	Y	Y	U	Y	Y	Y	Y	Y	N	Y	Y	Y	Y

[a]Question 1: Was true randomization used for assignment of participants to treatment groups?

[b]Question 2: Was allocation to treatment groups concealed?

[c]Question 3: Were treatment groups similar at the baseline?

[d]Question 4: Were participants blind to treatment assignment?

[e]Question 5: Were those delivering treatment blind to treatment assignment?

[f]Question 6: Were outcomes assessors blind to treatment assignment?

[g]Question 7: Were treatment groups treated identically other than the intervention of interest?

[h]Question 8: Was follow-up complete and if not, were differences between groups in terms of their follow-up adequately described and analyzed?

[i]Question 9: Were participants analyzed in the groups to which they were randomized?

[j]Question 10: Were outcomes measured in the same way for treatment groups?

[k]Question 11: Were outcomes measured in a reliable way?

[l]Question 12: Was appropriate statistical analysis used?

[m]Question 13: Was the trial design appropriate, and any deviations from the standard randomized controlled trial design (individual randomization, parallel groups) accounted for in the conduct and analysis of the trial?

[n]Y: Yes.

[o]N: No.

[p]U: Unclear.

Table 3. Risk of bias for randomized controlled trials based on the modified Cochrane Collaboration tool.

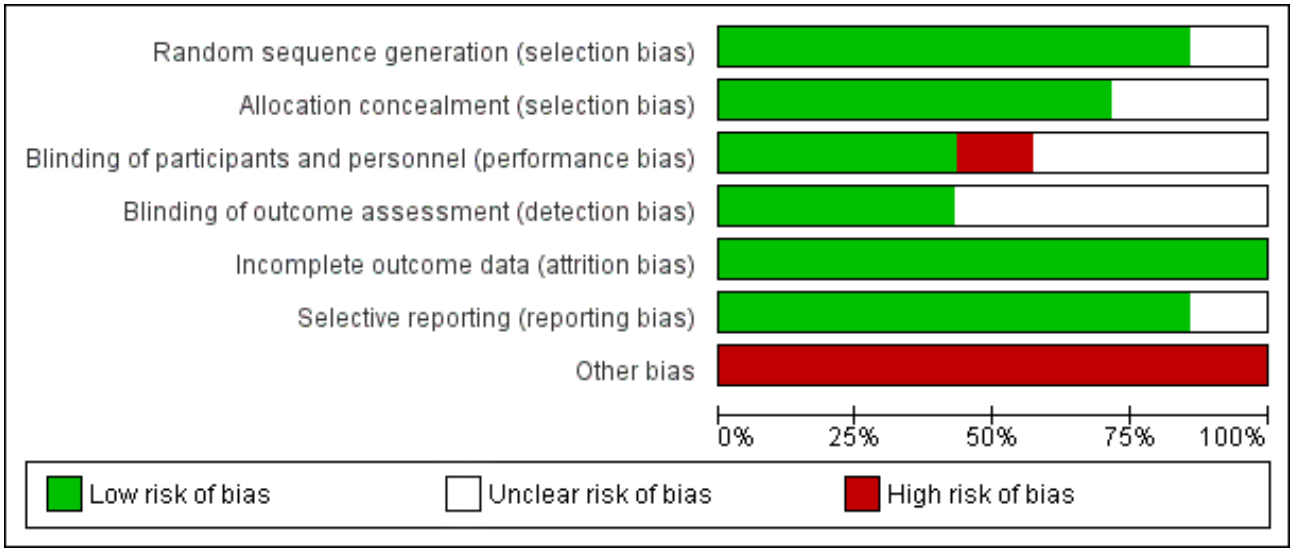
Study	Selection bias		Other bias	Reporting bias: selective reporting	Performance bias: blinding (participants and personnel)	Detection bias: blinding (outcome assessment)	Attrition bias: incomplete outcome data
	Random sequence generation	Allocation concealment					
Eggenberger et al [59]	L[a]	L	H[b]	U[c]	L	U	L
Gschwind et al [60]	L	L	H	L	L	U	L
Gschwind et al [61]	L	L	H	L	U	L	L
Lim et al [62]	U	U	H	L	U	U	L
Oesch et al [63]	L	L	H	L	H	L	L
Ozaki et al [64]	L	U	H	L	U	U	L
Hong et al [65]	L	L	H	L	L	L	L

[a]L: low risk.

[b]H: high risk.

[c]U: unclear risk.

Figure 2. Risk of bias in accordance with the authors' judgment (RevMan version 5.3.).



QES Design

Fourteen studies were QESs [66-79]; only one of these was a multicenter study [73]. Twelve of these studies used the same group of participants [66,68-77,79]. One study used two groups with different participants [67] and another had a control group and an intervention group [78]. However, the expected similarity between the groups was not clear because there was no description of the clinical conditions of the participants, including function of the inner ear. Loss to follow-up was also not mentioned for any of these studies (Table 4).

Table 4. Methodological quality of quasiexperimental studies based on the Joanna Briggs Institute Critical Appraisal Checklist.

Study	Q1[a]	Q2[b]	Q3[c]	Q4[d]	Q5[e]	Q6[f]	Q7[g]	Q8[h]	Q9[i]
Zacaria et al [66]	Y[j]	Y	Y	N[k]	Y	Y	Y	Y	Y
Hall et al [67]	Y	U[l]	Y	N	Y	Y	Y	Y	Y
Howcroft et al [68]	Y	Y	Y	N	Y	Y	Y	Y	Y
Lee et al [69]	Y	Y	Y	N	Y	Y	Y	Y	Y
Ponti et al [70]	Y	Y	Y	N	Y	Y	Y	Y	Y
Similä et al [71]	Y	Y	Y	N	Y	Y	Y	Y	Y
Shahzad et al [72]	Y	Y	Y	N	Y	Y	Y	Y	Y
Brodie et al [73]	Y	Y	Y	N	Y	Y	Y	Y	Y
Howcroft et al [74]	Y	Y	Y	N	Y	Y	Y	Y	Y
Chigateri et al [75]	Y	Y	Y	N	Y	Y	Y	Y	Y
Qiu et al [76]	Y	Y	Y	N	Y	Y	Y	Y	Y
Hiesh et al [77]	Y	Y	Y	N	Y	Y	Y	Y	Y
Maneproom et al [78]	Y	U	Y	Y	Y	Y	Y	Y	Y
Nightingale et al [79]	Y	Y	Y	N	Y	Y	Y	Y	Y

[a]Question 1: Is it clear in the study what is the "cause" and what is the "effect" (ie, there is no confusion about which variable comes first)?

[b]Question 2: Were the participants included in any similar comparisons?

[c]Question 3: Were the participants included in any comparisons receiving similar treatment/care other than the exposure or intervention of interest?

[d]Question 4: Was there a control group?

[e]Question 5: Were there multiple measurements of the outcome, both pre and post the intervention/exposure?

[f]Question 6: Was follow-up completed and if not, were differences between groups in terms of their follow-up adequately described and analyzed?

[g]Question 7: Were the outcomes of participants included in any comparisons measured in the same way?

[h]Question 8: Were outcomes measured in a reliable way?

[i]Question 9: Was appropriate statistical analysis used?

[j]Y: yes.

[k]N: no.

[l]U: unclear.

Multiple different measurements of the outcomes were used (Table 4). However, the instrumental inner ear tests were not used as the gold-standard test. This lack of comparison was considered to be a weakness of all of the included QESs. The most commonly applied tests were the TUGT and walking over different distances. One study assessed 1-week daily-life walking [73]. Only one study explored the activities of daily living [75] (Table 5).

The outcomes were considered to have been measured in a reliable manner and were properly analyzed.

Table 5. Quasiexperimental studies focused on screening/assessment.

Reference	Population and setting	Participants, N (male/female)	Technology/sensor used (location)	Completed sensor-based procedures	Comparison with fall history or a gold-standard test
Zacaria et al [66]	Hospital	38 (20/18)	WIS[a] accelerometer and gyrosensor (L2 vertebra)	TUGT[b] (sensor analysis of performance of each phase)	Classification of faller/nonfaller based on total duration for TUGT completion
Howcroft et al [68]	CDEP[c]	100 (44/56)	Wearable pressure-sensing insoles (pressure sensors-plantar); WIS, 4 triaxial accelerometers (head, pelvis, shanks)	7.62 m walk and 7.62 m walk with cognitive load (ST[d] and DT[e] gait)	Classification of faller/nonfaller based on retrospective fall occurrence
Lee et al [69]	CDEP	65 (16/49)	WIS (1 triaxial accelerometer belt around waist, pelvis, sacrum, L3-L5 vertebrae)	TUGT	Short-form BBS[f] (7 activities), TUGT
Ponti et al [70]	CDEP	36 (11/25)	WIS (1 triaxial accelerometer, waist)	ST TUGT, DT manual TUGT, DT cognitive TUGT	Faller/nonfaller based on retrospective fall occurrence, FES[g]
Similä et al [71]	Senior house/senior physical exercise group	35 (0/35)	WIS (2 accelerometers L3-L5 vertebrae, right hip)	BBS + TUGT + 4 m walk and follow-up after 1 year	Background questionnaire, interview, balance platform assessment with Kinect recording
Shahzad et al [72]	CDEP	23 (7/16)	WIS (1 triaxial accelerometer, L3-L5 vertebrae)	TUGT, STS-5[h], AST[i]	BBS
Brodie et al [73]	CDEP	96 (39/57)	Pendant sensor 3D accelerometer and barometer (sternum)	1-week daily life walking	Fall history, comparison with TUGT and 10 m walk test
Howcroft et al [74]	CDEP	75 (31/44)	Wearable pressure-sensing insoles (pressure sensor, plantar), WIS 4 triaxial accelerometers (head, pelvis, shanks)	7.62 m walk under ST, 7.62 m walk under DT (verbal-task cognitive load)	Classification of faller/ nonfaller based on prospective fall occurrence
Chigateri et al [75]	Frail elderly people from independent-living retirement homes	23 (6/17)	WIS 1 triaxial accelerometer (L5 vertebra)	TUGT, STS[j], activities of daily living	Synchronized videos with accelerometer (identification of the beginning of TUGT and of walking episode)
Qiu et al [76]	CDEP/social welfare centers	196 (0/196)	WIS 5 sensors with 3-axis-acceleration, 3-axis angular velocity, 3-axis magnetism each (low back, upper legs, lower legs)	Sensory integration test, limits of stability forward reach, STS-5, TUGT, motor function	Classification of faller based on self-reported history
Hiesh et al [77]	Healthy elderly people	30 (12/18)	Smartphone technology, 1 accelerometer (sternum)	Balance tests standing on a force plate and holding a smartphone against the chest: eyes open/closed DT, semitandem, tandem stance, single-leg stance	Comparison between force plate and smartphone data
Nightingale et al [79]	Local community centers/health care provider offices/senior centers	51 (unknown)	OptoGait system photoelectric technology	10 m walk	TUGT

[a]WIS: wearable inertial sensor.

[b]TUGT: Timed Up and Go Test.

[c]CDEP: community-dwelling elderly people.

[d]ST: single task.
[e]DT: dual task.
[f]BBS: Berg Balance Scale.
[g]FES: Fall Efficacy Scale.
[h]STS-5: Five Times Sit-to-Stand test.
[i]AST: Alternative Step Test.
[j]STS: Sit-to-Stand test.

Focus of eHealth Services

The 12 QESs were focused on screening and assessment. These studies compared the use of sensors in participants with a history of falling, or with clinical tests and functional scales [66,68-77,79] (Table 5). No instrumental test of the inner ear was performed as a gold-standard test. All RCTs [59-65] and two QESs [67,78] were focused on balance treatment or rehabilitation (Table 6). Again, no instrumental test of the inner ear was used in pre and postrehabilitation assessments.

Table 6. Studies focused on treatment/rehabilitation.

Reference	Study type	Setting and population	Participants, N (male/female)	Tested technology	Tested sensor architecture	Sensor-based procedures	Outcome measurements
Eggenberger et al [59]	RCT[a]	CDEP[b] and retirement homes	71 (25/46): dance group n=24, memory group n=22, control group n=25	VR[c] video game dancing + Impact Dance Platform treadmill walking + computer screen + training software	VR video game dancing + pressure sensitive platform, treadmill + training software	VR video game dancing with simultaneous cognitive-physical training; treadmill walking with simultaneous verbal memory training; treadmill walking (control)	Gait analysis: ST[d]/DT[e] 7.3 m walking, Short Physical Performance Battery, fall frequency, 6-minute walk test, measure of fall fear
Gschwind et al [60]	RCT	CDEP	153 (60/93): intervention group n=78, control group n=75	iStoppFalls system: computer + Google TV set top box + Microsoft Kinect + senior mobility monitor + android tablet	Kinect-based system (3D depth sensor), Senior Mobility Monitor (3D accelerometer, barometer)	16-week home-based balance exergames and muscle strength exercises + education booklet (intervention) or education booklet + usual activities (control)	Estimated risk of falling; mobility, self-care, usual activities, pain, discomfort, anxiety, depression, health questionnaires, cognitive performance, walking task, STS-5[f], TUGT[g], technology use
Gschwind et al [61]	RCT	CDEP	124 (42/82): step-mat training group n=39, Microsoft-Kinect group n=24, control group n=61	Input device, computer, USB modem, TV, exergames, Microsoft Kinect or electronic mat	Pressure-sensitive electronic mat, Kinect-based system (3D depth sensor)	Unsupervised 16-week home exercise using exergames or educational booklet about evidence-based health and fall prevention advice + usual activities (control)	Risk of falling, health and disability measure, STS-5, TUGT, cognitive performance
Hall et al [67]	QES[h]	CDEP	16 (0/16): group A n=8, group B n=8	Nintendo Wii Fit System: computer interface + monitor + Wii balance board + games Ski Slalom/ Table Tilt	*Balance* Board: force platform	Wii Fit balance test + games, followed by SOT[i] and LOS[j] test, CDP[k] (group A); SOT and LOS test, CDP, followed by Wii Fit balance test + games (group B)	Dynamic Gait Index, TUGT, gait speed
Lim et al [62]	RCT	CDEP	36 (11/25): intervention group n=18, control group n=18.	Wearable balance biofeedback (system (vibrotactile, auditory and visual biofeedback)	Biofeedback headband: 8 vibrotactile actuators, 2 bone-conducting acoustic transducers, 3 light-emitting diodes; and gyroscopes (lower back)	2-week training with real-time multimodal biofeedback of trunk sway or 2-week training without biofeedback (control)	Standing:1 leg (eyes open), feet together, firm surface and foam, tandem stance (EC[l]); self-paced 8 m walking (EC); 8 m walking with head turning; 8 tandem steps (EC)
Oesch et al [63]	RCT	Geriatric rehabilitation center	54 (29/25): intervention group n=26, control group n=28	Windows Kinect (exergame)	Kinect-based system (3D depth sensor)	10-day self-regulated training with exergames or 10-day self-regulated conventional training with instruction leaflets (control)	Adherence, motivation, enjoyment, sensor-based walking test

Reference	Study type	Setting and population	Participants, N (male/female)	Tested technology	Tested sensor architecture	Sensor-based procedures	Outcome measurements
Ozaki et al [64]	RCT	Prefrail or frail CDEP	27(7/20): intervention group n=14, control group n=13	BEAR[m] system: Stand-and-ride transport robot + wearable helmet and suspending device + software	Stand-and-ride transport robot with two inverted wheel motors	6-week based BEAR training first group or 6-week based conventional balance training first group (control)	Gait speed, tandem gait speed, functional reach test, TUGT
Hong et al [65]	RCT	CDEP	23 (0/23): intervention group n=10, control group n=13	tablet, web app, signaling server module network address translator traversal module	Web Real-Time Communication (WebRTC) technology	12-week telepresence exercise sessions or maintained lifestyle (control)	Senior fitness test, BBS[n], fall-related self-efficacy, FES[o], Fear of falling questionnaire
Maneproom et al [78]	QES	Senior housing	64 (13/51): intervention group n=32, control group n=32	robot, robot-installed fall prevention software	8-inch touchscreen installed at robot head	robot-installed fall prevention software + personal coaching + fall prevention handbook or fall prevention handbook only (control)	TUGT, BBS, fall prevention questionnaire

[a]RCT: randomized controlled trial.

[b]CDEP: community-dwelling elderly people.

[c]VR: virtual reality.

[d]ST: single task.

[e]DT: dual task.

[f]STS-5: Five Times Sit-to-Stand test.

[g]TUGT: Timed Up and Go Test

[h]QES: quasiexperimental study.

[i]SOT: Sensory Organization Test.

[j]LOS: Limits of Stability.

[k]CDP: computerized dynamic posturography.

[l]EC: eyes closed.

[m]BEAR: Balance Exercise Assist Robot.

[n]BBS: Berg Balance Scale

[o]FES: Fall Efficacy Scale.

Population Characteristics

The included studies had large differences in sample sizes, ranging from 23 to 153 participants [59-65] among RCTs and from 16 to 196 participants [66-79] among QESs. As shown in Tables 5 and 6, many studies included a small sample size that was described as a limitation.

The age range was 60-91 years for the RCTs and 60-92 years for the QESs. Most of the studies included more women than men. In four studies, only women participated [65,67,71,76]. The decision to only recruit women was explained in one study as “to avoid the influence of gender differences on risk of falling” [76]. Two studies excluded the few male participants [67,71] and the remaining article did not describe the reason for the exclusive participation of women [65]. One study did not describe the age range or the gender distribution of the participants [79].

The participants (≥60 years old) were recruited from the community [60-62,64,65,67-70,72-74,76,77], gerontology services [71,75,78], both [59,79], or at a hospital [66]. One study included participants who were referred for geriatric inpatient rehabilitation [63].

Most of the studies did not describe the characteristics of health conditions of the sample [59,62,64-66,68-72,74-77,79]. Only some authors provided quantitative data about the participants’ medication use [60,61,73,78] and their comorbidities [60,61,63,67,73,78]. One study [60] highlighted the following comorbidities of the participants: heart problems, high blood pressure, osteoporosis, lower back pain, hip pain, knee and/or leg pain, and foot pain. Two studies excluded participants with self-reported balance disorders [62,63]. Two other studies included frail or prefrail elderly adults [64,75].

Balance Disorder and Identification of Faller Participants

The included studies used functional balance tests, with or without sensors, to evaluate balance and risk of falling. An objective identification via exploration and quantification of the function of the inner ear by instrumental tests was not employed in any of the considered studies, as mentioned above.

Therefore, the presence of prebyvestibulopathy or other balance disorders was not known.

Some authors highlighted the potential of sensor-based tests in identifying early balance deficits [71] and in evaluating the risk of falling [66,72,73,76,77]. Improved balance and gait with technology-based training were mentioned in some studies [59,62,64].

The identification of faller participants based on retrospective [68,70,73,76] or prospective occurrence of falling [74] was employed to compare the technology results. The benefits of virtual training in reducing the risk of falling was also described [60,61,65,78].

No study focusing on detection of falling fully complied with the inclusion criteria of this review (RCTs or QESs, published in English between January 2015 and December 2019, restricted to the population 60 years or older).

eHealth Platform and Services

Different platforms were used for the provision of eHealth services. The main platforms identified were computer-based apps, either via the internet or mobile based platforms (Tables 5 and 6).

As mentioned above, 12 studies focused on screening or assessment [66,68-77,79] using different types (wearable inertial, wearable pressure, pendant, smartphone), quantities (range 1-5), and locations (head, sternum, lumbar vertebra, pelvis, hip, leg, shanks, foot) of sensors. Single or combined sensor-based tasks were employed. Only two studies [73,75] evaluated activities in a real-life environment (Table 5).

Nine studies explored balance rehabilitation [59-65,67,78] with different exercises and duration of training. The follow-up time was short (less than 6 months) in most of the studies, with the longest follow-up of 1 year [59]. The development of eHealth services was explored both inside and outside the laboratory environment (Table 6).

One study used a robot to provide information about training and fall prevention. However, the authors pointed out that the screen and the volume speaker were not adequate for use by elderly people [78].

The use of technical language and the presence of disabilities such as visual and hearing impairment were highlighted as the main barriers in using eHealth [78].

Health Benefits

Only one study did not report better adherence, enjoyment, motivation, and balance performance with virtual training. This was explained by the possible fragility of the sample included in the study and by the short duration of the training intervention [63].

The remaining papers emphasized the potential contribution of digital solutions to improve balance performance and risk of falling. The sensors used during balance tests improved the evaluation of balance and gait [66,68,69,71,72,79] and improved the identification of potential faller participants [70,73-77]. In addition, the use of eHealth devices for balance rehabilitation increased balance and gait performance [59,60,62,64,65,78], and reduced the risk of falling [60,61]. However, no long-term follow-up was reported. Virtual programs of falls prevention seemed to increase knowledge on the subject [78] (Table 7).

Table 7. Health benefits: conclusions from all studies.

Reference	Study type	Conclusions
Zacaria et al [66]	QES[a], screening	Single wearable sensor during TUGT[b]: an improved tool in evaluating fall risk
Howcroft et al [68]	QES, screening	Sensor-based gait assessment: potential of identification of gait changes
Lee et al [69]	QES, screening	Advantages of wearable sensor as an outside laboratory tool
Ponti et al [70]	QES, screening	Improved potential of identification of fallers with single sensor-based DT[c] TUGT
Similä et al [71]	QES, screening	Sensor-based walk test: a screening tool to identify early signs of balance deficits
Shahzad et al [72]	QES, screening	Importance of sensor-based TUGT, STS-5[d], and AST[e] on fall risk estimation
Brodie et al [73]	QES, multicenter screening	Better sensor-based daily-life gait assessment to discriminate fallers
Howcroft et al [74]	QES, screening	Sensor: potential to discriminate differences between ST[f] and DT gait and between prospective fallers and nonfallers
Chigateri et al [75]	QES, screening	Wearable accelerometer: useful for nonsedentary activity recognition and gait detection in frail older adults outside lab facilities
Qiu et al [76]	QES, screening	Potential use of wearable inertial sensor-based systems for elderly fall risk assessment
Hiesh et al [77]	QES, screening	Validity of smartphone for evaluation of postural stability and fall risk stratification in older adults
Nightingale et al [79]	QES, screening	Using Optogait system: TUGT as a tool for screening balance deficits
Eggenberger et al [59]	RCT[g], rehabilitation	Virtual reality game dancing with simultaneous cognitive-physical training and treadmill walking with simultaneous verbal memory training: potential to enhance gait variables
Gschwind et al [60]	RCT, rehabilitation	iStoppFalls program reduced physiological fall risk and improved postural sway
Gschwind et al [61]	RCT, rehabilitation	Step-mat-training and Microsoft-Kinect exergames reduced fall risks, Step-mat-training improved specific cognitive functions; neither intervention improved balance control
Hall et al [67]	QES, rehabilitation	WiiFit feasible to safely use, Ski Slalom game similar effect as computerized dynamic posturography
Lim et al [62]	RCT, rehabilitation	Balance training with biofeedback: most beneficial for the most difficult tasks but with few long-term benefits
Oesch et al [63]	RCT, rehabilitation	Superior results of conventional training with respect to adherence, enjoyment, and motivation; no difference of balance during walking between conventional and training with exergames
Ozaki et al [64]	Crossover trial without a washout term, rehabilitation	BEAR[h] training more effective for improving dynamic balance and lower extremity muscle strength
Hong et al [65]	RCT, rehabilitation	Telepresence exercise program: effective to improve balance and reduce fear of fall; no significant difference of fall efficacy between intervention (telepresence exercise sessions) and control group (maintained lifestyle)
Maneproom et al [78]	QES, rehabilitation	Robotic fall prevention program increased fall prevention knowledge, promoted exercises, and improved balance

[a]QES: quasiexperimental study.

[b]TUGT: Timed Up and Go Test.

[c]DT: dual task.

[d]STS-5: Five Times Sit-to-Stand test.

[e]AST: Alternative Step Test.

[f]ST: single task.

[g]RCT: randomized controlled trial.

[h]BEAR: Balance Exercise Assist Robot.

Fall Prevention Literacy

None of the studies explored the previous health literacy of the participants. Only two papers described the educational level of the participants [60,78].

One study compared use of a fall prevention software to a conventional handbook to evaluate the improvement of knowledge on fall prevention. Both the intervention and control groups showed improvement in knowledge, without a significant difference [78].

Discussion

Principal Findings

Population aging, and the associated vulnerability to the development of multiple chronic pathologies and balance disorders, have motivated research and the implementation of new strategies for the provision of health care. eHealth devices have been studied to help assess balance and gait performance, risk of falling in and outside a laboratory setting, and to perform in-home balance rehabilitation. In this review, we confirmed the potential of eHealth to complement the health care of elderly people. However, most of these studies were not designed to provide clinical guidelines.

Despite growing interest about this subject in the last 20 years, we decided to focus on studies published in the last 5 years (RCTs and QESs), taking into consideration both continuous advances in technological innovation and the opportunity to apply new clinical applications in balance disorder and risk of falling for the elderly population.

Unlike other reviews, our eligibility criteria ruled out many initially retrieved articles, especially studies with participants under 60 years old, those without reporting the age of participants, or with participants having a functional limitation due to neurologic disease. Therefore, only 21 articles fully complied with the requirements of this review [59-79].

Except for one study [63], the others showed the potential of eHealth to evaluate balance assessment and risk of falling of elderly people and to promote balance training. The eHealth devices allowed collecting additional information about the balance, gait, and risk of falling of elderly people, and to monitor their daily activities.

In particular, eHealth seems to provide an opportunity for increasing medical-patient interactions and to reduce access inequities [30]. In 1996, Viierre et al [80] had already mentioned the potential of eHealth in this field: “remote medical diagnosis and treatment facilities could make the few vestibular disorder specialists much more available to patients.” However, as observed in other reviews [36,37,42-47], the differences in methodologies and of variables included in the studies did not allow for a proper comparison to guide clinical applicability.

First, there was a broad range of sample sizes, which were generally quite small (ranging from 16 to 196 participants). A small sample of participants is considered a limitation for extrapolating the results, especially for the exploration of risk of falling.

Second, there were missing data about the clinical conditions of the participants. Except for one study [60], several volunteers were recruited from the community and were defined as “healthy” elderly people only based on a self-reported assessment. There were also participants recruited from geriatric services without reference to their clinical conditions. Despite the exclusion of participants with self-reported balance disorders in two studies [62,63], we consider that the exclusion rules should be more rigorous and based on objective data such as instrumental inner ear tests. We have to take into consideration that elderly people can have instability due to many conditions, including the normative aging process, and therefore the outcomes from a balance rehabilitation intervention could be sensitive to these differences [38-41].

Third, different research methodologies were used for screening and assessment. We observed a wide range of clinical tests, functional scales, faller classifications, and sensor-based tasks among the included studies. The lack of homogeneity of these variables limited an appropriate comparison among the studies. Moreover, functional inner ear tests were not used as the gold-standard test. We consider this as a weakness common to all studies.

Fourth, different types of sensors were used for screening and assessment. Similar to the findings of other reviews [36,37,44-47], the studies employed mainly accelerometers, with variations in both number and body location.

Fifth, as observed previously [42,43], studies focusing on treatment and rehabilitation used different devices, training durations, and follow-up times. Some authors employed supervised training. In one study, this was used a telepresence-based exercise platform [65]. Others employed in-home self-regulated exercises training [61], thereby avoiding the need for participants to travel to the rehabilitation center. None of the studies described pre and postintervention data about the function of the inner ear. The studies did not verify the long-term effect of training, especially with respect to fall occurrence. Only two studies explored a sensor used in real-life activities [73,75], which is relevant since it allowed for a better evaluation of the remote interaction and monitoring of daily activities.

Additionally, we observed a constraint related to the use of devices that are not fully adequate to match the abilities of elderly people [78]. We also highlight the importance of providing a better definition of the eHealth user profile to improve adherence.

Future studies in this field should consider the above topics as a starting point, as well as for health policy implementations on eHealth apps for elderly people with balance disorders.

The use of eHealth can play an important role as a complementary method to provide health care services, encouraging health promotion and patient participation, as well as allowing for the remote management of balance disorders.

Recommendations

Based on this review, we can provide the following recommendations to improve studies and applications of eHealth for preventing fall risk in the elderly population.

First, this review highlights the need for further research on the use of eHealth devices in proper clinical settings. This represents an opportunity to be explored, reaching out to elderly people with balance and risk of falling.

Second, despite several efforts to explore balance among the elderly, there is still a need for better characterization and description of the health condition of the population under study. In particular, we recommend future studies to include the results of functional tests of the inner ear as a gold-standard test or for comparison of the outcome before and after remote balance rehabilitation. Most of the interventions were developed with only functional balance tests. Future studies should also focus on the real-life environment, allowing for additional information of the daily activities among elderly participants.

Third, a longer follow-up time is important to evaluate the long-term benefits of eHealth tools on the balance performance and risk of falling of elderly people.

Finally, the eHealth devices should be user-friendly to improve adherence among elderly people.

Limitations

This review was limited to articles written in the English language and available on the PubMed, Google Scholar, Embase, and SciELO databases for the last 5 years; therefore, it is possible that relevant studies were missed.

Conclusions

The inclusion of eHealth services can play a critical role for the better provision of health care to elderly people with a balance disorder and risk of falling. The differences in populations, methodologies, eHealth devices, and follow-up times of the included studies did not allow for a clear comparison between results, therefore limiting the possibility of obtaining valid guidance for clinical applicability. More rigorous studies are recommended.

Acknowledgments

This work was partially supported by Fundação para a Ciência e a Tecnologia (FCT) for funds to Global Health and Tropical Medicine (GHTM) (UID/04413/2020 to LVL).

Conflicts of Interest

None declared.

References

1. Action plan on specific action on innovation in support of personalized health management, starting with a falls prevention initiative. European Innovation Partnership on Active and Healthy Ageing (EIP-AHA). 2013. URL: https://ec.europa.eu/eip/ageing/library/action-plan-specific-action-innovation-support-personalized-health-management-starting-falls_en.html [accessed 2020-01-05]
2. Global strategy and action plan on ageing and health. World Health Organization. 2017. URL: https://www.who.int/ageing/WHO-GSAP-2017.pdf?ua=1 [accessed 2020-01-05]
3. Health at a glance 2019 - OECD INDICATORS. Organization for Economic Co-operation and Development (OECD). URL: https://oecd-ilibrary.org/social-issues-migration-health/health-at-a-glance-2019_4dd50c09-en [accessed 2020-05-17]
4. Ageing Europe - Looking at the lives of older people in the EU. Eurostat - Statistics Explained. 2020. URL: https://ec.europa.eu/eurostat/statistics-explained/index.php/Ageing_Europe_-_looking_at_the_lives_of_older_people_in_the_EU [accessed 2020-11-29]
5. Salzman B. Gait and balance disorders in older adults. Am Fam Physician 2010 Jul 01;82(1):61-68 [FREE Full text] [Medline: 20590073]
6. Agrawal Y, Van de Berg R, Wuyts F, Walther L, Magnusson M, Oh E, et al. Presbyvestibulopathy: Diagnostic Criteria Consensus document of the classification committee of the Bárány Society. J Vestib Res 2019;29(4):161-170. [doi: 10.3233/VES-190672] [Medline: 31306146]
7. WHO global report on falls prevention in older age. World Health Organization. 2007. URL: https://www.who.int/ageing/publications/Falls_prevention7March.pdf [accessed 2020-01-05]
8. Berry SD, Miller RR. Falls: epidemiology, pathophysiology, and relationship to fracture. Curr Osteoporos Rep 2008 Dec;6(4):149-154 [FREE Full text] [doi: 10.1007/s11914-008-0026-4] [Medline: 19032925]
9. Haagsma JA, Olij BF, Majdan M, van Beeck EF, Vos T, Castle CD, et al. Falls in older aged adults in 22 European countries: incidence, mortality and burden of disease from 1990 to 2017. Inj Prev 2020 Oct;26(Supp 1):i67-i74. [doi: 10.1136/injuryprev-2019-043347] [Medline: 32111726]
10. Integrated care for older people Guidelines on community-level interventions to manage declines in intrinsic capacity. World Health Organization Institutional Repository for Information Sharing. 2017. URL: https://apps.who.int/iris/handle/10665/258981 [accessed 2020-01-05]

11. Neuhauser H. The epidemiology of dizziness and vertigo. Handb Clin Neurol 2016;137:67-82. [doi: 10.1016/B978-0-444-63437-5.00005-4] [Medline: 27638063]
12. Wojtczak R, Narożny W, Kuczkowski J, Siebert J. Epidemiology of dizziness in northern Poland - The first Polish neurootologic survey of the general population. Ann Agric Environ Med 2017 Sep 21;24(3):502-506. [doi: 10.5604/12321966.1228401] [Medline: 28954498]
13. Fernández L, Breinbauer HA, Delano PH. Vertigo and dizziness in the elderly. Front Neurol 2015;6:144. [doi: 10.3389/fneur.2015.00144] [Medline: 26167157]
14. Kerber KA, Callaghan BC, Telian SA, Meurer WJ, Skolarus LE, Carender W, et al. Dizziness symptom type prevalence and overlap: A US nationally representative survey. Am J Med 2017 Dec;130(12):1465.e1-1465.e9. [doi: 10.1016/j.amjmed.2017.05.048] [Medline: 28739195]
15. Lin HW, Bhattacharyya N. Balance disorders in the elderly: epidemiology and functional impact. Laryngoscope 2012 Aug;122(8):1858-1861. [doi: 10.1002/lary.23376] [Medline: 22645067]
16. Maarsingh OR, Dros J, Schellevis FG, van Weert HC, Bindels PJ, Horst HEVD. Dizziness reported by elderly patients in family practice: prevalence, incidence, and clinical characteristics. BMC Fam Pract 2010 Jan 11;11:2 [FREE Full text] [doi: 10.1186/1471-2296-11-2] [Medline: 20064231]
17. Gillespie LD, Robertson MC, Gillespie WJ, Sherrington C, Gates S, Clemson LM, et al. Interventions for preventing falls in older people living in the community. Cochrane Database Syst Rev 2012 Sep 12(9):CD007146. [doi: 10.1002/14651858.CD007146.pub3] [Medline: 22972103]
18. Ekvall Hansson E, Magnusson M. Vestibular asymmetry predicts falls among elderly patients with multi-sensory dizziness. BMC Geriatr 2013 Jul 22;13:77 [FREE Full text] [doi: 10.1186/1471-2318-13-77] [Medline: 23875891]
19. Kristinsdottir EK, Jarnlo GB, Magnusson M. Asymmetric vestibular function in the elderly might be a significant contributor to hip fractures. Scand J Rehabil Med 2000 Jun;32(2):56-60. [Medline: 10853718]
20. Kristinsdottir EK, Nordell E, Jarnlo GB, Tjäder A, Thorngren KG, Magnusson M. Observation of vestibular asymmetry in a majority of patients over 50 years with fall-related wrist fractures. Acta Otolaryngol 2001 Jun;121(4):481-485. [Medline: 11508508]
21. Baldursdottir B, Petersen H, Jonsson PV, Mogensen B, Whitney SL, Ramel A, et al. Sensory impairments and wrist fractures: A case-control study. J Rehabil Med 2018 Feb 13;50(2):209-215 [FREE Full text] [doi: 10.2340/16501977-2312] [Medline: 29271981]
22. Heinrich S, Rapp K, Rissmann U, Becker C, König HH. Cost of falls in old age: a systematic review. Osteoporos Int 2010 Jun;21(6):891-902. [doi: 10.1007/s00198-009-1100-1] [Medline: 19924496]
23. Kerber KA. Vertigo and dizziness in the emergency department. Emerg Med Clin North Am 2009 Feb;27(1):39-50 [FREE Full text] [doi: 10.1016/j.emc.2008.09.002] [Medline: 19218018]
24. Kovacs E, Wang X, Grill E. Economic burden of vertigo: a systematic review. Health Econ Rev 2019 Dec 27;9(1):37 [FREE Full text] [doi: 10.1186/s13561-019-0258-2] [Medline: 31883042]
25. Saber Tehrani AS, Coughlan D, Hsieh YH, Mantokoudis G, Korley FK, Kerber KA, et al. Rising annual costs of dizziness presentations to U.S. emergency departments. Acad Emerg Med 2013 Jul;20(7):689-696. [doi: 10.1111/acem.12168] [Medline: 23859582]
26. Bhaskar S, Bradley S, Chattu VK, Adisesh A, Nurtazina A, Kyrykbayeva S, et al. Telemedicine across the Globe-Position Paper from the COVID-19 Pandemic Health System Resilience PROGRAM (REPROGRAM) International Consortium (Part 1). Front Public Health 2020;8:556720. [doi: 10.3389/fpubh.2020.556720] [Medline: 33178656]
27. Gammon D, Berntsen GKR, Koricho AT, Sygna K, Ruland C. The chronic care model and technological research and innovation: a scoping review at the crossroads. J Med Internet Res 2015 Feb 06;17(2):e25 [FREE Full text] [doi: 10.2196/jmir.3547] [Medline: 25677200]
28. Kim K, Gollamudi SS, Steinhubl S. Digital technology to enable aging in place. Exp Gerontol 2017 Feb;88:25-31. [doi: 10.1016/j.exger.2016.11.013] [Medline: 28025126]
29. Global Observatory of eHealth. Global diffusion of eHealth: Making universal health coverage achievable: Report of the third Global Observatory on eHealth. World Health Organization. 2016. URL: https://apps.who.int/iris/bitstream/handle/10665/252529/9789241511780-eng.pdf;jsessionid=4857CD1FAD00D6536337F2868A07FD81?sequence=1#page=16 [accessed 2019-11-07]
30. Lapão L, Dussault G. The contribution of eHealth and mHealth to improving the performance of the health workforce: a review. Public Health Panorama.: World Health Organization Regional Office for Europe; 2017. URL: https://apps.who.int/iris/handle/10665/325262 [accessed 2019-11-07]
31. eHealth at WHO. World Health Organization. 2019 Dec 3. URL: https://www.who.int/ehealth/about/en/ [accessed 2019-12-03]
32. Gordon NP, Crouch E. Digital information technology use and patient preferences for internet-based health education modalities: cross-sectional survey study of middle-aged and older adults with chronic health conditions. JMIR Aging 2019 Apr 04;2(1):e12243 [FREE Full text] [doi: 10.2196/12243] [Medline: 31518291]
33. Merkel S, Hess M. The use of internet-based health and care services by elderly people in Europe and the importance of the country context: multilevel study. JMIR Aging 2020 Jun 03;3(1):e15491 [FREE Full text] [doi: 10.2196/15491] [Medline: 32490837]

34. Ammenwerth E, Duftschmid G, Al-Hamdan Z, Bawadi H, Cheung NT, Cho K, et al. International comparison of six basic eHealth indicators across 14 countries: an eHealth benchmarking study. Methods Inf Med 2020 Dec;59(S 02):e46-e63 [FREE Full text] [doi: 10.1055/s-0040-1715796] [Medline: 33207386]
35. Yelnik A, Bonan I. Clinical tools for assessing balance disorders. Neurophysiol Clin 2008 Dec;38(6):439-445. [doi: 10.1016/j.neucli.2008.09.008] [Medline: 19026963]
36. Montesinos L, Castaldo R, Pecchia L. Wearable inertial sensors for fall risk assessment and prediction in older adults: a systematic review and meta-analysis. IEEE Trans Neural Syst Rehabil Eng 2018 Mar;26(3):573-582. [doi: 10.1109/TNSRE.2017.2771383] [Medline: 29522401]
37. Bet P, Castro PC, Ponti MA. Fall detection and fall risk assessment in older person using wearable sensors: A systematic review. Int J Med Inform 2019 Oct;130:103946. [doi: 10.1016/j.ijmedinf.2019.08.006] [Medline: 31450081]
38. Moreira Bittar RS, Bovino Pedalini ME, Oiticica Ramalho J, Yoshimura R. Critical analysis of vestibular rehabilitation outcome according to dizziness etiology. Braz J Otorhinolaryngol 2007 Nov;73(6):760-764. [doi: 10.1016/s1808-8694(15)31172-1]
39. Han BI, Song HS, Kim JS. Vestibular rehabilitation therapy: review of indications, mechanisms, and key exercises. J Clin Neurol 2011 Dec;7(4):184-196 [FREE Full text] [doi: 10.3988/jcn.2011.7.4.184] [Medline: 22259614]
40. Lacour M, Bernard-Demanze L. Interaction between vestibular compensation mechanisms and vestibular rehabilitation therapy: 10 recommendations for optimal functional recovery. Front Neurol 2014;5:285. [doi: 10.3389/fneur.2014.00285] [Medline: 25610424]
41. Rossi-Izquierdo M, Gayoso-Diz P, Santos-Pérez S, Del-Río-Valeiras M, Faraldo-García A, Vaamonde-Sánchez-Andrade I, et al. Vestibular rehabilitation in elderly patients with postural instability: reducing the number of falls-a randomized clinical trial. Aging Clin Exp Res 2018 Nov;30(11):1353-1361. [doi: 10.1007/s40520-018-1003-0] [Medline: 30008159]
42. Skjæret N, Nawaz A, Morat T, Schoene D, Helbostad JL, Vereijken B. Exercise and rehabilitation delivered through exergames in older adults: An integrative review of technologies, safety and efficacy. Int J Med Inform 2016 Jan;85(1):1-16. [doi: 10.1016/j.ijmedinf.2015.10.008] [Medline: 26559887]
43. Choi SD, Guo L, Kang D, Xiong S. Exergame technology and interactive interventions for elderly fall prevention: A systematic literature review. Appl Ergon 2017 Nov;65:570-581. [doi: 10.1016/j.apergo.2016.10.013] [Medline: 27825723]
44. Sun R, Sosnoff JJ. Novel sensing technology in fall risk assessment in older adults: a systematic review. BMC Geriatr 2018 Jan 16;18(1):14 [FREE Full text] [doi: 10.1186/s12877-018-0706-6] [Medline: 29338695]
45. Rucco R, Sorriso A, Liparoti M, Ferraioli G, Sorrentino P, Ambrosanio M, et al. Type and location of wearable sensors for monitoring falls during static and dynamic tasks in healthy elderly: a review. Sensors (Basel) 2018 May 18;18(5):1613 [FREE Full text] [doi: 10.3390/s18051613] [Medline: 29783647]
46. Nguyen H, Mirza F, Naeem MA, Baig MM. Falls management framework for supporting an independent lifestyle for older adults: a systematic review. Aging Clin Exp Res 2018 Nov;30(11):1275-1286. [doi: 10.1007/s40520-018-1026-6] [Medline: 30196346]
47. Leirós-Rodríguez R, García-Soidán JL, Romo-Pérez V. Analyzing the use of accelerometers as a method of early diagnosis of alterations in balance in elderly people: a systematic review. Sensors (Basel) 2019 Sep 09;19(18):3883 [FREE Full text] [doi: 10.3390/s19183883] [Medline: 31505828]
48. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JPA, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. PLoS Med 2009 Jul 21;6(7):e1000100 [FREE Full text] [doi: 10.1371/journal.pmed.1000100] [Medline: 19621070]
49. O'Connor D, Green S, Higgins JPT. Chapter 5: Defining the review questiondeveloping criteria for including studies. In: Higgins JPT, Green S, editors. Cochrane Handbook of Systematic Reviews of Intervention. Version 5.1.0 (updated March 2011). London, UK: The Cochrane Collaboration; 2011.
50. Active ageing: a policy framework. World Health Organization. 2002. URL: https://www.who.int/ageing/publications/active_ageing/en/ [accessed 2019-11-05]
51. Knudson D. Fundamentals of Biomechanics. 2nd edition. New York: Springer; 2007.
52. Rosso AL, Studenski SA, Chen WG, Aizenstein HJ, Alexander NB, Bennett DA, et al. Aging, the central nervous system, and mobility. J Gerontol A Biol Sci Med Sci 2013 Nov;68(11):1379-1386 [FREE Full text] [doi: 10.1093/gerona/glt089] [Medline: 23843270]
53. A health telematics policy in support of WHO's Health-for-all strategy for global health development: report of the WHO Group Consultation on Health Telematics, 11-16 December, Geneva, 1997. World Health Organization. 1998. URL: https://apps.who.int/iris/handle/10665/63857 [accessed 2021-04-15]
54. Eysenbach G. What is e-health? J Med Internet Res 2001;3(2):E20 [FREE Full text] [doi: 10.2196/jmir.3.2.e20] [Medline: 11720962]
55. Deldar K, Bahaadinbeigy K, Tara SM. Teleconsultation and clinical decision making: a systematic review. Acta Inform Med 2016 Jul 16;24(4):286-292 [FREE Full text] [doi: 10.5455/aim.2016.24.286-292] [Medline: 27708494]
56. Jones DP. Biomedical Sensors. New York: Momentum Press; 2010.
57. Critical appraisal tools. The Joanna Briggs Institute. URL: https://jbi.global/critical-appraisal-tools [accessed 2019-01-05]

58. Julian PT Higgins, Jelena Savovic, Matthew J Page, Jonathan AC Sterne on behalf of the ROB2 Development Group. Revised Cochrane risk-of-bias tool for randomized trials (Rob2). URL: https://sites.google.com/site/riskofbiastool/welcome/rob-2-0-tool/current-version-of-rob-2 [accessed 2019-01-05]
59. Eggenberger P, Theill N, Holenstein S, Schumacher V, de Bruin ED. Multicomponent physical exercise with simultaneous cognitive training to enhance dual-task walking of older adults: a secondary analysis of a 6-month randomized controlled trial with 1-year follow-up. Clin Interv Aging 2015;10:1711-1732 [FREE Full text] [doi: 10.2147/CIA.S91997] [Medline: 26604719]
60. Gschwind YJ, Eichberg S, Ejupi A, de Rosario H, Kroll M, Marston HR, et al. ICT-based system to predict and prevent falls (iStoppFalls): results from an international multicenter randomized controlled trial. Eur Rev Aging Phys Act 2015;12:10 [FREE Full text] [doi: 10.1186/s11556-015-0155-6] [Medline: 26865874]
61. Gschwind YJ, Schoene D, Lord SR, Ejupi A, Valenzuela T, Aal K, et al. The effect of sensor-based exercise at home on functional performance associated with fall risk in older people - a comparison of two exergame interventions. Eur Rev Aging Phys Act 2015;12:11 [FREE Full text] [doi: 10.1186/s11556-015-0156-5] [Medline: 26865875]
62. Lim SB, Horslen BC, Davis JR, Allum JHJ, Carpenter MG. Benefits of multi-session balance and gait training with multi-modal biofeedback in healthy older adults. Gait Posture 2016 Jun;47:10-17. [doi: 10.1016/j.gaitpost.2016.03.017] [Medline: 27264396]
63. Oesch P, Kool J, Fernandez-Luque L, Brox E, Evertsen G, Civit A, et al. Exergames versus self-regulated exercises with instruction leaflets to improve adherence during geriatric rehabilitation: a randomized controlled trial. BMC Geriatr 2017 Mar 23;17(1):77 [FREE Full text] [doi: 10.1186/s12877-017-0467-7] [Medline: 28330455]
64. Ozaki K, Kondo I, Hirano S, Kagaya H, Saitoh E, Osawa A, et al. Training with a balance exercise assist robot is more effective than conventional training for frail older adults. Geriatr Gerontol Int 2017 Nov;17(11):1982-1990. [doi: 10.1111/ggi.13009] [Medline: 28295912]
65. Hong J, Kong H, Yoon H. Web-based telepresence exercise program for community-dwelling elderly women with a high risk of falling: randomized controlled trial. JMIR Mhealth Uhealth 2018 May 28;6(5):e132 [FREE Full text] [doi: 10.2196/mhealth.9563] [Medline: 29807877]
66. Zakaria NA, Kuwae Y, Tamura T, Minato K, Kanaya S. Quantitative analysis of fall risk using TUG test. Comput Methods Biomech Biomed Engin 2015;18(4):426-437. [doi: 10.1080/10255842.2013.805211] [Medline: 23964848]
67. Hall CD, Clevenger CK, Wolf RA, Lin JS, Johnson TM, Wolf SL. Feasibility of a low-cost, interactive gaming system to assess balance in older women. J Aging Phys Act 2016 Jan;24(1):111-118. [doi: 10.1123/japa.2014-0184] [Medline: 26181324]
68. Howcroft J, Kofman J, Lemaire ED, McIlroy WE. Analysis of dual-task elderly gait in fallers and non-fallers using wearable sensors. J Biomech 2016 May 03;49(7):992-1001. [doi: 10.1016/j.jbiomech.2016.01.015] [Medline: 26994786]
69. Lee C, Sun T, Jiang B, Choi V. Using wearable accelerometers in a community service context to categorize falling behavior. Entropy 2016 Jul 13;18(7):257 [FREE Full text] [doi: 10.3390/e18070257]
70. Ponti M, Bet P, Oliveira CL, Castro PC. Better than counting seconds: Identifying fallers among healthy elderly using fusion of accelerometer features and dual-task Timed Up and Go. PLoS One 2017;12(4):e0175559 [FREE Full text] [doi: 10.1371/journal.pone.0175559] [Medline: 28448509]
71. Similä H, Immonen M, Ermes M. Accelerometry-based assessment and detection of early signs of balance deficits. Comput Biol Med 2017 Jun 01;85:25-32. [doi: 10.1016/j.compbiomed.2017.04.009] [Medline: 28432935]
72. Shahzad A, Ko S, Lee S, Lee J, Kim K. Quantitative assessment of balance impairment for fall-risk estimation using wearable triaxial accelerometer. IEEE Sensors J 2017 Oct 15;17(20):6743-6751. [doi: 10.1109/jsen.2017.2749446]
73. Brodie MA, Coppens MJ, Ejupi A, Gschwind YJ, Annegarn J, Schoene D, et al. Comparison between clinical gait and daily-life gait assessments of fall risk in older people. Geriatr Gerontol Int 2017 Nov;17(11):2274-2282. [doi: 10.1111/ggi.12979] [Medline: 28176431]
74. Howcroft J, Lemaire ED, Kofman J, McIlroy WE. Dual-task elderly gait of prospective fallers and non-fallers: a wearable-sensor based analysis. Sensors (Basel) 2018 Apr 21;18(4):1275 [FREE Full text] [doi: 10.3390/s18041275] [Medline: 29690496]
75. Chigateri NG, Kerse N, Wheeler L, MacDonald B, Klenk J. Validation of an accelerometer for measurement of activity in frail older people. Gait Posture 2018 Oct;66:114-117. [doi: 10.1016/j.gaitpost.2018.08.024] [Medline: 30172217]
76. Qiu H, Rehman RZU, Yu X, Xiong S. Application of wearable inertial sensors and a new test battery for distinguishing retrospective fallers from non-fallers among community-dwelling older people. Sci Rep 2018 Nov 05;8(1):16349 [FREE Full text] [doi: 10.1038/s41598-018-34671-6] [Medline: 30397282]
77. Hsieh KL, Roach KL, Wajda DA, Sosnoff JJ. Smartphone technology can measure postural stability and discriminate fall risk in older adults. Gait Posture 2019 Jan;67:160-165. [doi: 10.1016/j.gaitpost.2018.10.005] [Medline: 30340129]
78. Maneeprom N, Taneepanichskul S, Panza A, Suputtitada A. Effectiveness of robotics fall prevention program among elderly in senior housings, Bangkok, Thailand: a quasi-experimental study. Clin Interv Aging 2019;14:335-346. [doi: 10.2147/CIA.S182336] [Medline: 30863027]
79. Nightingale CJ, Mitchell SN, Butterfield SA. Validation of the Timed Up and Go Test for assessing balance variables in adults aged 65 and older. J Aging Phys Act 2019 Apr 01;27(2):230-233. [doi: 10.1123/japa.2018-0049] [Medline: 30117359]

80. Viirre E. Vestibular telemedicine and rehabilitation. Applications for virtual reality. Stud Health Technol Inform 1996;29:299-305. [Medline: 10163763]

Abbreviations

ICT: information and communication technologies
JBI: Joanna Briggs Institute
QES: quasiexperimental study
RCT: randomized controlled trial
TUGT: Timed Up and Go Test

Edited by R Kukafka; submitted 08.07.20; peer-reviewed by L Barbosa, H Wang, J Howcroft; comments to author 17.11.20; revised version received 10.01.21; accepted 25.02.21; published 28.04.21.

Please cite as:
Gaspar AGM, Lapão LV
eHealth for Addressing Balance Disorders in the Elderly: Systematic Review
J Med Internet Res 2021;23(4):e22215
URL: https://www.jmir.org/2021/4/e22215
doi:10.2196/22215
PMID:33908890

Review

Utility, Value, and Benefits of Contemporary Personal Health Records: Integrative Review and Conceptual Synthesis

Umar Ruhi[1], BSc, MBA, PhD; Ritesh Chugh[2], GCTertEd, GDInfSys, MInfSys, PhD

[1]Telfer School of Management, University of Ottawa, Ottawa, ON, Canada

[2]School of Engineering & Technology, Central Queensland University, Melbourne, Australia

Corresponding Author:
Umar Ruhi, BSc, MBA, PhD
Telfer School of Management
University of Ottawa
55 Laurier Ave East
Ottawa, ON, K1N 6N5
Canada
Phone: 1 6135625800 ext 1990
Email: umar.ruhi@uottawa.ca

Abstract

Background: Contemporary personal health record (PHR) technologies offer a useful platform for individuals to maintain a lifelong record of personally reported and clinically sourced data from various points of medical care.

Objective: This paper presents an integrative review and synthesis of the extant literature on PHRs. This review draws upon multiple lenses of analysis and deliberates value perspectives of PHRs at the product, consumer, and industry levels.

Methods: Academic databases were searched using multiple keywords related to PHRs for the years 2001-2020. Three research questions were formulated and used as selection criteria in our review of the extant literature relevant to our study.

Results: We offer a high-level functional utility model of PHR features and functions. We also conceptualize a consumer value framework of PHRs, highlighting the applications of these technologies across various health care delivery activities. Finally, we provide a summary of the benefits of PHRs for various health care constituents, including consumers, providers, payors, and public health agencies.

Conclusions: PHR products offer a myriad of content-, connectivity-, and collaboration-based features and functions for their users. Although consumers benefit from the tools provided by PHR technologies, their overall value extends across the constituents of the health care delivery chain. Despite advances in technology, our literature review identifies a shortfall in the research addressing consumer value enabled by PHR tools. In addition to scholars and researchers, our literature review and proposed framework may be especially helpful for value analysis committees in the health care sector that are commissioned for the appraisal of innovative health information technologies such as PHRs.

(J Med Internet Res 2021;23(4):e26877) doi:10.2196/26877

KEYWORDS

electronic personal health records; PHR; functionality synopsis; value analysis; consumer health informatics

Introduction

Among the many technology applications available to individuals today for managing their own health and wellness, electronic personal health records (PHRs) offer a valuable means to facilitate active participation of health care consumers, including patients and their caregivers. By virtue of their potential capabilities to help individuals track their health conditions, provide access to patient medical record information (PMRI), and offer communication tools to interact with health care providers, PHRs have been regarded as a paradigm shift toward consumer-centric and patient-oriented health and medical services [1,2].

Although the consumer adoption of PHR systems has been slower than originally expected [3,4], these technologies are gaining traction in many countries worldwide [5-9]. Government eHealth initiatives in many countries are currently focused on the implementation of these technologies to foster greater patient engagement with personal health information (PHI) management and care coordination. An example of such an initiative is the

Stage 3 Meaningful Use program under the US Health Information Technology for Economic and Clinical Health initiative. This program calls for improving patient engagement through functionality, such as patient access to medical records, patient communication tools, and interoperability with hospital electronic health records (EHRs) [10,11]. Across the border, in Canada, several federally funded projects sponsored by the Canada Health Infoway are also geared toward the deployment of consumer-focused digital health technologies, including patient health information records, patient-physician communication tools, and remote patient monitoring [12]. Along similar lines, the p-medicine and eHealthMonitor projects funded by the European Union also aim to support personalized medicine through technologies such as PHRs [7].

The overarching vision behind PHR technology offerings is to enable patient empowerment, reduce health care costs, and provide better continuity of care [3] through access to timely, reliable, and comprehensible health information for patients and streamlined communication between patients and health care providers [13,14]. The objective of this paper is to offer a review of the utility, value, and benefits of PHR systems through a discussion of their features and functions and to deliberate how PHR functionality can potentially translate into value for the health care consumer and benefits for the health care system as a whole. Our review comprises both an analysis and a synthesis-oriented exposition on the current landscape of PHR technologies. The *Methods* section outlines our review approach and the ensuing structure of this study.

Methods

Overview

In characterizing the type of review offered in this paper, our discussion aligns with an integrative review, in which literature pertinent to a subject area is critically analyzed and synthesized to theorize alternative perspectives of the subject [15]. Although integrative literature reviews may serve multiple purposes, their essence is to review existing literature to elicit new insights, inquiries, or answers through the integration and synthesis of existing literature [16]. Integrative reviews are commonly used in health research [17-19], and such reviews have been purported to enhance the development of health care theory, policy, and practice [19]. Our integrative review was structured along the following 5 recommended phases [19]: problem identification, literature search, data evaluation, data analysis, and presentation of the results. The procedures followed are summarized below.

Problem Identification

Specific to our study, the purpose of our review is to extensively research pertinent PHR literature and provide an assessment of the product utility, consumer value, and industry benefits of these systems. Toward this, and in line with integrative reviews, we developed a protocol for the search and selection of relevant literature [15,16], deliberated the capabilities of PHR systems using several lenses of analysis, and then classified and synthesized a typology to formulate conceptual frameworks that explicate the utility and value of PHR systems. Typologies that offer a conceptual classification of constructs are recommended as useful theory-building tools [20] and a valuable form of synthesis in integrative reviews [15].

To guide our review process, we formulated 3 research questions that we aimed to answer through our analysis and synthesis of the extant literature. Table 1 below is the three-step approach that we adapted for our review based on suggested guidelines for the reviews of emerging HITs [21,22]. The *Results* section of this paper discusses the findings and outcomes from our review, as noted in Table 1.

Table 1. Research questions and guidelines followed for the review and synthesis.

Review perspective	RQ[a]	HIT[b] assessment review guidelines	Review and synthesis outcomes
Product utility	RQ1. What features and functions are available in contemporary PHRs[c], and how has this functionality evolved over time?	1. Technology definition and literature search	• Literature selection procedure • PHR working definition • Functional utility model of PHR technologies
Consumer value	RQ2. What is the potential value of various PHR functionalities to health care consumers?	2. Conceptual analysis and framework formulation	• Functional utility model of PHR technologies • PHR consumer value framework
Industry benefits	RQ3. How can the mainstream deployment and use of PHR systems translate into benefits for the health care system as a whole?	3. Reflective synthesis and summary	• Value propositions and benefits of PHR systems

[a]RQ: research question.

[b]HIT: health information technology.

[c]PHR: personal health record.

Search Strategy

Our academic article search was conducted using digital library databases, including PubMed, Web of Science, ScienceDirect, and Scopus. In addition, to ensure the breadth and validity of our search results, we explored the publications cited in previous scoping and systematic reviews of PHRs [23-25] and included any relevant articles that had been overlooked in our own search.

Our search techniques used various terms and keywords related to PHRs, including acronyms as well as expanded terms, such as PHR, personal health record, EPHR (electronic personal health record), patient portal, personal medical record, personally controlled health record, PCHR, personal health information, and PHI.

Data Evaluation

Both authors independently screened titles, keywords, and abstracts to determine whether publications should be included in the review. Our review included studies that explicitly discussed features, functions, utility, value, and benefits of electronic PHRs, whereas it excluded publications focusing on paper-based PHRs or studies solely focusing on psychosocial aspects of end users' PHR adoption or technical system design practices for PHRs. Following the first round of screening, we refined our search criteria and examined articles pertaining to consumer health informatics as a general field of study. Our initial review indicated that some publications pertaining to consumer health informatics directly discuss the benefits of PHR technologies [13,26-30]. In the second round of screening, each author assessed mutually exclusive but collectively exhaustive subsets of all publications identified as potentially relevant, and we ensured that the articles were indeed pertinent to our review.

Data Analysis and Synthesis

Following the selection of relevant literature, our review process began with an iterative concept-centric analysis of the attributes and benefits of PHR systems. We analyzed the literature at the product level by identifying various features and functions of PHR systems described in the extant literature, at the consumer level by deliberating the value of various PHR system functionalities, and at the industry level by identifying the benefits provided by PHR technologies to various health care industry constituents. A codebook was created to facilitate the analysis and extraction of data into systematic categories. The authors collaborated on the conceptual synthesis of the 3 classification systems for functional utility, consumer value, and industry benefits. These conceptual classifications were refined iteratively through simplification, abstraction, and focusing procedures, constituting the constant comparison method commonly recommended for integrative reviews [19].

Presentation and Paper Structure

In the final phase of the integrative review, the results from our analysis and synthesis were summarized and depicted using visual models and concept matrices. These are presented and discussed in the *Results* section of this paper.

We first provide a working definition of PHRs that was used as a touchstone to guide our literature search and subsequent discussion. Drawing upon that definition, we retrieved relevant peer-reviewed publications and industry reports that discussed the functionality, utility, value, and benefits of PHR technologies.

The outcome of our review of PHRs from a product utility perspective comprised an evaluation of various features and functions of PHR systems. The output from this evaluation is conceptualized as a high-level functional model of PHRs that summarizes the myriad of features and functions available in contemporary PHR systems.

Next, we discuss the capabilities of PHR systems from a consumer value orientation by juxtaposing the functionality of PHR systems alongside health care delivery activities ranging from prevention to the diagnosis and ongoing management of illnesses.

Finally, the *Results* section provides an industry-level viewpoint that summarizes various value propositions and benefits related to the use of PHR technologies at the micro, meso, and macro levels. The synthesis offers a discussion of how the effective deployment and use of PHR technologies can potentially translate into benefits for different constituents in the health care delivery chain, including consumers, providers, payors, and public health agencies.

Results

Defining Characteristics of PHRs

As a working definition, this paper adopts one of the earliest and most commonly cited characterization of a PHR as "an electronic application through which individuals can access, manage and share their health information and that of others for whom they are authorized, in a private, secure and confidential environment" [31]. In addition, PHR are sometimes referred to as personally controlled health records (PCHRs) comprising information and communication technologies that can potentially help all types of end users in maintaining health and wellness and specifically facilitate patients in managing their ongoing illnesses [32].

To further delineate the representative attributes of PHRs, we also differentiate between 3 similar yet distinct technologies related to patient records: electronic medical records (EMRs), EHRs, and PHRs. Depending on the health care setting, although these 3 technologies may be used as components in an integrated health information system (HIS), each of them can be differentiated from the other based on its custodianship and level of patient centricity. EMRs are often considered as digital versions of paper charts in a clinician's office [33]. These patient medical records in this instance are provider-centric [33,34] and are rarely accessible to other health care providers or to the patients themselves. In contrast, EHR systems offer a broader view of a patient's care by facilitating integration with HIS beyond the organization that originally collected and compiled the patient information [33]. These systems can aggregate patient data from multiple health care facilities to create a unified patient record that can be accessed by health care providers [35,36]. Finally, PHRs function under the custodianship of patients or their caregivers, and these systems comprise full or partial health information about patients over their lifetime [35,37]. Hence, PHRs specifically pertain to digitally stored health care information about an individual patient under the control of that patient or their caregiver [7,38], whereas EMRs and EHRs are typically maintained by health care providers or payor organizations [39].

Drawing upon these characteristics of PHR systems, this study adopts a consumer-oriented perspective and uses the term PHR to refer to both the underlying patient record and its data elements as well as the software that provides functionality to maintain that record. As such, we do not differentiate between the data (PHR) and PHR-S (software components of PHR), as sometimes done in the industry standard documentation such as HL7 (Health Level Seven) [40,41]. Furthermore, although we note that there may be differences among PHR systems in terms of front-end technology features, back-end information sources, patients' scope of access, and storage locations of online records, we consider electronic access (desktop, web, or mobile) and patient control over health records to be the defining characteristics of PHR systems.

To help understand the functional scope of the current PHR systems, researchers have classified these technologies into 3 main categories: *standalone*, *tethered*, and *interconnected* [32,37]. The main attributes that differentiate these categories are data control, record portability, and system interoperability. The key differences among these categories are outlined below.

Standalone PHRs require users to manually enter data to populate their own health information and medical history. Hence, the content of these applications is under direct physical control and ownership of the consumer. These PHR systems require a considerable long-term commitment from end users who need to be motivated to maintain their PHI in an accurate and complete fashion [37,42]. Although these technologies may be portable in allowing users to access their PHI anytime and anywhere, they lack interoperability because data must be manually imported or exported from other HIS.

Tethered PHR systems are typically offered as extensions of a health care institution's own back-end EHR or EMR system, providing users access to parts of their own EHRs. These systems are also referred to as patient portals. In addition to providing access to patient data, these systems may also include additional functionality, such as communication tools for email, messaging, appointment scheduling, and prescription renewals [37]. Access to these PHRs is typically provided through a web portal interface [32,43]. The data in tethered PHR systems are under the control of the health care provider, hence limiting the portability of patient records, and these technologies may not be fully interoperable with other HIS.

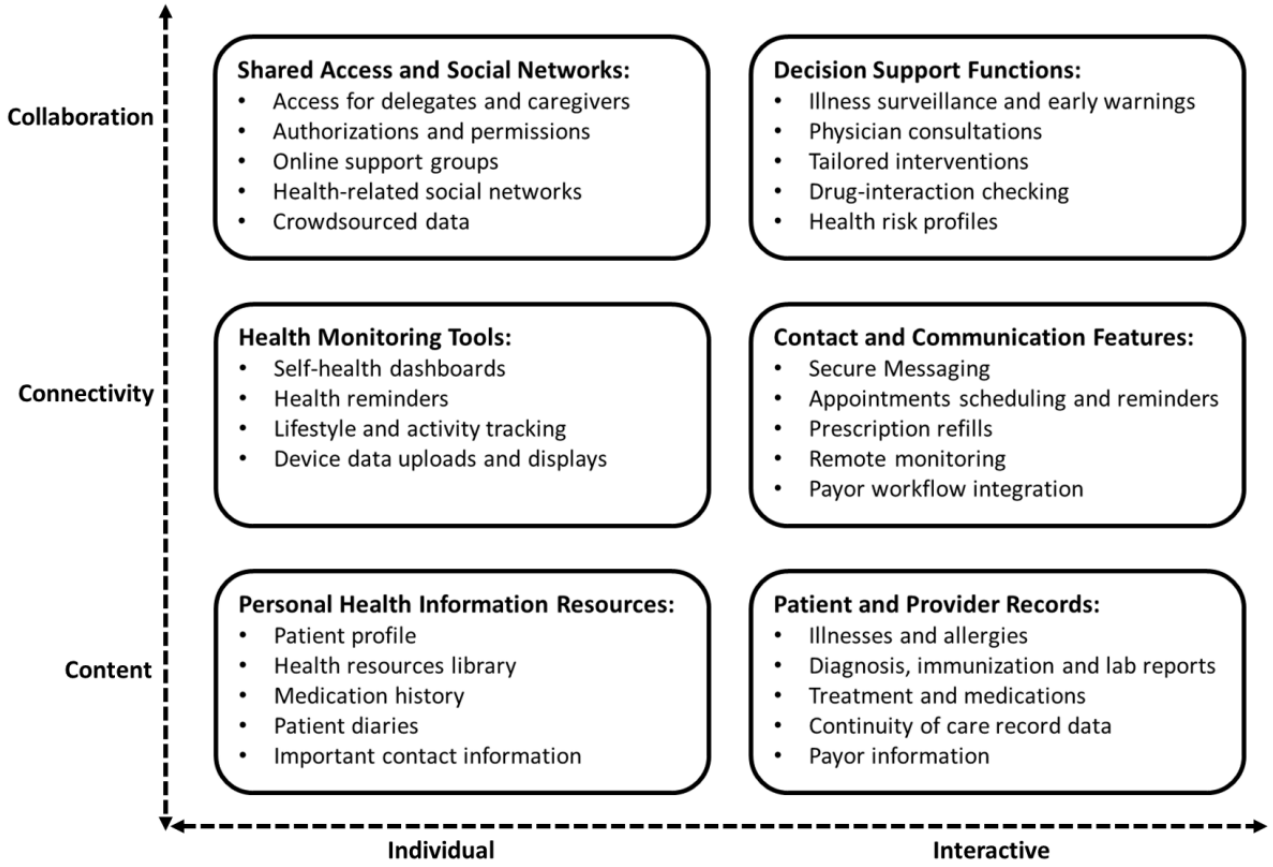
Interconnected PHRs are often described as the ideal or preferred type of PHR in terms of data control, record portability, and system interoperability [7,37,44,45]. These systems can usually be populated with patient information from a variety of sources, including physician EMRs, hospital EHRs, insurance carriers, health plan sponsors, labs, and pharmacies [29,37]. In addition, users can enter their own information in the selected areas of the PHR. These PHR systems offer consumers adequate control over parts of their health records and also alleviate the need for manual data entry. In addition, because of the established electronic linkages among some HIS, records can usually be easily transferred from one provider to another. Despite the limited offerings in this space, the functionality of these integrated systems is expected to translate into a wider range of convenience benefits and improved health outcomes for consumers as well as operational efficiencies for providers.

Product Utility of PHR Systems

On the basis of the discussion above on different types of PHR systems, one may be led to believe that tethered or interconnected PHRs offer considerably advanced functionality in comparison with standalone PHR systems. Although this is certainly true in the context of system capabilities that require back-end integration with provider EMR or EHR systems, there may be a range of other PHR features and functions that do not depend on such integration, and these can be offered through stand-alone PHRs just as well. For example, certain stand-alone PHR products provide a deeper functionality related to health resource libraries, patient-centered health monitoring, and linkages with web-based support groups—features that do not necessarily need high levels of system interoperability with other HIS.

In this section, we draw upon the extant academic literature as well as the current industry PHR software offerings to provide an overview of various functionalities that may be available in contemporary PHR systems. To aid our discussion, we organize the different PHR capabilities into different categories based on the consumers' modalities of use and functional characteristics of system features and functions. Figure 1 depicts the conceptualization of a high-level functional utility model of PHR systems. The elements of the model are briefly described below. It should be noted that our model does not aim to provide a system specification or technical architecture of PHR systems. These have been described elsewhere in the extant literature [30,45,46]. Our model simply aims to provide a scaffold to aid a high-level understanding of PHR systems at the consumer level.

Figure 1. A high-level functional utility model of personal health record systems.



In terms of modalities of use, our model differentiates between the *individual* and *interactive* modes of using PHR functions. Although individual modes of use entail highly personalized user-initiated tasks, interactive modes usually comprise bidirectional exchanges between the consumer and other care delivery constituents, including physicians, providers, or payors. Most of the individual tasks are performed in an asynchronous style with frequent individual interventions, whereas interactive exchanges usually occur in a dynamic and synchronous fashion.

Our model also characterizes PHR functions as being primarily *content*-, *connectivity*-, or *collaboration-oriented*. Content-rich features refer to PHR technologies that are primarily used for information management. Information can be pushed automatically or retrieved on an as-needed basis or it can be maintained within the PHR by users on their own. In connectivity-based applications, information is exchanged, and transactions are conducted in a 2-way flow between applications, devices, organizations, or people. Finally, collaboration-based mechanisms subsume other functional modes and offer tools for interpersonal exchanges and decision support, thereby enabling consumers to proactively manage their health and wellness. On the basis of these criteria for classification, we categorized various PHR features and functions into 6 groups: PHI resources, patient and provider records, health monitoring tools, contact and communication features, shared access and social networks, and decision support functions.

PHI Resources

At their core, most PHR systems comprise a repository of PHI that allows consumers to maintain their own profiles and medical history data. Various tools such as digital diaries to manage the lists of drugs and track personal data such as weight, glucose, and cholesterol levels allow consumers to exercise control over their medical information [47-49]. Additional functionality with links to web-based health information can help consumers create a library of health information resources pertinent to them [50,51].

Patient and Provider Records

For PHRs with electronic links to other HIS from providers, pertinent PMRI can be seamlessly added to the PHR system. Data from patient diagnosis, treatments, and medications can be added from physicians' EMRs or providers' EHR systems [50,52-57]. Patient health summary standards such as the CCR (continuity of care record) can provide guidelines for PHR pertinent data that would provide a holistic view of patient care and consequently improve the portability of patient health information [42,55,58].

Health Monitoring Tools

Beyond self-managed health information, many PHR technologies also facilitate connectivity with a range of medical and lifestyle tracking devices. Data from these devices can be uploaded to PHRs to enable consumers to keep track of their health and wellness [55,56,59-62]. In addition, these behavior management tools can help consumers track their health indicators via a dashboard style interface and also set up various notifications and alerts for any anomalies or items that require their attention [63-65].

Contact and Communication Features

The tools in this category are considered extremely useful by consumers for interconnected PHRs linked to provider EMR and EHR systems [7,45,52,59,60,64]. Although features such as patient-physician and patient-provider secure messaging and appointment scheduling provide convenience to users, other tools for prescription refills and insurance claims processing can help streamline process workflows for all constituents in health care delivery [66-68]. Advanced PHR offerings also provide telehealth features for patients to provide the results of basic health assessments from home and to transfer data from connected medical devices [37,64,69].

Shared Access and Social Networks

Most PHR systems provide a core set of collaboration tools to help consumers share their health information with other authorized people, including caregivers and designated family members. They do so by delegating access rights and permissions to the specific parts of their PHR [37,64,69]. More recently, social networking tools have been integrated with some PHRs to provide patients with more access to information from practitioners as well as from other patients with similar medical conditions [29,70]. The range of possibilities for social networking features in PHR offerings span a wide spectrum from basic moderated health discussion forums for questions and answers [59] to sophisticated sites that crowdsource patient data from connected devices to foster an active dialog among patients or to contribute to further research about illnesses [71].

Decision Support Functions

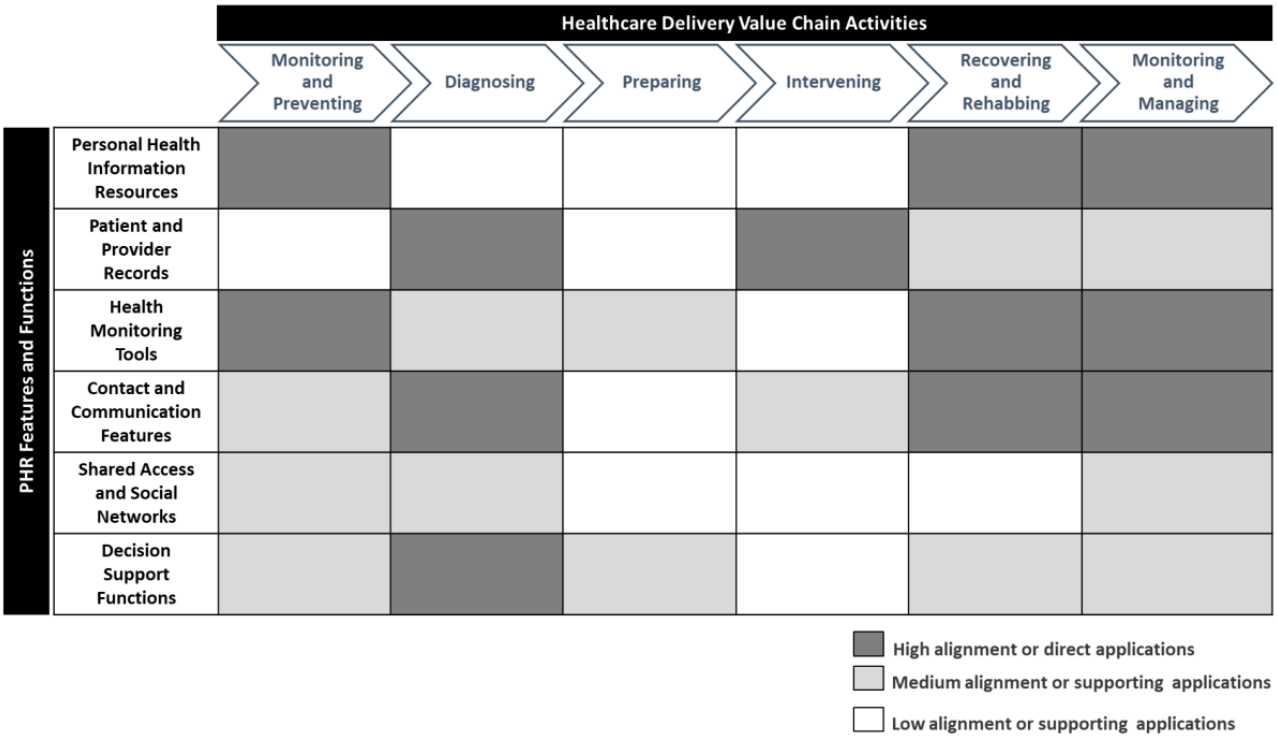
In interconnected PHR systems, collaborative interactions between patients and clinicians can be enabled through decision support features that include illness surveillance, virtual consultations, and computerized tailored interventions [29,72]. In addition, rule-based engines can also provide input to the decision-making processes through tools such as patient health risk profiles and drug-interaction checking [7,73-75]. By using patient data from other parts of the PHR and leveraging practitioner expertise, such tools can help in evaluating the harms and benefits of specific treatment options [37,63,76], issue health warnings through personalized clinical decision support (CDS) notifications [29,77], and recommend alternative treatments [37,73,78]. Recent studies have also shown that personalization-focused PHR functions such as tailored interventions with highly individualized communication, therapy, or medications can be extremely effective in inducing behavioral changes and improving patient health [79,80].

Consumer Value of PHR Systems

Drawing upon the review of the features and functions of PHR systems at the product level, this section discusses a consumer-centric viewpoint of the potential value that might be realized through the effective use of these technologies. To facilitate this viewpoint, we appropriate the conceptual framework of the *care delivery value chain* (CDVC) [81-83], which offers a systematic approach to delineate and analyze health care services and activities that jointly determine the overall success of health outcomes for consumers [81]. According to the CDVC framework, the value for the health care consumer is determined by the results and outcomes rather than the inputs and volume of the range of health care services provided [84]. Using a similar reasoning, we maintain that in the case of PHRs, the consumer value is determined not by the functions and features of technologies per se but also by their potential to address the needs, expectations, and preferences of consumers during the various health care delivery activities. The CDVC classifies these activities into 6 main areas: prevention, diagnosis, preparation, interventions, recovery, and ongoing management of illnesses [82,83].

Figure 2 presents a visual depiction of our conceptualization of the PHR consumer value framework that juxtaposes various PHR features and functions (described in the *Decision Support Functions* section) alongside the CDVC-defined 6 core activities in the health care delivery value chain. The relationships between PHR functionality and health care activities are depicted as lightly or darkly shaded intersections. The latter represents a high (direct) alignment or mapping between specific PHR functionality and health care delivery activity, as evidenced by strong support in the extant literature. To determine the strength of mapping between PHR functional categories and CDVC activities, we identified different use cases for PHR features and functions from the extant literature. These use cases are listed in Multimedia Appendix 1, which serves as a foundation for the summary depicted in Figure 2. A brief overview of the applications of PHR functionalities in different health care delivery activities is provided below.

Figure 2. Personal health record consumer value: a mapping of personal health record functionality to the health care delivery value chain. PHR: personal health record.



Monitoring and Preventing

These activities are primarily concerned with tracking an individual's current conditions and assessing health risks to proactively prevent or reduce the seriousness of illness or injury [81]. Comprehensive patient profile and medical history information sourced from personally reported data as well as patient clinical records can directly help in the early detection of illnesses or reducing the need for medical treatments [7,48]. Furthermore, the integrated use of health devices and self-management dashboards can help track lifestyle-or illness-related risk factors such as diet patterns, blood pressure, and glucose levels, which can help support preventive health care activities [48,55,59,85,86]. Social networks can also help support the proactive posture of patients by helping source relevant information about health conditions and treatments from practitioners and other patients [29,71]. Decision support tools offer an additional functionality to help with the proactive detection and resolution of potentially threatening health conditions through CDS advisories and virtual counseling from experts about potential future health risks [55,74,75,81].

Diagnosing

Diagnosing activities in the care cycle comprise a range of processes, such as laboratory testing, medical history evaluation, consultations with specialists, and the formulation of treatment plans. As depicted in Figure 2, there is a strong support in the extant literature that multiple PHR functionality groups have direct or indirect applications in diagnosis-related care activities. The ready availability of PMRI in interconnected PHR systems can help facilitate these activities [47,52,53,70]. These activities can also be enabled through decision support tools, such as virtual physician consultations and computerized tailored interventions [29,55,70,75]. Finally, in terms of supporting functionalities, personally tracked and self-reported personal information and family medical history shared through the PHR can also help in improving the overall quality of diagnostic processes [47,55,70,75,87].

Preparing

These activities refer to all setup procedures and processes that need to be completed before medical intervention. In the original CDVC framework, the authors note that this set of activities is often overlooked in the health care system [81,82]. Perhaps this is why our review also yielded very little direct evidence of PHR applications in this area. Nonetheless, we believe that PHR functionality related to health monitoring tools and decision support functions can help in supporting such preparatory activities. The former can help with the tracking of important health data from lifestyle health data dashboards and connected medical devices [61,62,65], whereas the latter can facilitate reviews of health profiles, interactions with specialists, and verification of potential drug interactions [29,73].

Intervening

Intervention processes and procedures are targeted at reversing or mitigating a health condition [81], and they typically include the initiation of therapy, treatment, or medication and the management of potential infections or associated illnesses. PHRs can play an important role in improving the quality of medical interventions by providing up-to-date medical history and PMRI across various points of care and by facilitating continuity of care [42,54,66,68]. In addition, contact and communication features in PHR systems can help patients receive regular and situational counseling on treatment and prognosis and can help attending physicians ensure treatment compliance [81,83].

Recovering and Rehabbing

These services are an essential component of care for all medical conditions [81], and the procedures are important for ensuring effective recovery and positive health outcomes for patients in the long term. PHR systems can help in achieving these health outcomes through an improved health monitoring of risk factors and lifestyle data [55-57,61,65,86,88,89] and by seamlessly connecting patients to providers for referrals, follow-ups, and prescription refills [63]. These features are also complemented well by decision support functions that can facilitate the development of tailored interventions and help in tracking and resolving treatment side effects [62,73,75,90].

Monitoring and Managing

Activities that constitute the final part of the care delivery chain aim to manage patient conditions and monitor therapy compliance on an ongoing basis [83]. As shown in Figure 2, PHI resources, health monitoring tools, and contact and communication features can potentially play a significant role in enabling these monitoring and managing activities [56,57,85,89,91,92]. Multiple studies pertaining to chronically ill patients using PHR technologies have demonstrated the usefulness of features such as personal logs and links to educational resources as well as tools such as web-based appointment scheduling, meeting reminders, and email communication with health care professionals [85,91,93]. Together, these tools can play a vital role in minimizing long-term health risks related to chronic illnesses [56,57,82,89,92].

Overall, the effective deployment and adoption of PHRs can potentially enable improved integration across health care activities that constitute the full cycle of care for a consumer. Such integration across the entire chain has been posited as the major driver of health care consumer value [83].

Our analytical framework highlights the similarities and complementarities among various PHR functions and features by conceptualizing the direct and supporting applications of PHRs in different health care activities. Furthermore, a circumspect inspection of the framework shows that the consumer value from contemporary PHRs primarily relates to care activities with a proactive health management orientation. Monitoring activities at the beginning and end of the care delivery chain has a high number of associated PHR applications, followed by activities related to diagnosis and recovery. From the viewpoint of various PHR functional categories, contact and communication features appear to have the most recurrent use cases, followed by personally and clinically sourced health record information as well as health monitoring tools. Our analysis supports previous research that indicates that PHR users consider connectivity features that facilitate health care processes to be the most useful tools [48,94].

Industry Benefits of PHR Systems

Having discussed the functionality of PHR systems and their applications in various health care activities, this section of the paper offers a summary of the benefits of PHR systems for the health care industry. In deliberating these benefits, we underscore our assertion that the value-producing potential of PHRs is not only dependent on the adoption of these technologies by consumers but also on the active participation for the provision and use of these technologies by multiple health care delivery constituents, such as hospitals, labs, pharmacies, insurance companies, and government agencies. Consequently, the effective deployment and adoption of PHRs can result in a variety of benefits for these constituents [95].

The lens of analysis used to outline the industry benefits of PHRs is based on an extension of the health care value chain described in the *Monitoring and Managing* section. However, rather than focusing on care delivery activities, we adopt a channel partner perspective and highlight the benefits of PHR technologies to various entities that comprise the health care system. Using standard health care industry nomenclature [96,97], we refer to providers as any clinicians, allied health professionals, or organizations that render direct health care services to consumers; payors as entities that finance the cost of health services (eg, insurance carriers or health plan sponsors); and public health agencies as government institutions concerned with research and policy issues for the social well-being of communities as a whole [97].

Table 2 offers a review summary of the industry benefits of PHRs by outlining the core value propositions and the principal benefits of these technologies to different constituents. Value propositions pertain to micro-level benefits for consumers, meso-level benefits for providers and payors, and macro-level benefits in the realm of public health. These are briefly discussed herewith.

Table 2. Value propositions and benefits of personal health record systems to health care delivery constituents.

Value propositions and benefits	Literature support	Health care delivery constituents			
		Consumers	Providers	Payors	Public health agencies
Consumer empowerment and patient engagement					
Promote consumer health education	[14,50,57,68,76]	✓[a]	✓		✓
Enable patients to become informed health care consumers	[5,42,44,66,98]	✓			✓
Enhance understanding of medical conditions	[29,42,57,66,74,75,99,100]	✓	✓		
Simplify and clarify patient instructions	[29,42,57,66,74,75,99,100]	✓	✓		
Provide a greater control over health outcomes	[3,56,101,102]	✓			
Offer convenient self-health management	[21,56,68,86,89,103,104]	✓			
Facilitate self-efficacy via cues for patient action	[13,102]	✓			
Health care communication					
Improve patient-physician or provider communication	[14,42,105]	✓	✓		
Timely information sharing for clinical decisions	[68,85,90,103]	✓	✓		
End-to-end care delivery involving multiple constituents	[44,55,68,89,106,107]	✓	✓	✓	✓
Process efficiencies and cost effectiveness					
Increased portability of patient records	[44,57,63,68,89,108,109]	✓	✓	✓	
Reduced cost of chronic disease management	[32,110]	✓	✓	✓	
Greater medical information validity and accuracy	[42,59,75,76,111,112]	✓	✓	✓	✓
Save patient, physician, and provider time	[14,32,57,85]	✓	✓	✓	
Reduced cost of duplication of tests and procedures	[42,59,113]	✓		✓	✓
Enhanced quality of care					
Increased patient safety considerations	[90,103,114]	✓	✓		✓
Improved handling of emergency situations	[115,116]	✓	✓		
Extended durability of patient data	[32,104,109]	✓	✓		✓
Early identification of patient risks and health susceptibilities	[76,115,117,118]	✓	✓	✓	
Public health outcomes					
Reduced burden on health care system and resources	[3,55,59,63]		✓	✓	✓
Enhanced care for underserved communities and populations	[3,59,63,95]	✓	✓		✓
Facilitate care in public health emergencies	[119,120]	✓	✓		✓
Support public health research	[71,90,121,122]				✓
New avenues for epidemiology surveillance and screenings	[71,90,121,122]				✓

[a]Mapping of value propositions and benefits of personal health records to various health care delivery constituents.

Consumers

From the perspective of *consumers*, PHR users are not only likely to be better informed about their health conditions [5,63,123], but they also actively participate and increasingly contribute toward their own health management activities [57,102]. PHRs have the potential to provide patients with a better understanding of health information and clearer health care instructions [74,100]. PHRs can also assist users in monitoring daily self-care activities and enable patients to collaborate and share their experiences with their providers and caregivers. As an integrated patient-centered technology, PHRs also offer the means for increased patient engagement through improved provider-patient and physician-patient communication [14,54,66,68,74], thus leading to greater personalization of care [3].

Providers

From the perspective of *providers*, a primary benefit of PHR systems is that these technologies address a significant gap in the current health information exchange mechanisms. In the absence of stable and formal technology standards that allow the transfer of patient records from one provider to another, PHRs can offer an alternative means to achieve this purpose [44,63,107]. Patients can access their health records as, when, and where needed. PHR systems can also help reduce health care costs and inefficiencies, especially those associated with inaccurate information and effort duplication [42,57,59,113]. Patients can directly verify health data, and a complete access to patient history from across providers can assist in avoiding unnecessary laboratory tests and medical procedures.

Payors

Providers and *payors* can also benefit from the patient adoption of PHRs because the use of these systems is likely to improve patient safety through an early identification of health risks [76,90,103,114], reduce the cost of chronic disease management [32,110], and enable health care institutions to better handle emergency situations [115,116]. Access to unified PMRI across health care providers can help alleviate medical treatment disruptions for patients with chronic conditions [124], and features such as remote monitoring and eHealth consultations can enable earlier and efficient hospital discharge and long-term patient monitoring processes [64,69].

Public Health Agencies

From the viewpoint of *public health agencies* at a macro level, the mainstream adoption of PHRs can lead to a variety of benefits for population health [95]. These technologies can help in reducing health care disparities across demographic, economic, and regional divides, and these technologies offer a means of access to high-quality health care for all [124]. By helping overcome structural barriers to quality health care, PHR technologies can potentially improve the health status of underserved communities and populations [69,124]. From a cost perspective, proactive care delivery made possible through these technologies can help in reducing the burden on public health institutions and resources [3,59,63,76].

Finally, from a population health research standpoint, consumer consent to sharing health care information and the subsequent widescale accumulation of PHR data have the potential to act as a valuable source of public health information for promoting healthy lifestyles and for detecting and preventing infectious diseases [119]. Through appropriate privacy and consent mechanisms, patient data available through an integrated health care network can be used to facilitate public health research on individuals and their communities as well as help with regional and global illness surveillance and screenings [55,90,121].

Overall, PHR systems can play a transformative role in facilitating complex information management processes across various health care delivery constituents. The mainstream deployment and adoption of these technologies has the potential to improve clinical and population health outcomes by streamlining medical and operational processes across the health care system.

Discussion

Although several previous studies on PHR technologies have alluded to a distinction among the functionality, utility, and value of these technologies [5,47,125], our review of the literature did not reveal any formal treatise of this subject at a theoretical or an empirical level. Our study aims to address this gap through a review and synthesis of the literature.

This paper presents a review of the extant literature on PHR systems, with the objective of providing an overall assessment of the functionality, utility, value, and benefits of contemporary PHRs. Toward this end, we offer a conceptual high-level functional utility model of PHRs outlining their features and functions categorized according to different use modalities. In addition, we deliberate on the value of PHR technologies to consumers by highlighting their applications across the spectrum of health care delivery activities. Finally, we provide a holistic summary of the value propositions and benefits of PHR systems to various health care industry constituents, including consumers, providers, payors, and public health agencies.

Our review indicates that PHR systems have made considerable progress over the past decade in terms of technology features and functions available at the product level. Compared with early PHR products that simply offer a basic functionality to maintain PHI [47,70], contemporary technologies offer a myriad of content-, connectivity-, and collaboration-based features and functions. Consequently, in recent years, the academic community has paid increasing attention to PHR functionality related to health monitoring tools, social networking features, and CDS functions.

From a value perspective, our analysis demonstrates that the value-generating potential of PHR systems arises from their role as an enabler for the integration of health care delivery activities across the full cycle of care for the health care consumer. These technologies can offer a useful mechanism for information exchange and care coordination among providers, thus leading to improved health outcomes for consumers. The consumer value framework conceptualized in this paper highlights that PHR functions have the potential to enhance patient experience through various touchpoints in health care delivery.

Our review also shows that although *consumers* are the primary beneficiaries of the functionality provided by PHR technologies, their overall value and benefits span across the activities and constituents of the health care delivery chain. At the consumer level, PHR systems can facilitate improved consumer health outcomes through the self-management of health and wellness as well as through enhanced quality of care. Moreover, these technologies can also generate channel partner value for providers and payors by enabling operational efficiency and reducing the cost of care. Finally, long-term and effective use of PHRs can also produce societal value in the form of improved public health outcomes.

This study offers several opportunities for research and potential practical applications. In terms of future research directions, we encourage researchers to undertake an empirical assessment of

our conceptualized functional utility and consumer value frameworks for PHR technologies. In particular, our literature review indicates a significant dearth of studies addressing the issue of consumer value of PHR offerings. Our study offers a possible starting point for this type of research. For health care practice, our review may be relevant to health care professionals associated with value analysis committees that are commissioned for the appraisal and recommendation of innovative HITs. A value framework such as the one proposed in this paper that integrates functional attributes, use cases, and applications in health care delivery activities can potentially be applied to the value assessment of other HITs as well.

Acknowledgments

The research reported here was supported by the Telfer School of Management Research Fund (SMRF grant) at the University of Ottawa.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Examples of personal health record use cases in the health care delivery value chain.
[PDF File (Adobe PDF File), 165 KB - jmir_v23i4e26877_app1.pdf]

References

1. Woods SS, Schwartz E, Tuepker A, Press NA, Nazi KM, Turvey CL, et al. Patient experiences with full electronic access to health records and clinical notes through the My HealtheVet Personal Health Record Pilot: qualitative study. J Med Internet Res 2013;15(3):e65 [FREE Full text] [doi: 10.2196/jmir.2356] [Medline: 23535584]
2. Young-Mason J. Information and knowledge is power: the online personal health record. Clin Nurse Spec 2013;27(2):105-106. [doi: 10.1097/NUR.0b013e3182818fea] [Medline: 23392069]
3. Greenhalgh T, Hinder S, Stramer K, Bratan T, Russell J. Adoption, non-adoption, and abandonment of a personal electronic health record: case study of HealthSpace. Br Med J 2010 Nov 16;341(nov16 1):5814 [FREE Full text] [doi: 10.1136/bmj.c5814] [Medline: 21081595]
4. Gartrell K, Storr CL, Trinkoff AM, Wilson ML, Gurses AP. Electronic personal health record use among registered nurses. Nurs Outlook 2015 May;63(3):278-287 [FREE Full text] [doi: 10.1016/j.outlook.2014.11.013] [Medline: 25982768]
5. Kahn JS, Aulakh V, Bosworth A. What it takes: characteristics of the ideal personal health record. Health Aff (Millwood) 2009 Mar;28(2):369-376 [FREE Full text] [doi: 10.1377/hlthaff.28.2.369] [Medline: 19275992]
6. Wolfson E. The personal health record. Healthline Networks Inc. URL: http://www.healthline.com/hlc/personal-health-record?micrositeId=30 [accessed 2019-06-18]
7. Genitsaridi I, Kondylakis H, Koumakis L, Marias K, Tsiknakis M. Evaluation of personal health record systems through the lenses of EC research projects. Comput Biol Med 2015 Apr;59:175-185. [doi: 10.1016/j.compbiomed.2013.11.004] [Medline: 24315661]
8. Pinciroli F, Pagliari C. Understanding the evolving role of the personal health record. Comput Biol Med 2015 Apr;59:160-163. [doi: 10.1016/j.compbiomed.2015.02.008] [Medline: 25726437]
9. Ford EW, Hesse BW, Huerta TR. Personal health record use in the United States: forecasting future adoption levels. J Med Internet Res 2016 Mar;18(3):e73 [FREE Full text] [doi: 10.2196/jmir.4973] [Medline: 27030105]
10. Goldzweig CL. Pushing the envelope of electronic patient portals to engage patients in their care. Ann Intern Med 2012 Oct 02;157(7):525-526. [doi: 10.7326/0003-4819-157-7-201210020-00013] [Medline: 23027322]
11. Appari A, Eric JM, Anthony DL. Meaningful use of electronic health record systems and process quality of care: evidence from a panel data analysis of U.S. acute-care hospitals. Health Serv Res 2013 Apr;48(2 Pt 1):354-375 [FREE Full text] [doi: 10.1111/j.1475-6773.2012.01448.x] [Medline: 22816527]
12. Adler-Milstein J, Sarma N, Woskie LR, Jha AK. A comparison of how four countries use health IT to support care for people with chronic conditions. Health Aff (Millwood) 2014 Sep;33(9):1559-1566. [doi: 10.1377/hlthaff.2014.0424] [Medline: 25201660]
13. Demiris G. New era for the consumer health informatics research agenda. Health Sys 2017 Dec 19;1(1):13-16. [doi: 10.1057/hs.2012.7]
14. Smith AB, Odlum M, Sikka M, Bakken S, Kanter T. Patient perceptions of pre-implementation of Personal Health Records (PHRs): a qualitative study of people living with HIV in New York City. J HIV/AIDS Soc Serv 2012 Oct;11(4):406-423. [doi: 10.1080/15381501.2012.735166]
15. Torraco RJ. Writing integrative literature reviews: guidelines and examples. Hum Resour Dev Rev 2016 Jul 24;4(3):356-367. [doi: 10.1177/1534484305278283]

16. Yorks L. What we know, what we don't know, what we need to know—integrative literature reviews are research. Hum Resour Dev Rev 2008 Jun 01;7(2):139-141. [doi: 10.1177/1534484308316395]
17. Kemppainen V, Tossavainen K, Turunen H. Nurses' roles in health promotion practice: an integrative review. Health Promot Int 2013 Dec;28(4):490-501. [doi: 10.1093/heapro/das034] [Medline: 22888155]
18. Stormacq C, Van den Broucke S, Wosinski J. Does health literacy mediate the relationship between socioeconomic status and health disparities? Integrative review. Health Promot Int 2019 Oct 01;34(5):1-17. [doi: 10.1093/heapro/day062] [Medline: 30107564]
19. Whittemore R, Knafl K. The integrative review: updated methodology. J Adv Nurs 2005 Dec;52(5):546-553. [doi: 10.1111/j.1365-2648.2005.03621.x] [Medline: 16268861]
20. Doty DH, Glick WH. Typologies as a unique form of theory building: toward improved understanding and modeling. Acad Manag Rev 1994 Apr;19(2):230-251. [doi: 10.5465/amr.1994.9410210748]
21. Kaelber DC, Pan EC. The value of personal health record (PHR) systems. AMIA Annu Symp Proc 2008 Nov 06:343-347 [FREE Full text] [Medline: 18999276]
22. Vincent A, Kaelber DC, Kaelber DC, Pan E, Shah S, Shah SS, et al. A patient-centric taxonomy for personal health records (PHRs). AMIA Annu Symp Proc 2008 Nov 06:763-767 [FREE Full text] [Medline: 18998912]
23. Archer N, Fevrier-Thomas U, Lokker C, McKibbon KA, Straus SE. Personal health records: a scoping review. J Am Med Inform Assoc 2011 Jul;18(4):515-522 [FREE Full text] [doi: 10.1136/amiajnl-2011-000105] [Medline: 21672914]
24. Irizarry T, DeVito DA, Curran CR. Patient portals and patient engagement: a state of the science review. J Med Internet Res 2015 Jun;17(6):e148 [FREE Full text] [doi: 10.2196/jmir.4255] [Medline: 26104044]
25. Pushpangadan S, Seckman C. Consumer perspective on personal health records: a review of the literature. Online J Nurs Informatics. URL: https://www.himss.org/resources/consumer-perspective-personal-health-records-review-literature [accessed 2019-06-18]
26. Eysenbach G. Consumer health informatics. Br Med J 2000 Jun 24;320(7251):1713-1716 [FREE Full text] [doi: 10.1136/bmj.320.7251.1713] [Medline: 10864552]
27. Eysenbach G, Jadad AR. Evidence-based patient choice and consumer health informatics in the internet age. J Med Internet Res 2001 Jun;3(2):E19 [FREE Full text] [doi: 10.2196/jmir.3.2.e19] [Medline: 11720961]
28. Jadad A. Consumer health informatics: supporting the hightouch. Ottawa, Canada: Consumer Health Informatics Summit; 2007. URL: http://www.ehealthinformation.ca/wp-content/uploads/2014/07/2007-Consumer-Health-Informatics-Summit.pdf [accessed 2021-04-22]
29. Johnson K, Jimison HB, Mandl KD. Consumer health informatics personal health records. In: Biomedical Informatics. London: Springer; 2013:517-539.
30. Marceglia S, Fontelo P, Ackerman M. Transforming consumer health informatics: connecting CHI applications to the health-IT ecosystem. J Am Med Inform Assoc 2015 Apr;22(e1):210-212 [FREE Full text] [doi: 10.1093/jamia/ocu030] [Medline: 25665702]
31. Americans want benefits of personal health records. Markle. 2003. URL: https://www.markle.org/publications/950-americans-want-benefits-personal-health-records [accessed 2021-04-22]
32. Tang PC, Ash JS, Bates DW, Overhage JM, Sands DZ. Personal health records: definitions, benefits, and strategies for overcoming barriers to adoption. J Am Med Inform Assoc 2006 Mar;13(2):121-126 [FREE Full text] [doi: 10.1197/jamia.M2025] [Medline: 16357345]
33. Garrett P, Seidman J. EMR vs EHR? What is the Difference? Health IT Buzz. 2011. URL: https://www.healthit.gov/buzz-blog/electronic-health-and-medical-records/emr-vs-ehr-difference [accessed 2019-06-18]
34. Raisinghani MS, Young E. Personal health records: key adoption issues and implications for management. Int J Electron Healthc 2008;4(1):67. [doi: 10.1504/ijeh.2008.018921]
35. Hodge T, Giokas D. EMR, EHR and PHR? Why all the confusion? Canada Health Infoway. 2011. URL: https://www.infoway-inforoute.ca/en/what-we-do/blog/digital-health-records/6852-emr-ehr-and-phr-why-all-the-confusion [accessed 2019-06-18]
36. Silow-Carroll S, Edwards JN, Rodin D. Using electronic health records to improve quality and efficiency: the experiences of leading hospitals. Issue Brief (Commonw Fund) 2012 Jul;17:1-40. [Medline: 22826903]
37. Detmer D, Bloomrosen M, Raymond B, Tang P. Integrated personal health records: transformative tools for consumer-centric care. BMC Med Inform Decis Mak 2008 Oct 06;8:45 [FREE Full text] [doi: 10.1186/1472-6947-8-45] [Medline: 18837999]
38. Kern J, Fister K, Polasek O. Active patient role in recording health data. In: Encyclopedia of Information Science and Technology, Second Edition. Hershey, Pennsylvania, United States: IGI Global; 2009:14-19.
39. Hoerbst A, Kohl CD, Knaup P, Ammenwerth E. Attitudes and behaviors related to the introduction of electronic health records among Austrian and German citizens. Int J Med Inform 2010 Feb;79(2):81-89. [doi: 10.1016/j.ijmedinf.2009.11.002] [Medline: 20031482]
40. Mon DT, EHR Work Group. HL7 PHR system functional model. In Healthcare Information Management Systems Society (HIMSS) Conference, New Orleans, LA, USA. 2013. URL: https://www.hl7.org/documentcenter/public/calendarofevents/himss/2013/Personal%20Health%20Record%20System%20Functional%20Model.pdf [accessed 2019-06-18]

41. PHR-S FM Personal Health Record System Functional Model (PHR-S FM). Health Level Seven International. URL: https://www.hl7.org/implement/standards/product_brief.cfm?product_id=88 [accessed 2019-06-18]
42. Endsley S, Kibbe DC, Linares A, Colorafi K. An introduction to personal health records. Fam Pract Manag 2006 May;13(5):57-62 [FREE Full text] [Medline: 16736906]
43. Anoshiravani A, Gaskin G, Kopetsky E, Sandborg C, Longhurst CA. Implementing an Interoperable Personal Health Record in Pediatrics: Lessons Learned at an Academic Children's Hospital. J Particip Med 2011 Jul 10;3:- [FREE Full text] [Medline: 21853160]
44. Gaskin GL, Longhurst C, Slayton R, Das A. Sociotechnical challenges of developing an interoperable personal health record. Appl Clin Inform 2017 Dec 16;02(04):406-419. [doi: 10.4338/aci-2011-06-ra-0035]
45. Barbarito F, Pinciroli F, Barone A, Pizzo F, Ranza R, Mason J, et al. Implementing the lifelong personal health record in a regionalised health information system: the case of Lombardy, Italy. Comput Biol Med 2015 Apr;59:164-174. [doi: 10.1016/j.compbiomed.2013.10.021] [Medline: 24238969]
46. Kikuchi S, Sachdeva S, Bhalla S. Cloud computing based PHR architecture using multi layers model. J Softw Eng Appl 2012;05(11):903-911. [doi: 10.4236/jsea.2012.531105]
47. Kim MI, Johnson K. Personal health records: evaluation of functionality and utility. J Am Med Inform Assoc 2002;9(2):171-180 [FREE Full text] [doi: 10.1197/jamia.m0978] [Medline: 11861632]
48. Cabitza F, Simone C, De Michelis G. User-driven prioritization of features for a prospective interpersonal health record: perceptions from the Italian context. Comput Biol Med 2015 Apr;59:202-210. [doi: 10.1016/j.compbiomed.2014.03.009] [Medline: 24768267]
49. Bachiri M, Idri A, Fernández-Alemán JL, Toval A. Mobile personal health records for pregnancy monitoring functionalities: analysis and potential. Comput Methods Programs Biomed 2016 Oct;134:121-135. [doi: 10.1016/j.cmpb.2016.06.008] [Medline: 27480737]
50. Emani S, Yamin CK, Peters E, Karson AS, Lipsitz SR, Wald JS, et al. Patient perceptions of a personal health record: a test of the diffusion of innovation model. J Med Internet Res 2012 Nov;14(6):e150 [FREE Full text] [doi: 10.2196/jmir.2278] [Medline: 23128775]
51. Schnipper JL, Gandhi TK, Wald JS, Grant RW, Poon EG, Volk LA, et al. Effects of an online personal health record on medication accuracy and safety: a cluster-randomized trial. J Am Med Inform Assoc 2012 Sep;19(5):728-734 [FREE Full text] [doi: 10.1136/amiajnl-2011-000723] [Medline: 22556186]
52. Weppner WG, Ralston JD, Koepsell TD, Grothaus LC, Reid RJ, Jordan L, et al. Use of a shared medical record with secure messaging by older patients with diabetes. Diabetes Care 2010 Nov;33(11):2314-2319 [FREE Full text] [doi: 10.2337/dc10-1124] [Medline: 20739686]
53. Tom JO, Mangione-Smith R, Solomon C, Grossman DC. Integrated personal health record use: association with parent-reported care experiences. Pediatrics 2012 Jul 11;130(1):183-190. [doi: 10.1542/peds.2011-1786] [Medline: 22689872]
54. Islam R, Kikuchi K, Sato Y, Izukura R, Yokota F, Nishikitani M, et al. Personal Health Record (PHR) system in portable health clinic. Stud Health Technol Inform 2020 Jun 16;270:1347-1348. [doi: 10.3233/SHTI200435] [Medline: 32570652]
55. Heart T, Ben-Assuli O, Shabtai I. A review of PHR, EMR and EHR integration: a more personalized healthcare and public health policy. Health Policy Technol 2017 Mar;6(1):20-25. [doi: 10.1016/j.hlpt.2016.08.002]
56. Seo D, Park YR, Lee Y, Kim JY, Park J, Lee J. The use of mobile personal health records for hemoglobin A1c regulation in patients with diabetes: retrospective observational study. J Med Internet Res 2020 Jun 02;22(6):e15372 [FREE Full text] [doi: 10.2196/15372] [Medline: 32484447]
57. Kipping S, Stuckey MI, Hernandez A, Nguyen T, Riahi S. A web-based patient portal for mental health care: benefits evaluation. J Med Internet Res 2016 Nov 16;18(11):e294 [FREE Full text] [doi: 10.2196/jmir.6483] [Medline: 27852556]
58. Kibbe DC, Phillips RL, Green LA. The continuity of care record. Am Fam Physician 2004 Oct 01;70(7):1220, 2-1220, 3 [FREE Full text] [Medline: 15508532]
59. Folker G. The chronic need for connectivity: helping today's aging heath care consumers help themselves. Consumer Health Informatics Summit. Ottawa, Canada: Consumer Health Informatics Summit; 2007. URL: http://www.ehealthinformation.ca/wp-content/uploads/2014/07/2007-Consumer-Health-Informatics-Summit.pdf [accessed 2021-04-22]
60. Ralston JD, Carrell D, Reid R, Anderson M, Moran M, Hereford J. Patient web services integrated with a shared medical record: patient use and satisfaction. J Am Med Inform Assoc 2007 Nov;14(6):798-806 [FREE Full text] [doi: 10.1197/jamia.M2302] [Medline: 17712090]
61. Kim J, Ryu B, Cho S, Heo E, Kim Y, Lee J, et al. Impact of personal health records and wearables on health outcomes and patient response: three-arm randomized controlled trial. JMIR Mhealth Uhealth 2019 Jan 04;7(1):e12070 [FREE Full text] [doi: 10.2196/12070] [Medline: 30609978]
62. Slevin P, Caulfield B. Patient-generated health data: looking toward future health care. In: Wearable Technology in Medicine and Health Care. Amsterdam: Elsevier; 2018:261-273.
63. Ball MJ, Smith C, Bakalar RS. Personal health records: empowering consumers. J Healthc Inf Manag 2007;21(1):76-86. [Medline: 17299929]

64. Mercer JG. Using web based tools to help patients achieve optimal clinical outcomes. Consumer Health Informatics Summit. Ottawa, Canada: Consumer Health Informatics Summit; 2007. URL: http://www.ehealthinformation.ca/wp-content/uploads/2014/07/2007-Consumer-Health-Informatics-Summit.pdf [accessed 2021-04-22]
65. Terry M. The personal health dashboard: consumer electronics is growing in the health and wellness market. Telemed J E Health 2009 Sep;15(7):642-645. [doi: 10.1089/tmj.2009.9947] [Medline: 19747072]
66. Hefner JL, MacEwan SR, Biltz A, Sieck CJ. Patient portal messaging for care coordination: a qualitative study of perspectives of experienced users with chronic conditions. BMC Fam Pract 2019 May 03;20(1):57 [FREE Full text] [doi: 10.1186/s12875-019-0948-1] [Medline: 31053063]
67. Plate JF, Ryan SP, Bergen MA, Hong CS, Attarian DE, Seyler TM. Utilization of an Electronic Patient Portal Following Total Joint Arthroplasty Does Not Decrease Readmissions. J Arthroplasty 2019 Feb;34(2):211-214. [doi: 10.1016/j.arth.2018.11.002] [Medline: 30497899]
68. Hill JN, Smith BM, Weaver FM, Nazi KM, Thomas FP, Goldstein B, et al. Potential of personal health record portals in the care of individuals with spinal cord injuries and disorders: provider perspectives. J Spinal Cord Med 2018 May 21;41(3):298-308 [FREE Full text] [doi: 10.1080/10790268.2017.1293760] [Medline: 28325112]
69. Kai E, Ahmed A. Technical challenges in providing remote health consultancy services for the unreached community. In: Proceedings of the 27th International Conference on Advanced Information Networking and Applications Workshops. 2013 Presented at: 27th International Conference on Advanced Information Networking and Applications Workshops; March 25-28, 2013; Barcelona, Spain p. 1016-1020. [doi: 10.1109/waina.2013.227]
70. Halamka JD, Mandl KD, Tang PC. Early experiences with personal health records. J Am Med Inform Assoc 2008 Jan;15(1):1-7 [FREE Full text] [doi: 10.1197/jamia.M2562] [Medline: 17947615]
71. Frost JH, Massagli MP. Social uses of personal health information within PatientsLikeMe, an online patient community: what can happen when patients have access to one another's data. J Med Internet Res 2008;10(3):e15 [FREE Full text] [doi: 10.2196/jmir.1053] [Medline: 18504244]
72. Bickmore TW, Schulman D, Sidner CL. A reusable framework for health counseling dialogue systems based on a behavioral medicine ontology. J Biomed Inform 2011 Apr;44(2):183-197 [FREE Full text] [doi: 10.1016/j.jbi.2010.12.006] [Medline: 21220044]
73. Goodman K. Decision support the personal health record. URL: http://www.projecthealthdesign.org/media/file/ELSI-decision-support.pdf [accessed 2019-06-18]
74. Davis S, MacKay L. Moving beyond the rhetoric of shared decision-making: designing personal health record technology with young adults with Type 1 diabetes. Can J Diabetes 2020 Jul;44(5):434-441. [doi: 10.1016/j.jcjd.2020.03.009] [Medline: 32616277]
75. Schera F, Schäfer M, Bucur A, van Leeuwen J, Ngantchjon EH, Graf N, et al. iManageMyHealth and iSupportMyPatients: mobile decision support and health management apps for cancer patients and their doctors. Ecancermedicalscience 2018 Jul 11;12:848 [FREE Full text] [doi: 10.3332/ecancer.2018.848] [Medline: 30079110]
76. Bateja R, Dubey S, Bhatt A. A patient-centric healthcare model based on health recommender systems. In: Recent Findings in Intelligent Computing Techniques. Singapore: Springer; 2018:269-276.
77. Sequist TD, Zaslavsky AM, Colditz GA, Ayanian JZ. Electronic patient messages to promote colorectal cancer screening: a randomized controlled trial. Arch Intern Med 2011 Apr 11;171(7):636-641 [FREE Full text] [doi: 10.1001/archinternmed.2010.467] [Medline: 21149743]
78. Sykes TA, Venkatesh V, Johnson JL. Enterprise system implementation and employee job performance: understanding the role of advice networks. MIS Q 2014 Jan 1;38(1):51-72. [doi: 10.25300/misq/2014/38.1.03]
79. Grant RW, Wald JS, Schnipper JL, Gandhi TK, Poon EG, Orav EJ, et al. Practice-linked online personal health records for type 2 diabetes mellitus: a randomized controlled trial. Arch Intern Med 2008 Sep 08;168(16):1776-1782 [FREE Full text] [doi: 10.1001/archinte.168.16.1776] [Medline: 18779465]
80. Krist AH, Woolf SH, Bello GA, Sabo RT, Longo DR, Kashiri P, et al. Engaging primary care patients to use a patient-centered personal health record. Ann Fam Med 2014 Sep;12(5):418-426 [FREE Full text] [doi: 10.1370/afm.1691] [Medline: 25354405]
81. Porter ME, Teisberg EO. Redefining health care: creating value-based competition on results. National Association of Chain Drug Stores. 2006. URL: https://www.hbs.edu/ris/Publication%20Files/20060502%20NACDS%20-%20Final%2005012006%20for%20On%20Point_db5ede1d-3d06-41f0-85e3-c11658534a63.pdf [accessed 2021-04-22]
82. Porter ME. Value-based health care delivery. Ann Surg 2008 Oct;248(4):503-509. [doi: 10.1097/SLA.0b013e31818a43af] [Medline: 18936561]
83. Nazi KM. The personal health record paradox: health care professionals' perspectives and the information ecology of personal health record systems in organizational and clinical settings. J Med Internet Res 2013 Apr;15(4):e70 [FREE Full text] [doi: 10.2196/jmir.2443] [Medline: 23557596]
84. Porter ME. What is value in health care? N Engl J Med 2010 Dec 23;363(26):2477-2481. [doi: 10.1056/nejmp1011024]

85. Hess R, Bryce CL, Paone S, Fischer G, McTigue KM, Olshansky E, et al. Exploring challenges and potentials of personal health records in diabetes self-management: implementation and initial assessment. Telemed J E Health 2007 Oct;13(5):509-517. [doi: 10.1089/tmj.2006.0089] [Medline: 17999613]
86. Wang CK. Security and privacy of personal health record, electronic medical record and health information. Probl Perspect Manag. 2015. URL: https://businessperspectives.org/images/pdf/applications/publishing/templates/article/assets/6962/PPM_2015_04_Wang.pdf [accessed 2019-06-18]
87. Zulman DM, Nazi KM, Turvey CL, Wagner TH, Woods SS, An LC. Patient interest in sharing personal health record information: a web-based survey. Ann Intern Med 2011 Dec 20;155(12):805-810. [doi: 10.7326/0003-4819-155-12-201112200-00002] [Medline: 22184687]
88. Williams JB, Weber-Jahnke JH. The regulation of personal health records in Canada. Can J Law Technol. URL: https://ojs.library.dal.ca/CJLT/article/download/4944/4452 [accessed 2019-06-18]
89. Manard W, Scherrer JF, Salas J, Schneider FD. Patient portal use and blood pressure control in newly diagnosed hypertension. J Am Board Fam Med 2016 Jul;29(4):452-459 [FREE Full text] [doi: 10.3122/jabfm.2016.04.160008] [Medline: 27390376]
90. Rothman B, Leonard JC, Vigoda MM. Future of electronic health records: implications for decision support. Mt Sinai J Med 2012 Dec 13;79(6):757-768. [doi: 10.1002/msj.21351] [Medline: 23239213]
91. Hess R. The diabetes patient portal: patient perspectives on structure and delivery. Diabetes Spectrum 2006 Apr 01;19(2):106-110. [doi: 10.2337/diaspect.19.2.106]
92. King DK, Toobert DJ, Portz JD, Strycker LA, Doty A, Martin C, et al. What patients want: relevant health information technology for diabetes self-management. Health Technol 2012 Mar 5;2(3):147-157. [doi: 10.1007/s12553-012-0022-7]
93. Winkelman WJ. Patient-perceived usefulness of online electronic medical records: employing grounded theory in the development of information and communication technologies for use by patients living with chronic illness. J Am Med Informatics Assoc 2005 Jan 31;12(3):306-314. [doi: 10.1197/jamia.m1712]
94. Agarwal R, Anderson C, Zarate J, Ward C. If we offer it, will they accept? Factors affecting patient use intentions of personal health records and secure messaging. J Med Internet Res 2013 Feb;15(2):e43 [FREE Full text] [doi: 10.2196/jmir.2243] [Medline: 23470453]
95. Rizk S. Utility of a comprehensive Personal Health Record (PHR) strategy. Qatar Foundation Annual Research Forum Proceedings 2013 Nov(2013):ICTO 011. [doi: 10.5339/qfarf.2013.ICTO-011]
96. Dorland's Medical Dictionary Online. URL: https://www.dorlandsonline.com/dorland/home [accessed 2021-04-22]
97. Burns LR. The business of healthcare innovation in the Wharton school curriculum. The Business of Healthcare Innovation 2012:1-24. [doi: 10.1017/CBO9781139176620.001]
98. Martineau M. The internet changes everything: Lessons from other industries. Ottawa, Canada: Consumer Health Informatics Summit; 2007. URL: http://www.ehealthinformation.ca/wp-content/uploads/2014/07/2007-Consumer-Health-Informatics-Summit.pdf [accessed 2021-04-22]
99. Ross SE, Moore LA, Earnest MA, Wittevrongel L, Lin C. Providing a web-based online medical record with electronic communication capabilities to patients with congestive heart failure: randomized trial. J Med Internet Res 2004 May 14;6(2):e12 [FREE Full text] [doi: 10.2196/jmir.6.2.e12] [Medline: 15249261]
100. Faisal S, Blandford A, Potts HW. Making sense of personal health information: challenges for information visualization. Health Informatics J 2013 Sep;19(3):198-217. [doi: 10.1177/1460458212465213] [Medline: 23981395]
101. McDaniel AM, Schutte DL, Keller LO. Consumer health informatics: from genomics to population health. Nurs Outlook 2008 Sep;56(5):216-223. [doi: 10.1016/j.outlook.2008.06.006] [Medline: 18922271]
102. Brennan PF, Downs S, Casper G. Project HealthDesign: rethinking the power and potential of personal health records. J Biomed Inform 2010 Oct;43(5 Suppl):S3-S5 [FREE Full text] [doi: 10.1016/j.jbi.2010.09.001] [Medline: 20937482]
103. Demiris G, Afrin LB, Speedie S, Courtney KL, Sondhi M, Vimarlund V, et al. Patient-centered applications: use of information technology to promote disease management and wellness. A white paper by the AMIA knowledge in motion working group. J Am Med Inform Assoc 2008 Jan;15(1):8-13 [FREE Full text] [doi: 10.1197/jamia.M2492] [Medline: 17947617]
104. Williams J, Weber-Jahnke J. Regulation of patient management software. Health Law J 2010;18:73-111. [Medline: 22125972]
105. Hassol A, Walker JM, Kidder D, Rokita K, Young D, Pierdon S, et al. Patient experiences and attitudes about access to a patient electronic health care record and linked web messaging. J Am Med Inform Assoc 2004 Nov;11(6):505-513 [FREE Full text] [doi: 10.1197/jamia.M1593] [Medline: 15299001]
106. Bates DW, Bitton A. The future of health information technology in the patient-centered medical home. Health Aff (Millwood) 2010 Apr;29(4):614-621. [doi: 10.1377/hlthaff.2010.0007] [Medline: 20368590]
107. AbuOun NA, Abdel-Hamid A, El-Nasr MA. The concept of a medical information system by rate of use (CHEBHOSP). Heal Informatics Int J 2015 May 31;4(1/2):1-11. [doi: 10.5121/hiij.2015.4201]
108. Studeny J, Coustasse A. Personal health records: is rapid adoption hindering interoperability? Perspect Health Inf Manag 2014;11:1e [FREE Full text] [Medline: 25214822]
109. AbuOun NA, Abdel-Hamid A, El-Nasr MA. A novel framework for electronic global health record access. Health Informatic Int J. 2015 May 31. URL: http://www.airccse.org/journal/hiij/papers/4215hiij01.pdf [accessed 2021-04-26]

110. Walker J, Pan E, Johnston D, Adler-Milstein J, Bates DW, Middleton B. The value of health care information exchange and interoperability. Health Aff (Millwood) 2005 Jan;Suppl Web Exclusives:5-10. [doi: 10.1377/hlthaff.w5.10] [Medline: 15659453]
111. Wuerdeman L, Volk L, Pizziferri L, Tsurikova R, Harris C, Feygin R, et al. How accurate is information that patients contribute to their Electronic Health Record? AMIA Annu Symp Proc 2005:834-838 [FREE Full text] [Medline: 16779157]
112. Nippak PM, Isaac WW, Geertsen A, Ikeda-Douglas CJ. Family attitudes towards an electronic personal health record in a long term care facility. J Hosp Adm 2015 Mar 22;4(3):9. [doi: 10.5430/jha.v4n3p9]
113. Veronin M. Enhancing personal health record adoption through the community pharmacy network: a service project. Innov Pharm 2015 Jan 01;6(4):-. [doi: 10.24926/iip.v6i4.403]
114. Bates DW, Wells S. Personal health records and health care utilization. J Am Med Assoc 2012 Nov 21;308(19):2034-2036. [doi: 10.1001/jama.2012.68169] [Medline: 23168828]
115. Fricton JR, Davies D. Personal health records to improve health information exchange and patient safety. In: Advances in Patient Safety: New Directions and Alternative Approaches (Vol. 4: Technology and Medication Safety). Bethesda MD, USA: National Center for Biotechnology Information, U.S. National Library of Medicine; 2008:1-10.
116. Menon AS, Greenwald S, Ma TJ, Kooshesh S, Duriseti R. Patient and physician willingness to use personal health records in the emergency department. West J Emerg Med 2012 May 1;13(2):172-175 [FREE Full text] [doi: 10.5811/westjem.2011.11.6844] [Medline: 22900108]
117. Cavoukian A, Rossos PG. Personal Health Information: A Practical Tool for Physicians Transitioning from Paper-Based Records to Electronic Health Records. Ontario, Canada: Information and Privacy Commissioner of Ontario; 2009. URL: https://www.ipc.on.ca/wp-content/uploads/Resources/phipa-toolforphysicians.pdf [accessed 2019-06-18]
118. Garrido T, Kanter M, Meng D, Turley M, Wang J, Sue V, et al. Race/ethnicity, personal health record access, and quality of care. Am J Manag Care 2015 Feb 01;21(2):103-113 [FREE Full text] [Medline: 25880485]
119. Weitzman ER, Kelemen S, Kaci L, Mandl KD. Willingness to share personal health record data for care improvement and public health: a survey of experienced personal health record users. BMC Med Inform Decis Mak 2012 May 22;12:39 [FREE Full text] [doi: 10.1186/1472-6947-12-39] [Medline: 22616619]
120. Bouri N, Ravi S. Going mobile: how mobile personal health records can improve health care during emergencies. JMIR Mhealth Uhealth 2014 Mar;2(1):e8 [FREE Full text] [doi: 10.2196/mhealth.3017] [Medline: 25098942]
121. Tomines A, Readhead H, Readhead A, Teutsch S. EGEMS (Wash DC) 2013 Oct 28;1(2):1019 [FREE Full text] [doi: 10.13063/2327-9214.1019] [Medline: 25848571]
122. Kostadinovska A, de Vries GJ, Geleijnse G, Zdravkova K. Employing personal health records for population health management. In: ICT Innovations 2014. Switzerland: Springer International Publishing; 2015:65-74.
123. Caligtan CA, Dykes PC. Electronic health records and personal health records. Semin Oncol Nurs 2011 Aug;27(3):218-228. [doi: 10.1016/j.soncn.2011.04.007] [Medline: 21783013]
124. Garvin J, Odom-Wesley B, Rudman WJ, Stewart RS. Healthcare disparities and the role of personal health records. The American Health Information Management Association. 2009. URL: https://library.ahima.org/doc?oid=91677#.YIEv5-gzayI [accessed 2021-04-22]
125. Gysels M, Richardson A, Higginson IJ. Does the patient-held record improve continuity and related outcomes in cancer care: a systematic review. Health Expect 2007 Mar;10(1):75-91 [FREE Full text] [doi: 10.1111/j.1369-7625.2006.00415.x] [Medline: 17324196]

Abbreviations

CDS: clinical decision support
CDVC: care delivery value chain
EHR: electronic health record
EMR: electronic medical record
HIS: health information system
HIT: health information technology
PHI: personal health information
PHR: personal health record
PMRI: patient medical record information

Edited by G Eysenbach; submitted 31.12.20; peer-reviewed by HL Patterson, S Riahi; comments to author 25.01.21; revised version received 30.01.21; accepted 15.04.21; published 29.04.21.

Please cite as:
Ruhi U, Chugh R
Utility, Value, and Benefits of Contemporary Personal Health Records: Integrative Review and Conceptual Synthesis
J Med Internet Res 2021;23(4):e26877
URL: https://www.jmir.org/2021/4/e26877
doi:10.2196/26877
PMID:33866308

Review

Exploring the Role of Persuasive Design in Unguided Internet-Delivered Cognitive Behavioral Therapy for Depression and Anxiety Among Adults: Systematic Review, Meta-analysis, and Meta-regression

Hugh C McCall[1,2], MA; Heather D Hadjistavropoulos[1,2], PhD; Christopher Richard Francis Sundström[3,4], PhD

[1]Department of Psychology, University of Regina, Regina, SK, Canada

[2]PSPNET, University of Regina, Regina, SK, Canada

[3]Department of Clinical Neuroscience, Centre for Psychiatry Research, Karolinska Institutet, Stockholm, Sweden

[4]Department of Psychology, Stockholm University, Stockholm, Sweden

Corresponding Author:
Hugh C McCall, MA
Department of Psychology
University of Regina
3737 Wascana Pkwy
Regina, SK, S4S 0A2
Canada
Phone: 1 306 585 4111
Email: hugh.c.mccall@gmail.com

Abstract

Background: Internet-delivered cognitive behavioral therapy (ICBT) is an effective treatment that can overcome barriers to mental health care. Various research groups have suggested that unguided ICBT (ie, ICBT without therapist support) and other eHealth interventions can be designed to enhance user engagement and thus outcomes. The persuasive systems design framework captures most design recommendations for eHealth interventions, but there is little empirical evidence that persuasive design is related to clinical outcomes in unguided ICBT.

Objective: This study aims to provide an updated meta-analysis of randomized controlled trials of unguided ICBT for depression and anxiety, describe the frequency with which various persuasive design principles are used in such interventions, and use meta-regression to explore whether a greater number of persuasive design elements predicts efficacy in unguided ICBT for depression and anxiety.

Methods: We conducted a systematic review of 5 databases to identify randomized controlled trials of unguided ICBT for depression and anxiety. We conducted separate random effects meta-analyses and separate meta-regressions for depression and anxiety interventions. Each meta-regression included 2 steps. The first step included, as a predictor, whether each intervention was transdiagnostic. For the meta-regression of ICBT for depression, the first step also included the type of control condition. The number of persuasive design principles identified for each intervention was added as a predictor in the second step to reveal the additional variance in effect sizes explained by persuasive design.

Results: Of the 4471 articles we identified in our search, 46 (1.03%) were eligible for inclusion in our analyses. Our meta-analyses showed effect sizes (Hedges *g*) ranging from 0.22 to 0.31 for depression interventions, depending on the measures taken to account for bias in the results. We found a mean effect size of 0.45 (95% CI 0.33-0.56) for anxiety interventions, with no evidence that the results were inflated by bias. Included interventions were identified as using between 1 and 13 persuasive design principles, with an average of 4.95 (SD 2.85). The meta-regressions showed that a greater number of persuasive design principles predicted greater efficacy in ICBT for depression (R^2 change=0.27; B=0.04; P=.02) but not anxiety (R^2 change=0.05; B=0.03; P=.17).

Conclusions: These findings show wide variability in the use of persuasive design in unguided ICBT for depression and anxiety and provide preliminary support for the proposition that more persuasively designed interventions are more efficacious, at least in the treatment of depression. Further research is needed to clarify the role of persuasive design in ICBT.

(J Med Internet Res 2021;23(4):e26939) doi:10.2196/26939

KEYWORDS

ICBT; internet; depression; anxiety; persuasive design; eHealth

Introduction

Background

Depression and anxiety are highly prevalent and represent the leading and the sixth leading causes of disability worldwide, respectively [1]. Despite the demonstrated efficacy of psychotherapeutic and pharmacological interventions for depression and anxiety [2-4], many people face structural barriers to accessing mental health care (eg, financial barriers, transportation barriers, inconvenience, and limited availability of services) [5,6]. Internet-delivered cognitive behavioral therapy (ICBT) is the most common type of internet intervention and an effective treatment for several common mental health problems, including depression and anxiety [7]. Unlike traditional cognitive behavioral therapy (CBT), ICBT enables users to access treatment materials privately at a time and location that is convenient for them, allowing it to be administered economically on a large scale and circumvent barriers to traditional forms of mental health care [8-10]. ICBT can be therapist guided or unguided. Guidance appears to improve adherence and clinical outcomes [11], but unguided ICBT is economical, highly scalable, and believed by many researchers to have considerable potential for improving public health [12-15].

Since the early 2000s, various research groups have suggested that eHealth interventions such as unguided ICBT can be designed in ways that improve user engagement and thus outcomes. In 2003, Fogg [16] presented the *functional triad* principle, suggesting that technology can function as a tool, a medium for relaying content, and a social actor to help facilitate behavior change. In 2009, Oinas-Kukkonen and Harjumaa [17] developed the *persuasive systems design* (PSD) framework, which elaborated on the functional triad and included 28 recommended design principles to produce more persuasive and engaging technological systems. They divided these principles into 4 categories: (1) *primary task support* principles, which facilitate the completion of the primary tasks of an intervention or other system; (2) *dialogue support* principles, through which an intervention or other system supports a user to help them enact their target behavior; (3) *system credibility support* principles, which facilitate a more credible and persuasive intervention or other system; and (4) *social support* principles, which leverage principles of social psychology to help users of an intervention or other system motivate one another. The 28 principles are described in Multimedia Appendix 1 [17].

Several other research groups have provided their own design recommendations for eHealth interventions. Despite using different terminology, most of these recommendations appear to align closely with the principles included in the PSD framework. Examples include recommendations related to *personalization* [18-22], *tailoring* [19,21,22], *reminders* [19,20], *self-monitoring* [18,23], *liking* [19,22,24], and various *dialogue support* principles [18,19,23]. A few design recommendations are not captured in the PSD framework (eg, time-limited access [20] and greater use of metaphors [22]), but to our knowledge, none of these have been proposed by 2 or more research groups; that is, the PSD framework appears to capture most common recommendations. Various groups' recommendations and the related PSD framework principles are displayed in Multimedia Appendix 2 [18-24].

In 2012, Kelders et al [25] used the PSD framework to assess whether the persuasive design principles used in 83 eHealth interventions for chronic conditions, lifestyle changes, and mental health predicted adherence. They conducted a meta-regression, finding that a greater number of dialogue support principles predicted greater adherence to eHealth interventions. However, to our knowledge, there is no empirical research demonstrating a relationship between persuasive design and symptom change in eHealth interventions.

Objectives and Hypothesis

This study aims to (1) present a systematic review and meta-analysis of randomized controlled trials of unguided ICBT for depression and anxiety among adults, (2) systematically examine the frequency with which various persuasive design principles are used in such interventions, and (3) use meta-regression to examine the extent to which persuasive design could explain the variability in effect sizes identified through the meta-analysis. Thus, the overarching objective of this study is to review the efficacy, the use of persuasive design, and the relationship between efficacy and persuasive design in unguided ICBT for depression and anxiety. We hypothesized that using a greater number of persuasive design principles would predict greater efficacy among the included studies.

Methods

Study Design

This study consisted of a systematic review, 2 meta-analyses, and 2 meta-regressions. The methods used in each phase of the study are described in the following sections. We registered the methodological protocol for this study on PROSPERO on October 24, 2019 (ID: 153466), before commencing the literature search, and kept a log of revisions to the original protocol throughout the course of this research (Multimedia Appendix 3 [26-28]). We followed the guidelines outlined in the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) statement in the preparation of this paper [29].

Systematic Review Methods

Eligibility Criteria

We searched for randomized controlled trials of unguided ICBT interventions for symptoms of depression and/or anxiety among adults that had been published in English in academic journals since 2000. We included trials of ICBT targeting symptoms of any type of depressive or anxiety disorder, as defined in the fifth edition of the Diagnostic and Statistical Manual of Mental

Disorders [30], with various kinds of control conditions (eg, waitlist, treatment as usual, and active control). Studies involving samples with a mean age of less than 18 years were excluded.

Although we excluded studies in which ICBT was delivered with guidance from a therapist or coach, we did not exclude studies involving diagnostic interviews or contact of a logistical nature between participants and research teams. Interventions that used a CBT model of treatment and were delivered via the internet were considered ICBT interventions regardless of whether the authors of trials identified them as such. We included interventions using third-wave CBT approaches (eg, mindfulness-based CBT and acceptance and commitment therapy) [31] because prior research has not demonstrated significant differences in outcomes between traditional CBT and third-wave approaches [32,33].

Literature Search

On October 29, 2019, we conducted a literature search on MEDLINE, PsycINFO, PubMed, Web of Science, and PsycArticles. To be identified, articles were required to include the words "CBT," "internet," "trial," and "depression" or "anxiety" or one of several similar phrases for each of these terms in their titles, keywords, or abstracts. The search terms are shown in detail in Multimedia Appendix 4. This search was updated on July 2, 2020.

Study Selection

After removing duplicates of studies identified in 2 or more databases, HCM and CRFS independently screened the studies in 3 stages: by title, by abstract, and by full text. Wherever the 2 screeners reached different decisions about whether to retain or exclude a study, that study was included in the next stage of screening. Differences in decisions on the full-text screening were resolved through discussion.

Data Extraction

HCM extracted several types of data from each study: study characteristics (eg, type of control condition and time between pretreatment and posttreatment measures), risk of bias [34], general intervention characteristics (eg, target symptoms and medium of delivery), persuasive design principles [17] as operationalized by Kelders et al [25], and efficacy data. Consistent with the approach of Kelders et al [25], we did not code principles in the *system credibility support* category of the PSD framework because they were reported very infrequently and would have been challenging to code objectively (eg, a system should have a "competent look and feel" and "provide endorsements from respected sources" [17]). In most cases, we coded persuasive design principles as present or absent based on the descriptions of interventions in the included studies, although we consulted other available sources of information when possible (eg, intervention websites and study protocols). The complete list of data items is provided in Multimedia Appendix 5. The persuasive design principle *tunneling*, which refers to the sequential presentation of treatment elements in a structured, linear manner, was not counted toward the total number of persuasive design principles in this study. This is because researchers have recently proposed that eHealth interventions can be made more engaging by providing users with greater flexibility and control concerning the modules or features they wish to use [19,22], which contrasts with the principle of *tunneling*.

Risk of Bias Assessment

We assessed the risk of bias among included studies using the Cochrane risk of bias tool [34]. We did not assess the risk domain *blinding participants and personnel* because it is not possible for participants to be blind to their conditions in psychotherapy research [35]. Furthermore, we did not assess the risk domain *blinding of outcome assessment* because all outcome measures were self-report measures, and participants could thus not be blinded. Self-report measures are generally considered equivalent to blind clinical observers in psychotherapy research, and research suggests that they do not result in inflated effect sizes [35].

Meta-analysis Methods

We conducted meta-analyses using Comprehensive Meta-Analysis software (Biostat Inc) [36]. As prior research suggests that ICBT for generalized and social anxiety is more efficacious than ICBT for depression [7,37], we conducted separate meta-analyses of ICBT for anxiety and ICBT for depression. Given the availability of symptom change data for both anxiety and depression, trials of ICBT designed to treat both conditions were included in both meta-analyses. We measured heterogeneity in the effect sizes of the included studies using the I^2 index and formally tested the degree of heterogeneity using the Q statistic [38]. In each of the 2 meta-analyses, we used a random effects model, used between-groups effect size (Hedges g) as the summary measure, and weighted each study by the inverse of the within-study variance of the primary outcome measure plus the between-study variance. Several studies evaluated 2 unguided ICBT interventions; in such cases, we treated the evaluation of each intervention as a separate study, except we divided the control group sample size by 2, such that each control group participant was included only once in the analyses [39]. We evaluated the risk of publication bias using funnel plots and accounted for publication bias using the trim and fill technique [40]. We explored the influence of study-level bias on outcomes by repeating the meta-analyses without studies deemed to be at high risk on one or more dimensions of the Cochrane tool for assessing risk of bias [34].

Meta-regression Methods

We conducted 2 meta-regressions using Comprehensive Meta-Analysis [36]—one for depression interventions and one for anxiety interventions—to determine the degree to which persuasive design principles could explain variance in effect sizes among studies. Paralleling the approach taken to the meta-analyses, we included trials of ICBT designed to treat both depression and anxiety in both meta-regressions, given the availability of symptom change data for both conditions. We also weighted each study by the inverse of the within-study variance of the primary outcome measure plus the between-study variance, as in the meta-analyses.

We used 3 predictor variables. Our main predictor of interest was the total number of persuasive design principles identified for each intervention. We were unable to include the number of persuasive design principles in each category of the PSD framework as separate predictors, as Kelders et al [25] did, because of the risk of overfitting, given the limited number of included studies. We also input a binary variable reflecting whether each intervention was transdiagnostic (ie, designed to treat symptoms of both depression and anxiety). We did this to account for the possibility that unguided ICBT focused on treating a narrower range of symptoms (ie, anxiety or depression) may be more efficacious for treating those symptoms than transdiagnostic unguided ICBT designed to treat a broader range of symptoms (ie, both depression and anxiety). Our final predictor was a binary variable reflecting whether each study used a control condition with active elements (eg, psychoeducation and mood monitoring) because a previous meta-analysis of unguided ICBT found a large mean effect size among studies using passive control conditions and a small mean effect size among studies using active control conditions [41]. However, the control condition type was not included as a predictor in the meta-regression of ICBT for anxiety because there were insufficient studies to justify an additional predictor variable (eg, because of the risk of overfitting), following most recommendations concerning acceptable *subjects per variable* ratios in linear regression analyses [26].

We conducted each meta-regression in 2 steps. The first step included transdiagnostic status and, for the meta-regression of ICBT for depression, the control condition type. In both meta-regressions, the number of persuasive design principles identified was then added in the second step. This 2-step approach was used to reveal the amount of additional variance persuasive design explained in the second step after accounting for the other variables in the first step.

We conducted 5 assumption tests at each step of each meta-regression. First, we examined Pearson *r* correlations and scatterplots to test the assumption of linearity of the relationship between each continuous predictor variable and Hedges *g* [42]. Second, we checked Cook distance values to identify any outlier studies that had unduly large influences on the results [43]. Third, we inspected the distribution of studentized residuals using a histogram to ensure that the residuals were normally distributed [42]. Fourth, we inspected scatterplots plotting studentized residuals against predicted values to test the assumption of homoscedasticity [42]. Finally, we examined variance inflation factors to check for multicollinearity [42].

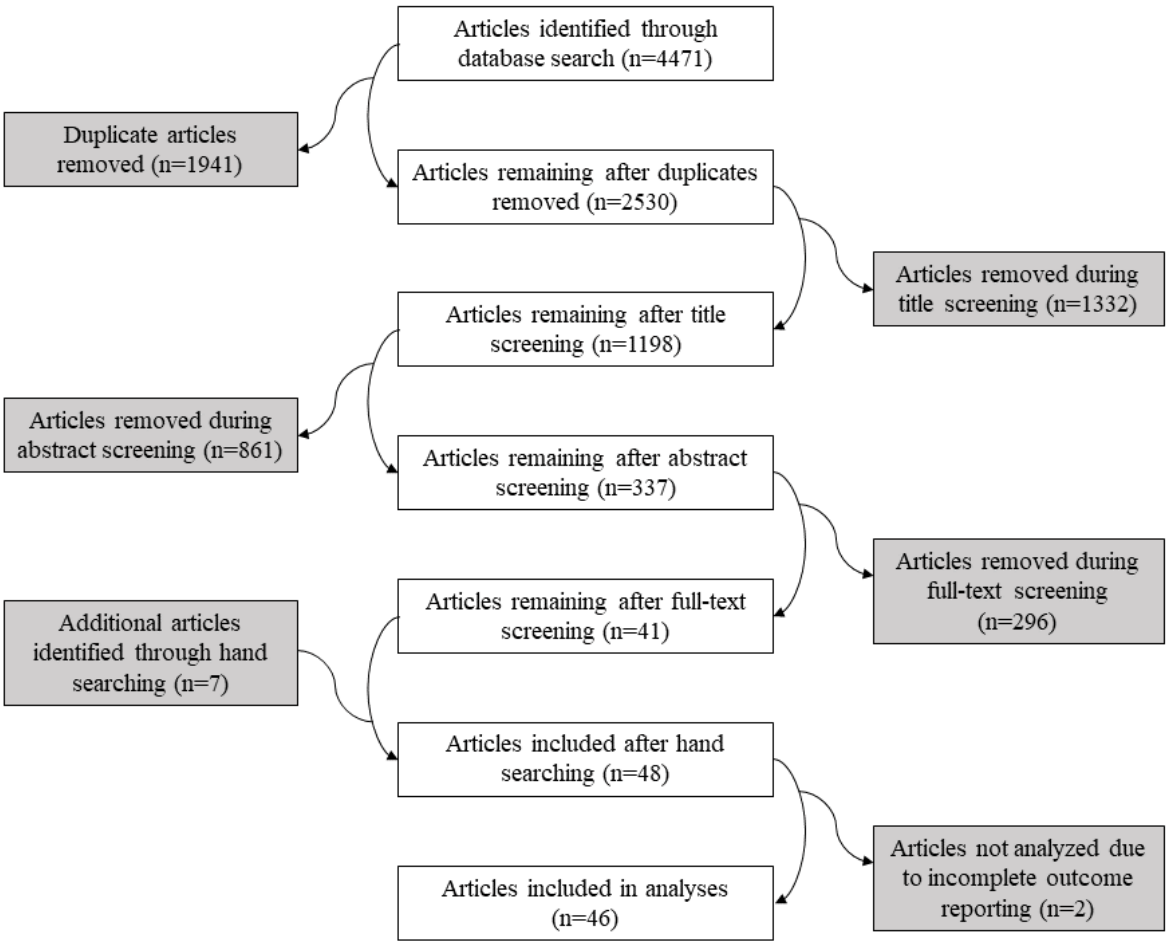
Results

Systematic Review Results

Study Selection

Between the original and updated literature searches, we identified 4471 articles, 39 of which were found eligible for analysis. Having found another 7 eligible articles through a hand search, we included a total of 46 articles. The flow of studies through the study selection process is shown in Figure 1. Separate flowcharts for the original and updated literature searches are shown in Multimedia Appendices 6 and 7, respectively. The 2 screeners (HCM and CRFS) made the same screening decision (ie, retain or remove) for 81.66% (2066/2530) of articles during the title screening, 86.39% (1035/1198) of articles during the abstract screening, and 81.9% (276/337) of articles during the full-text screening.

Figure 1. Flow of studies through the study selection process.



Study Characteristics

The 46 eligible studies included 16,632 participants, excluding participants assigned to experimental groups irrelevant to this study (eg, guided ICBT groups). Studies were most often published in or after 2017 (24/46, 52%); included samples drawn from the general population (26/46, 57%), clinical populations (14/46, 30%), or both (6/46, 13%); and most often used waitlist control conditions without active elements (28/46, 61%). The characteristics of each study are presented in Table 1.

Table 1. Study characteristics.

Category and study	Intervention	Participant, n[a]	Duration in weeks[b,c]	Control condition	Recruitment population	
					Clinical	Nonclinical
ICBT[d] for depression						
Berger et al, 2011 [44]	Deprexis	51	10	Waitlist		✓
Bücker et al, 2019 [45]	MOOD	125	6	Care as usual	✓	✓
Clarke et al, 2002 [46]	ODIN[e]	299	4	Health information website	✓	
Clarke et al, 2005 [47]	ODIN	175	5	Health information website	✓	
Clarke et al, 2009 [48]	__[f]	160	16	Health information website	✓	
Dahne et al, 2019 [49]	¡Aptívate!	33	8	Care as usual	✓	✓
Dahne et al, 2019 [49]	iCouch CBT	20	8	Care as usual	✓	✓
Dahne et al, 2019 [50]	Moodivate	33	8	Care as usual	✓	
Dahne et al, 2019 [50]	MoodKit	28	8	Care as usual	✓	
de Graaf et al, 2009 [51]	Colour Your Life	203	13.05	Care as usual		✓
Farrer et al, 2011 [52]	MoodGym and Bluepages	73	6	Care as usual	✓	
Gräfe et al, 2020 [53]	Deprexis	3805	12	Brochure and care as usual	✓	
Hur et al, 2018 [54]	Todac Todac[g]	34	3	Mood charting app	✓	✓
Lintvedt et al, 2013 [55]	MoodGym and Bluepages	163	8	Waitlist		✓
Löbner et al, 2018 [56]	MoodGym (German adapted, version III)	647	6	Care as usual	✓	
Lüdtke et al, 2018 [57]	Be good to yourself	88	4	Waitlist	✓	✓
Lüdtke et al, 2018 [58]	—	132	4	Care as usual	✓	
McDermott and Dozois, 2019 [59]	MoodGym	302	8	Attentional control		✓
Meyer et al, 2009 [60]	Deprexis	396	9	Waitlist		✓
Meyer et al, 2015 [61]	Deprexis	163	13.05	Waitlist	✓	✓
Mira et al, 2017 [62]	Sonreír es Divertido[h]	80	12	Waitlist		✓
Mohr et al, 2013 [63]	moodManager	68	6	Waitlist	✓	
Montero-Marin et al, 2016 [64]	Smiling is Fun	124	13.05	Improved treatment as usual	✓	
Moritz et al, 2012 [65]	Deprexis	210	8	Waitlist		✓
Morris et al, 2015 [66]	Panoply	166	3	Web-based expressive writing		✓
Noguchi et al, 2017 [67]	—	651	5	Waitlist		✓
Schure et al, 2019 [68]	Thrive	343	8	Depression information website		✓
Silverstone et al, 2017 [69]	MoodGym	109	12	Care as usual	✓	
Spek et al, 2007 [70]	—	202	10	Waitlist		✓
ICBT for depression and anxiety						

Category and study	Intervention	Participant, n[a]	Duration in weeks[b,c]	Control condition	Recruitment population	
					Clinical	Nonclinical
Bakker et al, 2018 [71]	MoodKit	120	4.29	Waitlist		✓
Bakker et al, 2018 [71]	MoodMission	114	4.29	Waitlist		✓
Kleiboer et al, 2015 [72]	Allesondercontrole[i]	213	6	Waitlist and web-based information		✓
Moberg et al, 2019 [73]	Pacifica	500	4.35	Waitlist		✓
Powell et al, 2013 [74]	MoodGym	3070	6	Waitlist		✓
Proudfoot et al, 2013 [75]	myCompass	459	8	Waitlist		✓
Shirotsuki et al, 2017 [76]	—	48	6	Mood monitoring		✓
Twomey et al, 2014 [77]	MoodGym	66	4.57	Waitlist	✓	
ICBT for anxiety						
Berger et al, 2017 [78]	Velibra	139	9	Waitlist	✓	
Boettcher et al, 2018 [79]	Challenger	139	7	Waitlist		✓
Boettcher et al, 2018 [79]	—	139	7	Waitlist		✓
Botella et al, 2010 [80]	Talk to Me	91	8.7	Waitlist		✓
Ciuca et al, 2018 [81]	PAXPD[j]	75	12	Waitlist	✓	✓
Donker et al, 2019 [82]	0Phobia	193	3	Waitlist		✓
Ivanova et al, 2016 [83]	Ångesthjälpen[k]	102	10	Waitlist		✓
Kenardy et al, 2003 [84]	—	83	6	Waitlist		✓
Lin et al, 2020 [85]	—	26	8	Waitlist		✓
McCall et al, 2018 [86]	Overcome Social Anxiety	101	17.4	Waitlist		✓
Oh et al, 2020 [87]	Todaki	41	4	Book on panic disorder	✓	
Powell et al, 2020 [88]	E-couch	2116	6	Waitlist		✓
Titov et al, 2008 [89]	Shyness	64	10	Waitlist		✓

[a]For the purpose of this table, n was calculated as the number of participants assigned to the intervention identified in each row plus the number of participants assigned to the control condition (ie, excluding participants assigned to use other interventions).

[b]Study duration expressed in days was divided by 7. Study duration expressed in months was multiplied by 4.35 (the average number of weeks in a month during a 365-day year).

[c]Some studies reported data from multiple posttreatment time points; for such studies, the duration, as shown in this table, is the number of weeks between pretreatment and whichever posttreatment time point was selected for use in the analyses reported in this study.

[d]ICBT: internet-delivered cognitive behavioral therapy.

[e]ODIN: Overcoming Depression on the Internet.

[f]Data were not reported.

[g]Todac Todac translates to “Tap Tap.”

[h]Sonreír es Divertido translates to “Smiling is Fun.”

[i]Allesondercontrole translates to “all is under control.”

[j]PAXPD: PAXonline Program for Panic Disorder.

[k]Ångesthjälpen translates to “The Anxiety Help.”

Risk of Bias

We evaluated the risk of bias among included studies using 5 of the 7 domains in the Cochrane risk of bias tool [34]. Of the 46 included studies, 14 (30%) were identified to be at high risk of bias in at least one domain, whereas only 4 (9%) were found to be at low risk of bias in all domains assessed. Most studies (28/46, 61%) were found to be at low or unclear risk in each domain. The risk of bias identified in each study is presented in Table 2.

Table 2. Risk of bias in included studies.

Category and study	Random sequence generation	Allocation concealment	Incomplete outcome data (attrition bias)	Selective reporting	Other bias
ICBT[a] for depression					
Berger et al, 2011 [44]	Low	Unclear	Low	Unclear	Low
Bücker et al, 2019 [45]	Low	Low	Low	Unclear	Low
Clarke et al, 2002 [46]	Low	Low	Low	Unclear	Low
Clarke et al, 2005 [47]	Low	Low	High	Unclear	Low
Clarke et al, 2009 [48]	Unclear	Unclear	Low	Unclear	Unclear
Dahne et al, 2019 [49]	Unclear	Unclear	Low	Unclear	Low
Dahne et al, 2019 [50]	Unclear	Unclear	Unclear	Unclear	Low
de Graaf et al, 2009 [51]	Low	Low	Low	Low	Low
Farrer et al, 2011 [52]	Unclear	Low	Low	Unclear	Low
Gräfe et al, 2020 [53]	Low	Unclear	Low	Low	Low
Hur et al, 2018 [54]	Low	Low	Low	Unclear	Low
Lintvedt et al, 2013 [55]	Low	Unclear	Low	Unclear	Low
Löbner et al, 2018 [56]	Low	Low	Low	Unclear	Low
Lüdtke et al, 2018 [57]	Unclear	Low	Low	Unclear	Low
Lüdtke et al, 2018 [58]	High	Low	Low	Unclear	Low
McDermott and Dozois, 2019 [59]	Unclear	Unclear	Low	Unclear	Unclear
Meyer et al, 2009 [60]	Low	High	High	Unclear	Low
Meyer et al, 2015 [61]	Low	Low	Low	Unclear	Low
Mira et al, 2017 [62]	Low	Low	Low	Unclear	Low
Mohr et al, 2013 [63]	Low	Low	Low	Unclear	Low
Montero-Marin et al, 2016 [64]	Low	Low	Low	Low	Low
Moritz et al, 2012 [65]	Unclear	Unclear	Low	Unclear	Low
Morris et al, 2015 [66]	Unclear	Unclear	Low	Unclear	Low
Noguchi et al, 2017 [67]	Low	Low	Low	Unclear	Low
Schure et al, 2019 [68]	Unclear	Unclear	Low	Low	Low
Silverstone et al, 2017 [69]	High	High	High	Unclear	Low
Spek et al, 2007 [70]	Unclear	Low	Low	Unclear	Low
ICBT for depression and anxiety					
Bakker et al, 2018 [71]	High	High	High	Unclear	Low
Kleiboer et al, 2015 [72]	Low	Low	High	Unclear	Unclear
Moberg et al, 2019 [73]	Unclear	Unclear	High	Unclear	Low
Powell et al, 2013 [74]	Low	Low	Low	Low	Low
Proudfoot et al, 2013 [75]	Low	Low	High	Unclear	Low
Shirotsuki et al, 2017 [76]	Unclear	Low	Low	Unclear	Low
Twomey et al, 2014 [77]	Low	High	High	Unclear	Low
ICBT for anxiety					
Berger et al, 2017 [78]	Low	Low	Low	Unclear	Low
Boettcher et al, 2018 [79]	Low	Low	Low	Unclear	Low
Botella et al, 2010 [80]	Unclear	Unclear	High	Unclear	Unclear
Ciuca et al, 2018 [81]	Low	Low	Low	Unclear	Low

Category and study	Random sequence generation	Allocation concealment	Incomplete outcome data (attrition bias)	Selective reporting	Other bias
Donker et al, 2019 [82]	Low	Low	Low	Low	Low
Ivanova et al, 2016 [83]	Low	Low	High	Low	Low
Kenardy et al, 2003 [84]	Unclear	Unclear	Low	Unclear	Unclear
Lin et al, 2020 [85]	Low	Unclear	High	Unclear	Unclear
McCall et al, 2018 [86]	Low	High	Low	Unclear	Low
Oh et al, 2020 [87]	Unclear	Unclear	Low	Low	Low
Powell et al, 2020 [88]	Low	Low	High	Low	Low
Titov et al, 2008 [89]	Low	Unclear	Low	Unclear	Low

[a]ICBT: internet-delivered cognitive behavioral therapy.

Intervention Characteristics

In total, 37 unguided ICBT interventions were evaluated in the 46 included studies. Of these 37 interventions, 15 (41%) were designed to treat depression exclusively and 9 (24%) were designed to treat depression and anxiety or stress. Other interventions were designed to treat social anxiety (6/37, 16%), panic (2/37, 5%), fear of public speaking (1/37, 3%), generalized anxiety (1/37, 3%), acrophobia (1/37, 3%), or symptoms of multiple anxiety disorders (2/37, 5%). Most interventions (23/37, 62%) were described as traditional CBT interventions, but many interventions (9/37, 24%) were described as being strongly influenced by elements of third-wave CBT (eg, mindfulness) or other therapeutic approaches (eg, positive psychology), and several interventions were based on behavioral activation (2/37, 5%), cognitive therapy (2/37, 5%), or problem-solving therapy (1/37, 3%). Half of the interventions (19/37, 51%) were delivered via a web browser, but many interventions were delivered via a mobile app (11/37, 30%) or both a browser and an app (5/37, 14%). Of the 37 interventions, it was unclear how 2 (5%) interventions were delivered. The characteristics of each intervention are presented in Table 3.

Table 3. Intervention characteristics.

Study	Name of the intervention	Target symptoms	Theoretical approach	Composition	Delivery medium	Number of persuasive design principles identified
Dahne et al, 2019 [49]	¡Aptívate!	Depression	Behavioral activation	Unclear	Mobile app	8
Donker et al, 2019 [82]	0Phobia	Acrophobia	CBT[a]	6 modules	Mobile app	6
Kleiboer et al, 2015 [72]	Allesondercontrole[b]	Depression and anxiety	Problem-solving therapy	5 lessons	Web browser	1
Ivanova et al, 2016 [83]	Ångesthjälpen[c]	Panic and social anxiety	Acceptance and commitment therapy	8 modules	App, browser, and CD	4
Lüdtke et al, 2018 [57]	Be Good to Yourself	Depression	CBT with third-wave elements	4 modules	Mobile app	4
Boettcher et al, 2018 [79]	Challenger	Social anxiety	CBT	N/A[d]	Mobile app	13
de Graaf et al, 2009 [51]	Colour Your Life	Depression	CBT	9 modules	Web browser	2
Berger et al, 2011 [44]; Gräfe et al, 2020 [53]; Meyer et al, 2009 [60]; Meyer et al, 2015 [61]; and Moritz et al, 2012 [65]	Deprexis	Depression	CBT and other approaches	12 modules	Web browser	5
Powell et al, 2020 [88]	E-couch	Social anxiety	CBT	6 modules	App and browser	2
Dahne et al, 2019 [49]	iCouch CBT	Depression and anxiety	CBT	Unclear	Mobile app	1
Bücker et al, 2019 [45]	MOOD	Depression	CBT with third-wave elements	9 modules	Web browser	5
Farrer et al, 2011 [52]; Lintvedt et al, 2013 [55]; Löbner et al, 2018 [56]; McDermott and Dozois, 2019 [59]; and Twomey et al, 2014 [77]	MoodGym[e]	Depression and anxiety	CBT and other approaches	5 modules	Web browser	1
Dahne et al, 2019 [50]	Moodivate	Depression	Behavioral activation	7 modules	Mobile app	8
Bakker et al, 2018 [71] and Dahne et al, 2019 [50]	MoodKit	Depression and anxiety	CBT	4 features	Mobile app	5
Mohr et al, 2013 [63]	moodManager	Depression	CBT	18 lessons	Web browser	4
Bakker et al, 2018 [71]	MoodMission	Depression and anxiety	CBT	N/A	Mobile app	4
Proudfoot et al, 2013 [75]	myCompass	Depression, anxiety, and stress	CBT and other approaches	12 modules	App and browser	6
McCall et al, 2018 [86]	Overcome Social Anxiety	Social anxiety	CBT	7 modules	Web browser	8

Study	Name of the intervention	Target symptoms	Theoretical approach	Composition	Delivery medium	Number of persuasive design principles identified
Clarke et al, 2002 [46] and Clarke et al, 2005 [47]	ODIN[f]	Depression	Cognitive therapy	7 modules	Web browser	5
Moberg et al, 2019 [73]	Pacifica	Depression, anxiety, and stress	CBT and other approaches	Unclear	Mobile app	8
Morris et al, 2015 [66]	Panoply	Depression	Cognitive therapy	N/A	Web browser	8
Ciuca et al, 2018 [81]	PAXPD[g]	Panic disorder	CBT	16 modules	Web browser	3
Titov et al, 2008 [89]	Shyness	Social anxiety	CBT	6 lessons	Web browser	8
Mira et al, 2017 [62] and Montero-Marin et al, 2016 [64]	Sonreír es Divertido[h]	Depression	CBT and other approaches	10 modules	Web browser	5
Botella et al, 2010 [80]	Talk to Me	Fear of public speaking	CBT	Unclear	Web browser	6
Schure et al, 2019 [68]	Thrive	Depression	CBT	3 modules	App and browser	3
Hur et al, 2018 [54]	Todac Todac[i]	Depression	CBT	3 modules	Mobile app	7
Oh et al, 2020 [87]	Todaki	Panic	CBT	4 modes	Mobile app	7
Berger et al, 2017 [78]	Velibra	Various anxiety disorders	CBT and other approaches	6 sessions	Web browser	5
Boettcher et al, 2018 [79]	Not reported	Social anxiety disorder	CBT	9 modules	Unclear	2
Clarke et al, 2009 [48]	Not reported	Depression	CBT	4 sections	Web browser	8
Kenardy et al, 2003 [84]	Not reported	Anxiety	CBT	6 sessions	Web browser	3
Lin et al, 2020 [85]	Not reported	Social anxiety	CBT	8 modules	Web browser	10
Lüdtke et al, 2018 [58]	Not reported	Depression	CBT	1 module	App and browser	3
Noguchi et al, 2017 [67]	Not reported	Depression and stress	CBT	Unclear	Web browser	1
Shirotsuki et al, 2017 [76]	Not reported	Depression- and anxiety-related symptoms	CBT	6 modules	e-learning system and guidebook	2
Spek et al, 2007 [70]	Not reported	Depression	CBT	8 modules	Web browser	2

[a]CBT: cognitive behavioral therapy.

[b]Allesondercontrole translates to “all is under control.”

[c]Ångesthjälpen translates to “The Anxiety Help.”

[d]N/A: not applicable.

[e]Bluepages was offered as a complement to MoodGym in studies by Farrer et al [52] and Lintvedt et al [55] but was omitted from this table (and all analyses) because it is a psychoeducation package and not an internet-delivered cognitive behavioral therapy intervention.

[f]ODIN: Overcoming Depression on the Internet.

[g]PAXPD: PAXonline Program for Panic Disorder.

[h]Sonreír es Divertido translates to “Smiling is Fun.”

[i]Todac Todac translates to “Tap Tap.”

Persuasive Design

On average, interventions included 4.95 (SD 2.85) persuasive design principles (excluding *tunneling*). The total number of persuasive design elements ranged from 1 to 13. Principles in the *primary task support* category were the most common (mean 2.86, SD 1.32), followed by principles in the *dialogue support* category (mean 1.27, SD 1.19) and *social support* category (mean 0.81, SD 1.60). The number of interventions in which each persuasive design principle was identified is presented in Table 4.

Table 4. Persuasive design principles identified.

Persuasive design principle	Brief description[a]	Interventions used, n (%)
Primary task support		
Reduction	Divides target behavior into simple steps	35 (95)
Tunneling	Delivers content in a step-by-step format	29 (78)
Tailoring	Provides content adapted to user group	2 (5)
Personalization	Provides content that is adapted to one user	18 (49)
Self-monitoring	Provides ability to monitor progress or status	20 (54)
Simulation	Provides ability to observe relevant behavior	6 (16)
Rehearsal	Provides ability to rehearse a behavior	25 (68)
Dialogue support		
Praise	Offers praise to participant	8 (22)
Rewards	Offers reward to participant	5 (14)
Reminders	Provides reminders	13 (35)
Suggestion	Provides suggestions	15 (41)
Similarity	Is designed to look familiar	0 (0)
Liking	Is visually designed to be attractive	1 (3)
Social role	Acts as if it has a social role	5 (14)
Social support		
Social learning	Facilitates learning from other users	7 (19)
Social comparison	Facilitates comparison with other users	5 (14)
Normative influence	Provides normative information on target behavior	2 (5)
Social facilitation	Facilitates awareness of others using intervention	5 (14)
Cooperation	Stimulates users to cooperate	8 (22)
Competition	Stimulates users to compete	0 (0)
Recognition	Shows users who adopted target behavior	3 (8)

[a]These descriptions were adapted from the operational definitions provided by Kelders et al [25].

Meta-analysis Results

Meta-analysis of Unguided ICBT for Depression

We conducted a meta-analysis of 37 comparisons across 34 trials of unguided ICBT for depression. There was statistically significant heterogeneity in Hedges *g* among the studies (Q=89.85, df=36; P<.001). An I^2 statistic of 59.93 indicated that a moderate proportion of variability was attributable to true heterogeneity rather than sampling error [90,91]. The weighted mean between-subjects effect size was small to moderate (Hedges g=0.31; SE 0.04; 95% CI 0.24-0.38). The forest plot for this meta-analysis is shown in Figure 2. Weighted mean effect sizes after excluding studies deemed to be at high risk of bias and after adjusting for publication bias using the trim and fill technique [40] are presented in Table 5.

Figure 2. Meta-analysis of unguided internet-delivered cognitive behavioral therapy for depression.

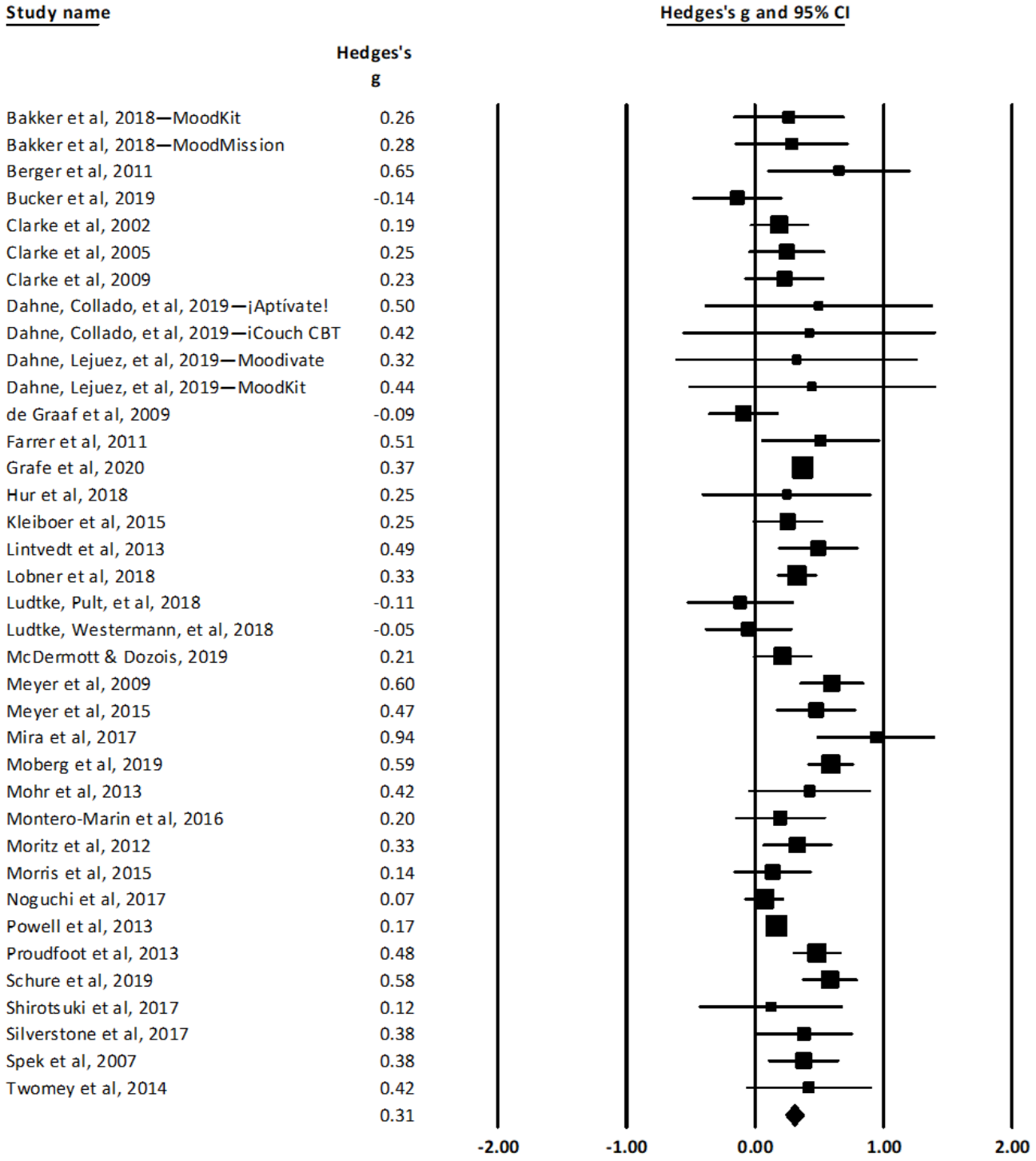


Table 5. Summary statistics of meta-analyses with and without bias corrections.

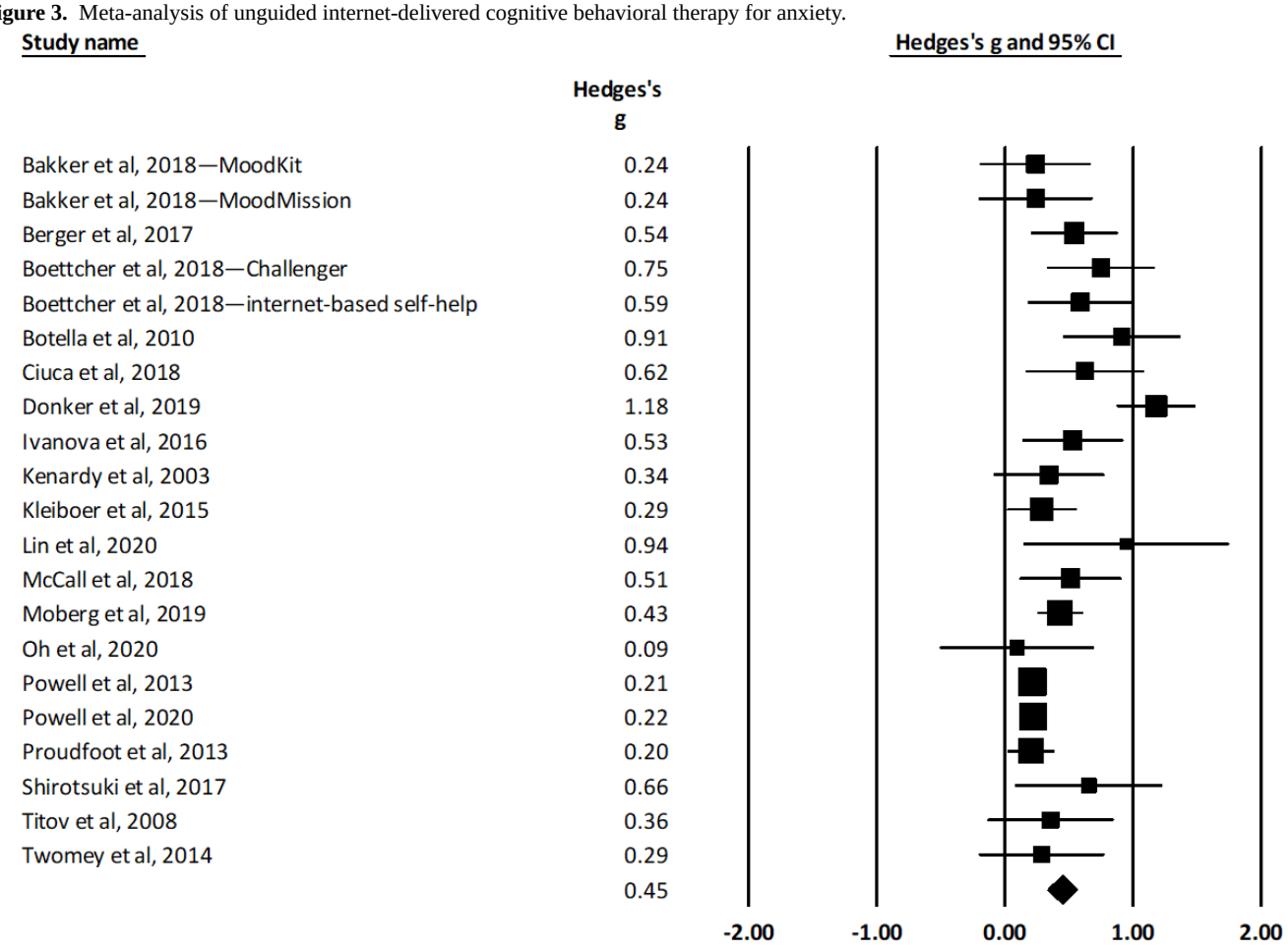
Meta-analysis	Hedges *g* (95% CI)
Meta-analysis of ICBT[a] for depression	
All studies of ICBT for depression	0.31 (0.24-0.38)
All studies with trim and fill adjustment	0.23 (0.16-0.31)
Studies with high risk of bias excluded	0.28 (0.20-0.36)
Studies with high risk of bias excluded, with trim and fill adjustment	0.22 (0.14-0.31)
Meta-analysis of ICBT for anxiety	
All studies of ICBT for anxiety	0.45 (0.33-0.56)
All studies with trim and fill adjustment	0.45 (0.33-0.56)
Studies with high risk of bias excluded	0.54 (0.29-0.79)
Studies with high risk of bias excluded, with trim and fill adjustment	0.54 (0.29-0.79)

[a]ICBT: internet-delivered cognitive behavioral therapy.

Meta-analysis of Unguided ICBT for Anxiety

We included 19 studies that reported 21 comparisons in a meta-analysis of unguided ICBT for anxiety. The results indicated statistically significant heterogeneity of Hedges *g* among studies (Q=68.47, df=20; P<.001). The corresponding I^2 statistic of 70.79 suggested that a substantial proportion of the variability represented true heterogeneity [90,91]. The weighted mean between-subjects effect size was moderate (Hedges g=0.45; SE 0.06; 95% CI 0.33-0.56). A forest plot is shown in Figure 3. Additional weighted mean effect sizes accounting for publication- and study-level bias are presented in Table 5.

Figure 3. Meta-analysis of unguided internet-delivered cognitive behavioral therapy for anxiety.



Meta-regression Results

Meta-regression of Unguided ICBT for Depression

The meta-regression of ICBT for depression, like the meta-analysis of ICBT for depression, included 34 studies reporting 37 comparisons. We used 3 predictors in this meta-regression: the total number of persuasive design principles (mean 3.90, SD 2.33), whether each intervention was designed to treat symptoms of both depression and anxiety (19/37, 51%) or only depression (18/37, 49%), and whether each study used an active control condition (13/37, 35%) or a passive control condition (24/37, 65%). The results for both steps of the meta-regression are presented in Table 6. With the possible exception of very minor heteroscedasicity of residuals at one or both steps, all assumptions were met, as detailed in Multimedia Appendix 8.

Table 6. Meta-regression of unguided internet-delivered cognitive behavioral therapy for depression.

Step and variable	Model				Predictors				
	Model summary		R^2	R^2 change	B [a]	SE	95% CI	P value	Variance inflation factor
	Q (df)	P value							
Step 1	1.05 (2)	.59	0.0	N/A[b]					
Constant					0.28	0.07	0.15 to 0.41	<.001	3.03
Active control condition					−0.02	0.08	−0.19 to 0.14	.79	1.11
Transdiagnostic intervention					0.07	0.08	−0.09 to 0.23	.40	1.11
Step 2	6.74 (3)	.08	0.27	0.27					
Constant					0.10	0.10	−0.09 to 0.29	.32	8.63
Active control condition					−0.01	0.07	−0.16 to 0.13	.85	1.14
Transdiagnostic intervention					0.13	0.08	−0.02 to 0.28	.09	1.36
Persuasive design principles					0.04	0.02	0.01 to 0.07	.02	1.22

[a]Unstandardized β coefficient.

[b]N/A: not applicable.

Meta-regression of Unguided ICBT for Anxiety

Similar to the meta-analysis of ICBT for anxiety, the meta-regression of ICBT for anxiety included 19 studies reporting 21 comparisons. We used 2 predictors: the total number of persuasive design principles (mean 5.05, SD 3.17) and whether each intervention was designed to treat symptoms of both depression and anxiety (8/21, 38%) or only anxiety (13/21, 62%). The results for both steps of the meta-regression are presented in Table 7. The assumption of normality of residuals may not have been met fully at both steps, although the residuals roughly approximated normal distributions. The assumption of homoscedasticity of the residuals was violated in step 1. The assumption tests for this meta-regression are detailed in Multimedia Appendix 9.

Table 7. Meta-regression of unguided internet-delivered cognitive behavioral therapy for anxiety.

Step and variable	Model				Predictors				
	Model summary		R^2	R^2 change	B [a]	SE	95% CI	P value	Variance inflation factor
	Q (df)	P value							
Step 1	4.80 (1)	.03	0.0	N/A[b]					
Constant					0.57	0.08	0.41 to 0.73	<.001	1.77
Transdiagnostic intervention					−0.27	0.12	−0.51 to −0.03	.03	1.00
Step 2	7.55 (2)	.02	0.05	0.05					
Constant					0.42	0.13	0.18 to 0.67	<.001	5.12
Transdiagnostic intervention					−0.23	0.12	−0.46 to −0.00	.049	1.07
Persuasive design principles					0.03	0.02	−0.01 to 0.06	.17	1.07

[a]Unstandardized β coefficient.

[b]N/A: not applicable.

Discussion

Principal Findings

Recent years have witnessed a proliferation of randomized trials of eHealth interventions, including many trials of unguided ICBT for depression and anxiety. Indeed, most of the studies included in this review were published in or after 2017. There was considerable diversity in the design of both studies (eg, study duration and type of control condition) and interventions (eg, mode of delivery and use of persuasive design principles).

The results of the meta-analysis of unguided ICBT for depression were consistent with the results of previous meta-analyses. We reported 4 mean effect sizes (Hedges *g*) for unguided ICBT for depression, ranging from 0.22 to 0.31, based on the corrections we made for publication bias and study-level bias. Previous meta-analyses of unguided ICBT for depression have found comparable mean effect sizes (Hedges *g* or Cohen *d*) ranging from 0.24 to 0.36 [12,92-95]. Our meta-analysis of unguided ICBT for anxiety yielded a mean effect size of 0.45. There was no evidence of publication bias, and the mean effect size was greater (Hedges *g*=0.54) after excluding studies found to be at a high risk of bias. Several previous meta-analyses of ICBT for symptoms of anxiety disorders found effect sizes between 0.70 and 1.12 [41,96-98]; however, all these meta-analyses included trials of guided ICBT interventions, which likely explains the greater mean effect sizes, at least in part. We are aware of only 1 meta-analysis that has included a subgroup analysis of *unguided* ICBT for anxiety—social anxiety, specifically—finding mean effect sizes (Hedges *g*) of 0.78 and 0.19 for studies using passive and active control conditions, respectively [41]. It is worth noting that our review included many transdiagnostic interventions designed to treat symptoms of both depression and anxiety. The meta-regressions showed that these interventions were significantly less efficacious for treating anxiety symptoms compared with interventions designed to treat anxiety symptoms only; however, their efficacy in treating depression did not significantly differ from interventions designed to treat symptoms of depression only.

We identified wide variability in the use of persuasive design in unguided ICBT for depression and anxiety, with several interventions using only 1 persuasive design principle and others using as many as 13. The intervention identified as having the greatest number of persuasive design principles (ie, 13), called *Challenger*, was specifically designed to be engaging, with many features inspired by the literature on gamification [79,99]. The mean number of persuasive design principles identified across interventions (4.95, excluding the principle of *tunneling*) was comparable with the mean of 5.4 principles identified by Kelders et al [25] among mental health interventions in their review. The mean number of persuasive design principles identified in the primary task support (mean 2.86, SD 1.32; excluding tunneling), dialogue support (mean 1.27, SD 1.19), and social support (mean 0.81, SD 1.60) categories were also roughly comparable with the corresponding means identified among mental health interventions by Kelders et al [25] (2.6, 1.6, and 1.3, respectively).

Persuasive design was a significant predictor of effect size in the meta-regression of ICBT for depression. The unstandardized β coefficient (*B*) of 0.04 suggested that for each additional persuasive design principle an intervention uses, one could predict the effect size (Hedges *g*) for that intervention to increase by 0.04, compared with a control condition in a randomized trial. However, meta-regression is an inherently observational procedure [100], and the results therefore could not show whether persuasive design *caused* certain ICBT interventions for depression to be more efficacious than others. Persuasive design did not predict efficacy in the meta-regression of ICBT for anxiety. However, it is worth noting that the meta-regression of ICBT for anxiety included far fewer studies than the

meta-regression of ICBT for depression and had limited statistical power to identify an effect. Indeed, persuasive design had an unstandardized β coefficient of 0.03 in the meta-regression of ICBT for anxiety, which—although not statistically significant—was comparable in magnitude with that of the meta-regression of ICBT for depression. The results of the meta-regression of unguided ICBT for anxiety should be interpreted cautiously because assumption tests showed that certain assumptions were unmet. Nonetheless, our results suggest that persuasive design is more closely related to outcomes in interventions for depression than anxiety. Given that persuasive design is purported to motivate engagement in treatment [17] and that lack of motivation is a hallmark of depression, it is possible that persuasive design is particularly important in ICBT for depression.

Overall, our findings support the hypothesis that persuasive design predicts efficacy in unguided ICBT, at least in the treatment of depression. Our findings also support the validity of the PSD framework [17] by showing that it is meaningfully related to treatment outcomes. Although the results do not demonstrate the importance of any specific persuasive design principles, they support the growing body of theory and data suggesting, broadly, that persuasive design matters in eHealth [18-24]. These findings are encouraging and timely. ICBT has become well established over the last two decades, having now been evaluated in hundreds of trials [101] and currently being funded by many governments around the world [102]. It is clear that ICBT is effective, and a natural next step in ICBT research will be to explore possible avenues for making it more effective. Our findings suggest that enhanced persuasive design may be one such avenue. Notably, because ICBT is highly scalable, particularly when it is unguided, even slight increases in effectiveness can have substantial and wide-reaching implications for public health.

Limitations

This study had several limitations. First, a considerable amount of data was unreported; in particular, it is likely that many interventions used persuasive design principles that were not described in the included studies. Second, although we were able to identify the principles in the PSD framework as present or absent, we did not have access to the interventions themselves, and we were unable to evaluate how effectively persuasive design principles were implemented. Third, we were unable to show, through our meta-regressions, whether specific persuasive design principles predicted efficacy. Finally, only 1 researcher was involved in data extraction; a second extractor would have helped reduce the risk of error, inconsistency, or bias.

Future Directions

Further research will be required to clarify the role of persuasive design in unguided ICBT and other eHealth interventions. First, dismantling studies comparing versions of interventions with and without certain persuasive design principles could evaluate the utility of specific principles. Factorial randomized trials of this kind would allow researchers to efficiently evaluate multiple persuasive design principles in a single study. Second, it would be helpful to explore how intervention users experience persuasive design, which could perhaps be achieved through qualitative research or the development of a self-report questionnaire assessing user experiences of persuasive design. Third, the literature would benefit from a more detailed description of persuasive design in unguided ICBT interventions based on a careful review of the interventions themselves (ie, rather than this study's review of descriptions of interventions from randomized trials). Finally, further research will be required to test our finding that persuasive design predicts efficacy in unguided ICBT for depression but not for anxiety.

Conclusions

The literature on ICBT and other eHealth interventions is evolving rapidly. This review has provided an updated meta-analysis of unguided ICBT for depression and anxiety, generally finding smaller effect sizes for depression than for anxiety. It has also documented the wide variability in the use of persuasive design in unguided ICBT and demonstrated through a meta-regression that persuasive design predicts efficacy in unguided ICBT for depression. Persuasive design is a promising avenue for further optimization of eHealth interventions, including ICBT, and an area of research that is worth investigating further.

Acknowledgments

The authors would like to thank Dr Gordon Asmundson, Dr Donald Sharpe, Dr Swati Mehta, and Dr Janine Olthuis for providing valuable feedback on this research. This research was made possible by scholarship funding provided to HCM by the Social Sciences and Humanities Research Council of Canada. This research was also supported by PSPNET, which is led by HDH and funded by the Canadian Government's Ministry of Public Safety and Emergency Preparedness. HDH holds funding from the Canadian Institutes of Health Research, Saskatchewan Health Research Foundation, Saskatchewan Centre for Patient-Oriented Research, and Craig Neilson Foundation. CRFS is currently employed through funding from AFA Försäkring.

Authors' Contributions

HCM and HDH formulated the idea for this study and developed the search terms. HCM conducted the literature search. HCM and CRFS conducted eligibility screening of the identified articles. HCM extracted and analyzed the data and wrote the first draft of the manuscript with support from HDH. All authors contributed to the revision of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The persuasive systems design framework.
[DOCX File , 33 KB - jmir_v23i4e26939_app1.docx]

Multimedia Appendix 2
Recommendations for making eHealth interventions more engaging and related persuasive design principles.
[DOCX File , 39 KB - jmir_v23i4e26939_app2.docx]

Multimedia Appendix 3
Log of revisions and clarifications to the original methodological protocol.
[DOCX File , 34 KB - jmir_v23i4e26939_app3.docx]

Multimedia Appendix 4
Search terms.
[DOCX File , 30 KB - jmir_v23i4e26939_app4.docx]

Multimedia Appendix 5
Data items.
[DOCX File , 31 KB - jmir_v23i4e26939_app5.docx]

Multimedia Appendix 6
Flow of studies in the original literature search (October 29, 2019).
[PNG File , 32 KB - jmir_v23i4e26939_app6.png]

Multimedia Appendix 7
Flow of studies in the revised literature search (July 2, 2020).
[PNG File , 31 KB - jmir_v23i4e26939_app7.png]

Multimedia Appendix 8
Assumption tests for meta-regression of unguided internet-delivered cognitive behavioral therapy for depression.
[DOCX File , 103 KB - jmir_v23i4e26939_app8.docx]

Multimedia Appendix 9
Assumption tests for meta-regression of unguided internet-delivered cognitive behavioral therapy for anxiety.
[DOCX File , 94 KB - jmir_v23i4e26939_app9.docx]

References

1. Depression and other common mental disorders: global health estimates. World Health Organization. 2017. URL: https://www.who.int/mental_health/management/depression/prevalence_global_health_estimates/en/ [accessed 2021-04-22]
2. Cuijpers P. Four decades of outcome research on psychotherapies for adult depression: an overview of a series of meta-analyses. Can Psychol / Psychologie Canadienne 2017;58(1):7-19. [doi: 10.1037/cap0000096]
3. Cuijpers P, Berking M, Andersson G, Quigley L, Kleiboer A, Dobson KS. A meta-analysis of cognitive-behavioural therapy for adult depression, alone and in comparison with other treatments. Can J Psychiatry 2013 Jul;58(7):376-385. [doi: 10.1177/070674371305800702] [Medline: 23870719]
4. Otte C. Cognitive behavioral therapy in anxiety disorders: current state of the evidence. Dialogues Clin Neurosci 2012 Jan;13(4):413-421. [doi: 10.31887/dcns.2011.13.4/cotte]
5. Andrade LH, Alonso J, Mneimneh Z, Wells JE, Al-Hamzawi A, Borges G, et al. Barriers to mental health treatment: results from the WHO World Mental Health surveys. Psychol Med 2013 Aug 09;44(6):1303-1317. [doi: 10.1017/s0033291713001943]
6. Mojtabai R, Olfson M, Sampson NA, Jin R, Druss B, Wang PS, et al. Barriers to mental health treatment: results from the National Comorbidity Survey Replication. Psychol Med 2011 Aug;41(8):1751-1761 [FREE Full text] [doi: 10.1017/S0033291710002291] [Medline: 21134315]

7. Andersson G, Carlbring P, Titov N, Lindefors N. Internet interventions for adults with anxiety and mood disorders: a narrative umbrella review of recent meta-analyses. Can J Psychiatry 2019 Jul 16;64(7):465-470 [FREE Full text] [doi: 10.1177/0706743719839381] [Medline: 31096757]
8. Christensen H, Hickie IB. E-mental health: a new era in delivery of mental health services. Med J Aust 2010 Jun 07;192(S11):S2-S3. [doi: 10.5694/j.1326-5377.2010.tb03684.x] [Medline: 20528702]
9. Furmark T, Carlbring P, Hedman E, Sonnenstein A, Clevberger P, Bohman B, et al. Guided and unguided self-help for social anxiety disorder: randomised controlled trial. Br J Psychiatry 2009 Nov;195(5):440-447. [doi: 10.1192/bjp.bp.108.060996] [Medline: 19880935]
10. Richards D, Enrique A, Palacios J, Duffy D. Internet-delivered cognitive behaviour therapy. In: Cognitive Behavioral Therapy and Clinical Applications. Online: IntechOpen; 2017.
11. Baumeister H, Reichler L, Munzinger M, Lin J. The impact of guidance on Internet-based mental health interventions — a systematic review. Internet Interv 2014 Oct;1(4):205-215. [doi: 10.1016/j.invent.2014.08.003]
12. Andersson G, Cuijpers P. Internet-based and other computerized psychological treatments for adult depression: a meta-analysis. Cogn Behav Ther 2009 Dec;38(4):196-205. [doi: 10.1080/16506070903318960] [Medline: 20183695]
13. Gerhards SA, de Graaf LE, Jacobs LE, Severens JL, Huibers MJ, Arntz A, et al. Economic evaluation of online computerised cognitive-behavioural therapy without support for depression in primary care: randomised trial. Br J Psychiatry 2010 Apr;196(4):310-318 [FREE Full text] [doi: 10.1192/bjp.bp.109.065748] [Medline: 20357309]
14. Lorenzo-Luaces L, Johns E, Keefe JR. The generalizability of randomized controlled trials of self-guided internet-based cognitive behavioral therapy for depressive symptoms: systematic review and meta-regression analysis. J Med Internet Res 2018 Nov 09;20(11):e10113 [FREE Full text] [doi: 10.2196/10113] [Medline: 30413400]
15. Andersson G, Titov N. Advantages and limitations of internet-based interventions for common mental disorders. World Psychiatry 2014 Feb;13(1):4-11 [FREE Full text] [doi: 10.1002/wps.20083] [Medline: 24497236]
16. Fogg BJ. Persuasive technology. Ubiquity 2002 Dec 01;2002(December):2. [doi: 10.1145/764008.763957]
17. Oinas-Kukkonen H, Harjumaa M. Persuasive systems design: key issues, process model, and system features. Commun Assoc Inf Syst 2009;24:-. [doi: 10.17705/1CAIS.02428]
18. Morrison LG, Yardley L, Powell J, Michie S. What design features are used in effective e-health interventions? A review using techniques from Critical Interpretive Synthesis. Telemed J E Health 2012 Mar;18(2):137-144. [doi: 10.1089/tmj.2011.0062] [Medline: 22381060]
19. Fleming TM, de Beurs D, Khazaal Y, Gaggioli A, Riva G, Botella C, et al. Maximizing the impact of e-therapy and serious gaming: time for a paradigm shift. Front Psychiatry 2016;7:65 [FREE Full text] [doi: 10.3389/fpsyt.2016.00065] [Medline: 27148094]
20. Helgadóttir FD, Menzies RG, Onslow M, Packman A, O'Brian S. Online CBT I: bridging the gap between eliza and modern online CBT treatment packages. Behav Change 2012 Feb 22;26(4):245-253. [doi: 10.1375/bech.26.4.245]
21. Schubart JR, Stuckey HL, Ganeshamoorthy A, Sciamanna CN. Chronic health conditions and internet behavioral interventions: a review of factors to enhance user engagement. Comput Inform Nurs 2011 Feb;29(2):81-92. [doi: 10.1097/NCN.0b013e3182065eed] [Medline: 21164337]
22. Ludden GD, van Rompay TJ, Kelders SM, van Gemert-Pijnen JE. How to increase reach and adherence of web-based interventions: a design research viewpoint. J Med Internet Res 2015 Jul 10;17(7):e172 [FREE Full text] [doi: 10.2196/jmir.4201] [Medline: 26163456]
23. Brouwer W, Kroeze W, Crutzen R, de Nooijer J, de Vries NK, Brug J, et al. Which intervention characteristics are related to more exposure to internet-delivered healthy lifestyle promotion interventions? A systematic review. J Med Internet Res 2011 Jan;13(1):e2 [FREE Full text] [doi: 10.2196/jmir.1639] [Medline: 21212045]
24. Neilsen AS, Wilson RL. Combining e-mental health intervention development with human computer interaction (HCI) design to enhance technology-facilitated recovery for people with depression and/or anxiety conditions: an integrative literature review. Int J Ment Health Nurs 2019 Feb 22;28(1):22-39. [doi: 10.1111/inm.12527] [Medline: 30133096]
25. Kelders SM, Kok RN, Ossebaard HC, Van Gemert-Pijnen JE. Persuasive system design does matter: a systematic review of adherence to web-based interventions. J Med Internet Res 2012 Nov 14;14(6):e152 [FREE Full text] [doi: 10.2196/jmir.2104] [Medline: 23151820]
26. Austin PC, Steyerberg EW. The number of subjects per variable required in linear regression analyses. J Clin Epidemiol 2015 Jun;68(6):627-636 [FREE Full text] [doi: 10.1016/j.jclinepi.2014.12.014] [Medline: 25704724]
27. Cuijpers P, Weitz E, Cristea IA, Twisk J. Pre-post effect sizes should be avoided in meta-analyses. Epidemiol Psychiatr Sci 2017 Aug;26(4):364-368 [FREE Full text] [doi: 10.1017/S2045796016000809] [Medline: 27790968]
28. Hoyt WT, Del Re AC. Effect size calculation in meta-analyses of psychotherapy outcome research. Psychother Res 2018 May;28(3):379-388. [doi: 10.1080/10503307.2017.1405171] [Medline: 29179665]
29. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JP, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. PLoS Med 2009 Jul 21;6(7):e1000100 [FREE Full text] [doi: 10.1371/journal.pmed.1000100] [Medline: 19621070]
30. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders: Dsm-5. 5th Ed. Washington, D.C: American Psychiatric Association; 2013:1-970.

31. Hayes SC. Acceptance and commitment therapy, relational frame theory, and the third wave of behavioral and cognitive therapies. Behavior Therapy 2004;35(4):639-665. [doi: 10.1016/S0005-7894(04)80013-3]
32. Hunot V, Moore T, Caldwell D, Furukawa T, Davies P, Jones H, et al. 'Third wave' cognitive and behavioural therapies versus other psychological therapies for depression. Cochrane Database Syst Rev 2013 Oct 18(10):CD008704. [doi: 10.1002/14651858.CD008704.pub2] [Medline: 24142844]
33. Hofmann SG, Sawyer AT, Fang A. The empirical status of the "new wave" of cognitive behavioral therapy. Psychiatr Clin North Am 2010 Sep;33(3):701-710 [FREE Full text] [doi: 10.1016/j.psc.2010.04.006] [Medline: 20599141]
34. Higgins JP, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, Cochrane Bias Methods Group, Cochrane Statistical Methods Group. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. Br Med J 2011 Oct 18;343(oct18 2):5928 [FREE Full text] [doi: 10.1136/bmj.d5928] [Medline: 22008217]
35. Munder T, Barth J. Cochrane's risk of bias tool in the context of psychotherapy outcome research. Psychother Res 2018 May 09;28(3):347-355. [doi: 10.1080/10503307.2017.1411628] [Medline: 29224503]
36. Comprehensive Meta-Analysis Software (CMA). URL: https://www.meta-analysis.com/ [accessed 2020-11-15]
37. Andrews G, Cuijpers P, Craske MG, McEvoy P, Titov N. Computer therapy for the anxiety and depressive disorders is effective, acceptable and practical health care: a meta-analysis. PLoS One 2010 Oct;5(10):e13196 [FREE Full text] [doi: 10.1371/journal.pone.0013196] [Medline: 20967242]
38. Huedo-Medina TB, Sánchez-Meca J, Marín-Martínez F, Botella J. Assessing heterogeneity in meta-analysis: Q statistic or I^2 index? Psychol Methods 2006;11(2):193-206. [doi: 10.1037/1082-989x.11.2.193]
39. Borenstein M, Hedges LV, Higgins JP, Rothstein H. Regression in Meta-Analysis. 2017. URL: https://www.meta-analysis.com/downloads/MRManual.pdf [accessed 2020-12-31]
40. Duval S, Tweedie R. A nonparametric “Trim and Fill” method of accounting for publication bias in meta-analysis. J Am Stat Assoc 2000 Mar;95(449):89-98. [doi: 10.1080/01621459.2000.10473905]
41. Kampmann IL, Emmelkamp PM, Morina N. Meta-analysis of technology-assisted interventions for social anxiety disorder. J Anxiety Disord 2016 Dec;42:71-84. [doi: 10.1016/j.janxdis.2016.06.007] [Medline: 27376634]
42. Hair Jr JF, Black WC, Babin BJ, Anderson RE. Multivariate Data Analysis. 7. Ed. Harlow: Pearson; 2014:1-7.
43. Viechtbauer W, Cheung MW. Outlier and influence diagnostics for meta-analysis. Res Synth Methods 2010 Apr 04;1(2):112-125. [doi: 10.1002/jrsm.11] [Medline: 26061377]
44. Berger T, Hämmerli K, Gubser N, Andersson G, Caspar F. Internet-based treatment of depression: a randomized controlled trial comparing guided with unguided self-help. Cogn Behav Ther 2011 Dec;40(4):251-266. [doi: 10.1080/16506073.2011.616531] [Medline: 22060248]
45. Bücker L, Schnakenberg P, Karyotaki E, Moritz S, Westermann S. Diminishing effects after recurrent use of self-guided internet-based interventions in depression: randomized controlled trial. J Med Internet Res 2019 Oct 02;21(10):e14240 [FREE Full text] [doi: 10.2196/14240] [Medline: 31579014]
46. Clarke G, Reid E, Eubanks D, O'Connor E, DeBar LL, Kelleher C, et al. Overcoming depression on the internet (ODIN): a randomized controlled trial of an Internet depression skills intervention program. J Med Internet Res 2002 Dec;4(3):E14 [FREE Full text] [doi: 10.2196/jmir.4.3.e14] [Medline: 12554545]
47. Clarke G, Eubanks D, Reid E, Kelleher C, O'Connor E, DeBar LL, et al. Overcoming Depression on the Internet (ODIN) (2): a randomized trial of a self-help depression skills program with reminders. J Med Internet Res 2005 Jun;7(2):e16 [FREE Full text] [doi: 10.2196/jmir.7.2.e16] [Medline: 15998607]
48. Clarke G, Kelleher C, Hornbrook M, Debar L, Dickerson J, Gullion C. Randomized effectiveness trial of an internet, pure self-help, cognitive behavioral intervention for depressive symptoms in young adults. Cogn Behav Ther 2009 May;38(4):222-234 [FREE Full text] [doi: 10.1080/16506070802675353] [Medline: 19440896]
49. Dahne J, Collado A, Lejuez CW, Risco CM, Diaz VA, Coles L, et al. Pilot randomized controlled trial of a Spanish-language Behavioral Activation mobile app (¡Aptívate!) for the treatment of depressive symptoms among united states Latinx adults with limited English proficiency. J Affect Disord 2019 May 01;250:210-217. [doi: 10.1016/j.jad.2019.03.009] [Medline: 30870770]
50. Dahne J, Lejuez C, Diaz VA, Player MS, Kustanowitz J, Felton JW, et al. Pilot randomized trial of a self-help behavioral activation mobile app for utilization in primary care. Behav Ther 2019 Jul;50(4):817-827 [FREE Full text] [doi: 10.1016/j.beth.2018.12.003] [Medline: 31208690]
51. de Graaf LE, Gerhards SA, Arntz A, Riper H, Metsemakers JF, Evers SM, et al. Clinical effectiveness of online computerised cognitive-behavioural therapy without support for depression in primary care: randomised trial. Br J Psychiatry 2009 Jul;195(1):73-80 [FREE Full text] [doi: 10.1192/bjp.bp.108.054429] [Medline: 19567900]
52. Farrer L, Christensen H, Griffiths KM, Mackinnon A. Internet-based CBT for depression with and without telephone tracking in a national helpline: randomised controlled trial. PLoS One 2011 Nov;6(11):e28099 [FREE Full text] [doi: 10.1371/journal.pone.0028099] [Medline: 22140514]
53. Gräfe V, Moritz S, Greiner W. Correction to: health economic evaluation of an internet intervention for depression (deprexis), a randomized controlled trial. Health Econ Rev 2020 Jul 30;10(1):24 [FREE Full text] [doi: 10.1186/s13561-020-00280-1] [Medline: 32734480]

54. Hur JW, Kim B, Park D, Choi S. A scenario-based cognitive behavioral therapy mobile app to reduce dysfunctional beliefs in individuals with depression: a randomized controlled trial. Telemed J E Health 2018 Dec;24(9):710-716. [doi: 10.1089/tmj.2017.0214] [Medline: 29323626]
55. Lintvedt OK, Griffiths KM, Sørensen K, Østvik AR, Wang CE, Eisemann M, et al. Evaluating the effectiveness and efficacy of unguided internet-based self-help intervention for the prevention of depression: a randomized controlled trial. Clin Psychol Psychother 2013 Sep 02;20(1):10-27. [doi: 10.1002/cpp.770] [Medline: 21887811]
56. Löbner M, Pabst A, Stein J, Dorow M, Matschinger H, Luppa M, et al. Computerized cognitive behavior therapy for patients with mild to moderately severe depression in primary care: A pragmatic cluster randomized controlled trial (@ktiv). J Affect Disord 2018 Oct 01;238:317-326. [doi: 10.1016/j.jad.2018.06.008] [Medline: 29902736]
57. Lüdtke T, Pult LK, Schröder J, Moritz S, Bücker L. A randomized controlled trial on a smartphone self-help application (Be Good to Yourself) to reduce depressive symptoms. Psychiatry Res 2018 Nov;269:753-762. [doi: 10.1016/j.psychres.2018.08.113] [Medline: 30273901]
58. Lüdtke T, Westermann S, Pult LK, Schneider BC, Pfuhl G, Moritz S. Evaluation of a brief unguided psychological online intervention for depression: a controlled trial including exploratory moderator analyses. Internet Interv 2018 Sep;13:73-81. [doi: 10.1016/J.INVENT.2018.06.004]
59. McDermott R, Dozois DJ. A randomized controlled trial of internet-delivered CBT and attention bias modification for early intervention of depression. J Exp Psychopathol 2019 Apr 18;10(2):204380871984250. [doi: 10.1177/2043808719842502]
60. Meyer B, Berger T, Caspar F, Beevers CG, Andersson G, Weiss M. Effectiveness of a novel integrative online treatment for depression (Deprexis): randomized controlled trial. J Med Internet Res 2009 May 11;11(2):e15 [FREE Full text] [doi: 10.2196/jmir.1151] [Medline: 19632969]
61. Meyer B, Bierbrodt J, Schröder J, Berger T, Beevers CG, Weiss M, et al. Effects of an internet intervention (Deprexis) on severe depression symptoms: randomized controlled trial. Internet Interv 2015 Mar;2(1):48-59. [doi: 10.1016/j.invent.2014.12.003]
62. Mira A, Bretón-López J, García-Palacios A, Quero S, Baños RM, Botella C. An internet-based program for depressive symptoms using human and automated support: a randomized controlled trial. Neuropsychiatric Dis Treat 2017 Mar;Volume 13:987-1006. [doi: 10.2147/ndt.s130994]
63. Mohr DC, Duffecy J, Ho J, Kwasny M, Cai X, Burns MN, et al. A randomized controlled trial evaluating a manualized TeleCoaching protocol for improving adherence to a web-based intervention for the treatment of depression. PLoS One 2013 Aug;8(8):e70086 [FREE Full text] [doi: 10.1371/journal.pone.0070086] [Medline: 23990896]
64. Montero-Marín J, Araya R, Pérez-Yus MC, Mayoral F, Gili M, Botella C, et al. An internet-based intervention for depression in primary care in spain: a randomized controlled trial. J Med Internet Res 2016;18(8):e231 [FREE Full text] [doi: 10.2196/jmir.5695] [Medline: 27565118]
65. Moritz S, Schilling L, Hauschildt M, Schröder J, Treszl A. A randomized controlled trial of internet-based therapy in depression. Behav Res Ther 2012 Aug;50(7-8):513-521. [doi: 10.1016/j.brat.2012.04.006] [Medline: 22677231]
66. Morris RR, Schueller SM, Picard RW. Efficacy of a Web-based, crowdsourced peer-to-peer cognitive reappraisal platform for depression: randomized controlled trial. J Med Internet Res 2015 Mar;17(3):e72 [FREE Full text] [doi: 10.2196/jmir.4167] [Medline: 25835472]
67. Noguchi R, Sekizawa Y, So M, Yamaguchi S, Shimizu E. Effects of five-minute internet-based cognitive behavioral therapy and simplified emotion-focused mindfulness on depressive symptoms: a randomized controlled trial. BMC Psychiatry 2017 Dec 04;17(1):85 [FREE Full text] [doi: 10.1186/s12888-017-1248-8] [Medline: 28259151]
68. Schure MB, Lindow JC, Greist JH, Nakonezny PA, Bailey SJ, Bryan WL, et al. Use of a fully automated internet-based cognitive behavior therapy intervention in a community population of adults with depression symptoms: randomized controlled trial. J Med Internet Res 2019 Nov 18;21(11):e14754 [FREE Full text] [doi: 10.2196/14754] [Medline: 31738173]
69. Silverstone PH, Rittenbach K, Suen VY, Moretzsohn A, Cribben I, Bercov M, et al. Depression outcomes in adults attending family practice were not improved by screening, stepped-care, or online CBT during a 12-week study when compared to controls in a randomized trial. Front Psychiatry 2017;8:32 [FREE Full text] [doi: 10.3389/fpsyt.2017.00032] [Medline: 28373846]
70. Spek V, Nyklícek I, Smits N, Cuijpers P, Riper H, Keyzer J, et al. Internet-based cognitive behavioural therapy for subthreshold depression in people over 50 years old: a randomized controlled clinical trial. Psychol Med 2007 Dec;37(12):1797-1806. [doi: 10.1017/S0033291707000542] [Medline: 17466110]
71. Bakker D, Kazantzis N, Rickwood D, Rickard N. A randomized controlled trial of three smartphone apps for enhancing public mental health. Behav Res Ther 2018 Oct;109:75-83. [doi: 10.1016/j.brat.2018.08.003] [Medline: 30125790]
72. Kleiboer A, Donker T, Seekles W, van SA, Riper H, Cuijpers P. A randomized controlled trial on the role of support in internet-based problem solving therapy for depression and anxiety. Behav Res Ther 2015 Sep;72:63-71. [doi: 10.1016/j.brat.2015.06.013] [Medline: 26188373]
73. Moberg C, Niles A, Beermann D. Guided self-help works: randomized waitlist controlled trial of Pacifica, a mobile app integrating cognitive behavioral therapy and mindfulness for stress, anxiety, and depression. J Med Internet Res 2019 Jun 08;21(6):e12556 [FREE Full text] [doi: 10.2196/12556] [Medline: 31199319]

74. Powell J, Hamborg T, Stallard N, Burls A, McSorley J, Bennett K, et al. Effectiveness of a web-based cognitive-behavioral tool to improve mental well-being in the general population: randomized controlled trial. J Med Internet Res 2013;15(1):e2 [FREE Full text] [doi: 10.2196/jmir.2240] [Medline: 23302475]
75. Proudfoot J, Clarke J, Birch M, Whitton AE, Parker G, Manicavasagar V, et al. Impact of a mobile phone and web program on symptom and functional outcomes for people with mild-to-moderate depression, anxiety and stress: a randomised controlled trial. BMC Psychiatry 2013;13:312 [FREE Full text] [doi: 10.1186/1471-244X-13-312] [Medline: 24237617]
76. Shirotsuki K, Nonaka Y, Abe K, Adachi S, Adachi S, Kuboki T, et al. The effect for Japanese workers of a self-help computerized cognitive behaviour therapy program with a supplement soft drink. Biopsychosoc Med 2017 Sep 19;11(1):23 [FREE Full text] [doi: 10.1186/s13030-017-0109-5] [Medline: 28932258]
77. Twomey C, O'Reilly G, Byrne M, Bury M, White A, Kissane S, et al. A randomized controlled trial of the computerized CBT programme, MoodGYM, for public mental health service users waiting for interventions. Br J Clin Psychol 2014 Nov;53(4):433-450. [doi: 10.1111/bjc.12055] [Medline: 24831119]
78. Berger T, Urech A, Krieger T, Stolz T, Schulz A, Vincent A, et al. Effects of a transdiagnostic unguided internet intervention ('velibra') for anxiety disorders in primary care: results of a randomized controlled trial. Psychol Med 2017 Jan;47(1):67-80. [doi: 10.1017/S0033291716002270] [Medline: 27655039]
79. Boettcher J, Magnusson K, Marklund A, Berglund E, Blomdahl R, Braun U, et al. Adding a smartphone app to internet-based self-help for social anxiety: a randomized controlled trial. Comput Hum Behav 2018 Oct;87(6):98-108. [doi: 10.1016/j.chb.2018.04.052]
80. Botella C, Gallego MJ, Garcia-Palacios A, Guillen V, Baños RM, Quero S, et al. An internet-based self-help treatment for fear of public speaking: a controlled trial. Cyberpsychol Behav Soc Netw 2010 Aug;13(4):407-421. [doi: 10.1089/cyber.2009.0224] [Medline: 20712499]
81. Ciuca AM, Berger T, Crişan LG, Miclea M. Internet-based treatment for panic disorder: a three-arm randomized controlled trial comparing guided (via real-time video sessions) with unguided self-help treatment and a waitlist control. PAXPD study results. J Anxiety Disord 2018 May;56:43-55. [doi: 10.1016/j.janxdis.2018.03.009] [Medline: 29625834]
82. Donker T, Cornelisz I, van Klaveren C, van Straten A, Carlbring P, Cuijpers P, et al. Effectiveness of self-guided app-based virtual reality cognitive behavior therapy for acrophobia: a randomized clinical trial. JAMA Psychiatry 2019 Jul 01;76(7):682-690 [FREE Full text] [doi: 10.1001/jamapsychiatry.2019.0219] [Medline: 30892564]
83. Ivanova E, Lindner P, Ly KH, Dahlin M, Vernmark K, Andersson G, et al. Guided and unguided acceptance and commitment therapy for social anxiety disorder and/or panic disorder provided via the internet and a smartphone application: a randomized controlled trial. J Anxiety Disord 2016 Dec;44:27-35. [doi: 10.1016/j.janxdis.2016.09.012] [Medline: 27721123]
84. Kenardy J, McCafferty K, Rosa V. Internet-delivered indicated prevention for anxiety disorders: a randomized controlled trial. Behav Cognit Psychother 2003;31(3):279-289. [doi: 10.1017/S1352465803003047]
85. Lin LY, ` K, Kishimoto T, Rodriguez M, Qian M, Yang Y, et al. An internet-based intervention for individuals with social anxiety and different levels of Taijin Kyofusho in China. J Cross Cult Psychol 2020 May 21;51(5):387-402. [doi: 10.1177/0022022120920720]
86. McCall HC, Richardson CG, Helgadottir FD, Chen FS. Evaluating a web-based social anxiety intervention among university students: randomized controlled trial. J Med Internet Res 2018 Dec 21;20(3):e91 [FREE Full text] [doi: 10.2196/jmir.8630] [Medline: 29563078]
87. Oh J, Jang S, Kim H, Kim J. Efficacy of mobile app-based interactive cognitive behavioral therapy using a chatbot for panic disorder. Int J Med Inform 2020 Aug;140:104171. [doi: 10.1016/j.ijmedinf.2020.104171] [Medline: 32446158]
88. Powell J, Williams V, Atherton H, Bennett K, Yang Y, Davoudianfar M, et al. Effectiveness and cost-effectiveness of a self-guided internet intervention for social anxiety symptoms in a general population sample: randomized controlled trial. J Med Internet Res 2020 Jan 10;22(1):e16804 [FREE Full text] [doi: 10.2196/16804] [Medline: 31821151]
89. Titov N, Andrews G, Choi I, Schwencke G, Mahoney A. Shyness 3: randomized controlled trial of guided versus unguided internet-based CBT for social phobia. Aust N Z J Psychiatry 2008 Dec;42(12):1030-1040. [doi: 10.1080/00048670802512107] [Medline: 19016091]
90. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. Br Med J 2003 Sep 6;327(7414):557-560 [FREE Full text] [doi: 10.1136/bmj.327.7414.557] [Medline: 12958120]
91. Higgins J, Thomas J. Cochrane Handbook for Systematic Reviews of Interventions. URL: https://training.cochrane.org/handbook [accessed 2021-01-06]
92. Karyotaki E, Riper H, Twisk J, Hoogendoorn A, Kleiboer A, Mira A, et al. Efficacy of self-guided internet-based cognitive behavioral therapy in the treatment of depressive symptoms: a meta-analysis of individual participant data. JAMA Psychiatry 2017 Apr 01;74(4):351-359. [doi: 10.1001/jamapsychiatry.2017.0044] [Medline: 28241179]
93. Richards D, Richardson T. Computer-based psychological treatments for depression: a systematic review and meta-analysis. Clin Psychol Rev 2012 Jun;32(4):329-342. [doi: 10.1016/j.cpr.2012.02.004] [Medline: 22466510]
94. Spek V, Cuijpers P, Nyklícek I, Riper H, Keyzer J, Pop V. Internet-based cognitive behaviour therapy for symptoms of depression and anxiety: a meta-analysis. Psychol Med 2007 Mar;37(3):319-328. [doi: 10.1017/S0033291706008944] [Medline: 17112400]

95. Wright JH, Owen J, Richards D, Eells T, Richardson T, Brown G, et al. Computer-assisted cognitive-behavior therapy for depression: a systematic review and meta-analysis. J Clin Psychiatry 2019 Mar 19;80(2):19 [FREE Full text] [doi: 10.4088/JCP.18r12188] [Medline: 30900849]
96. Andrews G, Basu A, Cuijpers P, Craske M, McEvoy P, English C, et al. Computer therapy for the anxiety and depression disorders is effective, acceptable and practical health care: An updated meta-analysis. J Anxiety Disord 2018 Apr;55:70-78 [FREE Full text] [doi: 10.1016/j.janxdis.2018.01.001] [Medline: 29422409]
97. Olthuis JV, Watt M, Bailey K, Hayden J, Stewart S. Therapist-supported internet cognitive behavioural therapy for anxiety disorders in adults. Cochrane Database Syst Rev 2015 Mar 05(3):CD011565. [doi: 10.1002/14651858.CD011565] [Medline: 25742186]
98. Richards D, Richardson T, Timulak L, McElvaney J. The efficacy of internet-delivered treatment for generalized anxiety disorder: a systematic review and meta-analysis. Internet Interv 2015 Sep;2(3):272-282. [doi: 10.1016/j.invent.2015.07.003]
99. Miloff A, Marklund A, Carlbring P. The challenger app for social anxiety disorder: new advances in mobile psychological treatment. Internet Interv 2015 Nov;2(4):382-391. [doi: 10.1016/j.invent.2015.08.001]
100. Baker WL, White CM, Cappelleri J, Kluger J, Coleman C, Health Outcomes, Policy,Economics (HOPE) Collaborative Group. Understanding heterogeneity in meta-analysis: the role of meta-regression. Int J Clin Pract 2009 Oct;63(10):1426-1434. [doi: 10.1111/j.1742-1241.2009.02168.x] [Medline: 19769699]
101. Andersson G, Titov N, Dear BF, Rozental A, Carlbring P. Internet-delivered psychological treatments: from innovation to implementation. World Psychiatry 2019 Jan 02;18(1):20-28. [doi: 10.1002/wps.20610] [Medline: 30600624]
102. Titov N, Dear B, Nielssen O, Staples L, Hadjistavropoulos H, Nugent M, et al. ICBT in routine care: a descriptive analysis of successful clinics in five countries. Internet Interv 2018 Sep;13:108-115 [FREE Full text] [doi: 10.1016/j.invent.2018.07.006] [Medline: 30206525]

Abbreviations

CBT: cognitive behavioral therapy
ICBT: internet-delivered cognitive behavioral therapy
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PSD: persuasive systems design

Edited by G Eysenbach; submitted 23.01.21; peer-reviewed by A Pencer; comments to author 12.02.21; revised version received 14.02.21; accepted 11.04.21; published 29.04.21.

Please cite as:
McCall HC, Hadjistavropoulos HD, Sundström CRF
Exploring the Role of Persuasive Design in Unguided Internet-Delivered Cognitive Behavioral Therapy for Depression and Anxiety Among Adults: Systematic Review, Meta-analysis, and Meta-regression
J Med Internet Res 2021;23(4):e26939
URL: https://www.jmir.org/2021/4/e26939
doi:10.2196/26939
PMID:33913811

Review

mHealth Interventions for Self-Harm: Scoping Review

Bethany Cliffe[1], BSc, MSc; Jessica Tingley[2*], BSc, MSc; Isobel Greenhalgh[2*], BSc, MRes; Paul Stallard[1,2], PhD

[1]Department for Health, University of Bath, Bath, United Kingdom

[2]Child and Adolescent Mental Health Services, Oxford Health NHS Foundation Trust, Bristol, United Kingdom

[*]these authors contributed equally

Corresponding Author:
Bethany Cliffe, BSc, MSc
Department for Health
University of Bath
Claverton Down
Bath,
United Kingdom
Phone: 44 01225 388388
Email: bc731@bath.ac.uk

Abstract

Background: Self-harm is a growing issue with increasing prevalence rates; however, individuals who self-harm do not often receive treatment. Mobile health (mHealth) interventions are a possible solution to some of the barriers that individuals face when seeking support, and they have also been found to be effective in improving mental health. Thus far, reviews of mHealth interventions for self-harm have been limited by study type. Therefore, we determined that a broader scoping review will provide a more exhaustive understanding of mHealth interventions for self-harm.

Objective: This scoping review aims to identify mHealth interventions for self-harm within the literature, understand the types and features of interventions that have been developed and evaluated, highlight research findings around mHealth interventions for self-harm, and determine what outcomes are typically used to assess the efficacy of interventions.

Methods: A search was conducted using Embase, PubMed, PsycINFO, PsycEXTRA, Web of Science, and the Cochrane Library. Studies were included if they described an mHealth intervention designed to have a direct (ie, if the intervention was designed for self-harm or for people who self-harm) or indirect (ie, if self-harm was measured as an outcome) treatment effect and if the paper was available in English. There were no exclusion criteria based on the study design.

Results: A total of 36 papers were included in the review, and most of them were randomized controlled trials published within the last 4 years. The interventions were mostly smartphone apps and calling or texting services, with 62% (21/34) having underlying therapeutic models to inform the intervention content. They were generally shown to be promising and appealing, but only 5 were widely available for use. Outcomes focused on a reduction of self-harm and suicidality, mood, and the users' experiences of the intervention. Samples were typically nondiverse, and there was limited variety in the study designs and in the measurements of self-harm recovery.

Conclusions: Promising and appealing mHealth interventions have been developed but are not widely available. Research could benefit from greater diversity as well as a broader and more nuanced understanding of recovery from self-harm.

(J Med Internet Res 2021;23(4):e25140) doi:10.2196/25140

KEYWORDS

mHealth; self-harm; digital interventions; self-injury; NSSI; mobile phone

Introduction

Self-harm

The National Institute for Health and Care Excellence defines self-harm as any act of self-injury or poisoning, irrespective of the motivation behind the act [1]. It is a growing concern that can have great physical, psychological [2], and societal [3] costs. Notably, self-harm has been identified as a significant and persistent predictor of suicide [4]. In recent years, the lifetime prevalence of self-harm in the general English population has seen a sharp increase from 2.4% in 2000 to 6.4% in 2014; this was most common in young adult females who reported an increase in prevalence from 6.5% to 19.7% [5]. Furthermore,

given that individuals are often reluctant to disclose their self-harm behaviors due to the shame and stigma associated with it [6], rates of self-harm may be even higher than what these figures suggest [7].

Help for Self-harm

Tørmoen et al [8] surveyed 11,440 young people aged 14-17 years in Norway and found that only 34% of those who had self-harmed had ever sought professional help, indicating that help-seeking is low among those who self-harm [9,10]. Concerns over being perceived as "attention-seeking" or "crazy" and difficulty talking about their self-harm behaviors have been identified as some of the barriers to seek support by adolescents [11]. Furthermore, a lack of effective interventions for self-harm creates barriers to receive support when an individual seeks professional help. The National Institute for Health and Care Excellence guidelines advise against the use of pharmacological treatments for self-harm, instead recommending psychological interventions tailored to self-harm that may involve problem solving, cognitive behavioral or psychodynamic elements [12]. Despite this, to date, there is limited high-quality evidence suggesting that psychological or pharmacological interventions for self-harm are effective [13,14]. Moreover, increased pressure on services and the resulting difficulties with the availability and accessibility of these interventions can further prevent individuals from receiving support [15].

Use of Mobile Health

The use of mobile health (mHealth) may help overcome the barriers to treatment accessibility and availability. mHealth is a branch of eHealth, defined by the Global Observatory for eHealth as "medical and public health practice supported by mobile devices, such as mobile phones, patient monitoring devices, personal digital assistants (PDAs), and other wireless devices" [16]. Given the ubiquity of mobile phone ownership [17,18], providing mental health support in this way has the potential to reach many individuals who may not be receiving help for self-harm. mHealth offers multiple possibilities, including self-help smartphone apps, SMS text messaging with a support service, physical symptom tracking through wearable technologies, and receiving virtual therapy [19]. Clough and Casey [20] found that mHealth users felt that receiving virtual mHealth therapy was more beneficial compared with face-to-face therapy, particularly highlighting the freedom they felt to be completely open and honest with their therapist. mHealth tools also have merit as standalone interventions, with some studies reporting reductions in symptoms of mental health difficulties, including anxiety [21], schizophrenia [22], depression [23], and borderline personality disorder [24].

Studies investigating the efficacy of mHealth interventions for managing self-harm have also been reviewed but they are limited. Witt et al [25] identified only one study that included outcome measures of an mHealth intervention for self-harm, whereas Melia et al [26] identified 2. These reviews focused on randomized controlled trials (RCTs) and pre- and poststudies. Arshad et al [27] focused more closely on self-harm and identified 22 studies; however, this was still limited by their decision to exclude qualitative studies and those where self-harm was not the primary outcome. A broader scoping review will help to identify more mHealth tools available for managing self-harm and broaden our knowledge of them.

Aims

This scoping review aims to (1) identify mHealth interventions for self-harm within the literature, (2) understand the types and features of interventions that have been developed and evaluated, (3) highlight research findings around mHealth interventions for self-harm, and (4) determine what outcomes are typically used to assess the efficacy of interventions.

Methods

Overview

A detailed methodology can be found in the review protocol [28]. The following databases were searched in April 2020: Embase, PubMed, PsycINFO, PsycEXTRA, Web of Science, and the Cochrane Library. The reference lists of all papers identified in the searches were also screened. Multimedia Appendix 1 describes the full and detailed search strategy.

After duplicates were removed, the titles and abstracts were initially screened according to the aims of this review and were progressed for a full screening if they met the following inclusion criteria: (1) the study described an mHealth intervention (eg, SMS text messaging, phone calls, or websites accessible through a mobile device) designed to have a direct (ie, if the intervention was designed for self-harm or for people who self-harm) or indirect (if self-harm was measured as an outcome) treatment effect and (2) the paper was in English.

The full texts of the papers that met these criteria were then screened. Before screening the title and abstract, a pilot screening was performed by all 3 reviewers (BC, JT, and IG) on 20 papers selected at random. An interrated reliability check of at least 75% agreement was required to progress the papers to full screening. Initially, 80% agreement was achieved, and the remaining papers were briefly discussed until 100% agreement was achieved. Each paper was screened by at least two reviewers, with a third reviewer resolving any inconsistencies. During the title and abstract screening, there was 98% agreement between the 2 reviewers on each paper, with the others being discussed and resolved again. The full texts of the progressed papers were screened for eligibility.

Data Charting Process

The reviewers extracted predefined data regarding the study details (eg, year and country), participants (eg, number, age, and ethnicity), type of mHealth intervention, study design, measures, and outcomes. Both before and after data extraction, consistency was checked between reviewers using a random sample of papers. Data extraction was an iterative process in which categories were added or amended in accordance with the aims of the review.

Results

Identify mHealth Interventions for Self-Harm Within the Literature

The search results are summarized in Figure 1. A total of 295 papers were identified. After duplicates were removed, 78% (229/295) of titles and abstracts were left to be screened. About 35% (79/229) of titles and abstracts progressed to a full-text screening, resulting in 54% (43/79) of papers being excluded. Of these, 51% (22/43) did not present an mHealth intervention for self-harm, 21% (9/43) were systematic reviews that did not identify any papers not already identified by our search, and 19% (8/43) were protocols for studies for which a full text had since become available. Of the remaining 4 papers, 2 (5%) were replies or comments on other studies, 1 (2%) was a content analysis of apps that were commercially available with no references to an evidence base, and 1 (2%) was a description of an app that had already been identified. A total of 36 papers met the inclusion criteria and are summarized in Table 1.

Figure 1. Flowchart of search results. mHealth: mobile health.

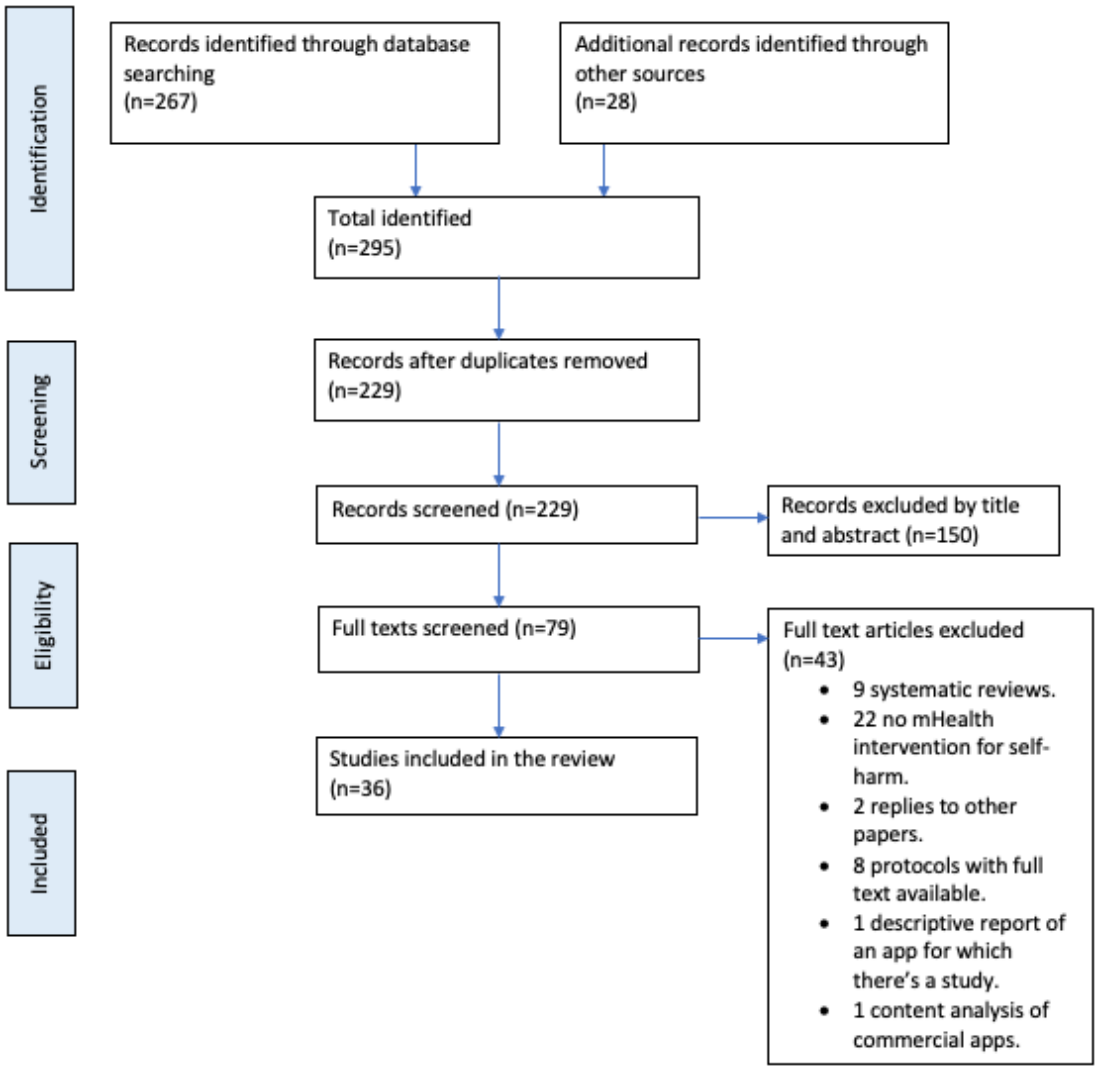


Table 1. Summary of papers (N=36).

Intervention details	Type of intervention	Design	Sample	Type of self-harm	Self-harm measure used	Development	Availability	Improvement in self-harm
Rebound (2019) [29]: *Accessible anytime*[a]	Social network with support from peer workers and clinical psychologists	RCT[b] (protocol)	Youth (aged 14-27 years) in recovery from major depressive disorder–Australia	Suicidal and nonsuicidal self-harm	Risk taking and self-harm inventory	Collaboration with consumers, youth representatives, and clinicians	Not widely available	No data
SIAM (2014) [30]: *9 texts sent, staggered for 6 months following A&E*[c] *discharge*	Supportive and monitoring texts	RCT (protocol)	Adults discharged after suicide attempt–France	Suicidal self-harm	Columbia-Suicide Severity Rating Scale	Developed by the authors	Not widely available	No data
A Virtual Hope Box (2015; 2017) [31,32]: *Accessible anytime*	Smartphone app to help with coping, distraction, relaxation, and positive thinking	Proof of concept; RCT	Army veterans at risk of self-harm or suicide–United States	Suicidal self-harm; suicidal and nonsuicidal self-harm	Semistructured interview; Beck Scale for Suicidal Ideation and Columbia-Suicide Severity Rating Scale	Developed by the Department of Defense National Centre for Telehealth and Technology, with feedback from target users	Free to download	Yes; no
EpxDepression (2017) [33]: *Daily EMA*[d] *for 2-4 months*	Text-based EMA that notifies care team of risk	Proof of concept	Adolescents–United States	Suicidal self-harm	Last question of PHQ[e]-9	Clinicians, researchers, and biostatisticians with feedback from target end users	Available for health care providers on subscription	Yes
Imaginator (2020) [34]: *Accessible anytime*	Smartphone app to promote self-management	RCT	Youth (aged 16-25 years) currently self-harming–United Kingdom	Suicidal and nonsuicidal self-harm	Strength of motivation for reducing self-harm scale, craving experience questionnaire for self-harm, self-harm imagery interview. Self-harm, frequency, severity, and self-efficacy for control were measured with items developed by the enhancer	Codeveloped with youth with lived experience of self-harm	Not widely available	Yes
Therapeutic Evaluative Conditioning (2016) [35]: *Accessible anytime*	Smartphone app to decondition self-harm	RCT	Adults with self-harm–international	Suicidal and nonsuicidal self-harm	Self-injurious thoughts and behaviors interview	Developed by the authors	No longer available	Yes, but did not persist at follow-up
BlueIce (2018) [36,37]: *Accessible anytime*	Smartphone app to help manage urges to self-harm	Pre-post phase 1 trial	Adolescents currently self-harming or at risk of self-harming attending CAMHS[f]—United Kingdom	Nonsuicidal self-harm	Semistructured interview; Self Harm data from clinical records	Coproduced with youth with lived experience of self-harm, with input from clinical staff, academics, and app developers	Available for CAMHS on subscription	Yes
Unnamed (2018) [38]: *Accessible anytime*	Smartphone app for mood monitoring and distraction	Development study	Aged 18-25 years—Australia	Nonsuicidal self-harm	N/A[g]	Designed with target users and clinicians	Not widely available	No data

Intervention details	Type of intervention	Design	Sample	Type of self-harm	Self-harm measure used	Development	Availability	Improvement in self-harm
ClinTouch CareLoop enhance management (2014) [39]: *Accessible anytime; 44 daily ecological momentary assessment alerts*	Smartphone app for mood monitoring and alerting care team of risk	RCT (protocol)	Adults with psychotic disorders—United Kingdom	Suicidal and nonsuicidal self-harm	Unclear	Codeveloped with service users, clinicians, academics, and health professionals	Free to download	No data
Unnamed (2016) [40]: *52 text messages sent twice weekly*	Supportive and informative texts	Intervention study	Adults with suicidal ideation—Japan	Nonsuicidal self-harm	Researchers developed their own questionnaire measuring the presence of self-harm	Psychiatry specialists	Not widely available	Yes
A-CHESS (2017) [41]: *Accessible anytime*	Smartphone app containing safety plan, social network, resources, and interaction dashboard with their therapist	Qualitative study of experience	Presented to emergency department following self-harm—Canada	Suicidal and nonsuicidal self-harm	Last question of PHQ-9, hospital presentations for self-harm	Developed by the network for improvement of addiction treatment with user feedback from focus groups	Available for health care providers on subscription	Yes
Safe Storage Intervention (2015) [42]: *Texts sent daily*	Social contact via phone conversations and texts	Cost-effective analysis (protocol)	Participants recruited to another study–Sri Lanka	Suicidal and nonsuicidal self-harm	Hospital presentations and coroner's data	No data	Not widely available	No data
Brief mobile treatment [43]: *10 staggered phone calls after discharge for 24 weeks; audio messages accessible anytime; text reminders sent weekly*	Phone calls to monitor mood, meditation audio messages, and text reminders of treatment elements	RCT	Aged 15-74 years admitted to hospital after self-harm—Sri Lanka	Suicidal self-harm	Beck Scale for Suicidal Ideation	No data	Not widely available	No
No name (2013) [44]: *Staggered texts that start daily and decline gradually to weekly over 3 months*	Supportive text messages	RCT	Adults in A&E after self-harm—Ireland	Suicidal and nonsuicidal self-harm	Unclear—self-harm repetition	No data	Not widely available	No data
No name (2017) [45]: *Accessible anytime*	Smartphone app to provide care messages, resources, health care contacts, and self-help exercises	RCT (protocol)	Adult in A&E after self-harm—Hong Kong	Suicidal and nonsuicidal self-harm	Hospital presentations and coroner's data	No data	Not widely available	No data
Unnamed (2018) [46]: *12-month treatment period, no further detail*	Smartphone app to augment delivery of problem-solving therapy	RCT (protocol)	Men in A&E after self-harm–Canada	Suicidal and nonsuicidal self-harm	Hospital presentations	No data	Not widely available	No data
TeenTEXT (2016) [47]: *Receiver sets schedule of when texts will be received*	Supportive texts written by the receiver	Feasibility study	Adolescents currently self-harming—United Kingdom	Nonsuicidal self-harm	N/A	Designed with service users with history of self-harm, carers, and clinicians	Not widely available	No data (terminated)

Intervention details	Type of intervention	Design	Sample	Type of self-harm	Self-harm measure used	Development	Availability	Improvement in self-harm
DBT[h] Coach (2016) [48]: *Accessible anytime*	Smartphone app that provides coaching in DBT skills	Pilot	Individuals seeking DBT—United States	Suicidal and nonsuicidal self-harm	SITBI[i], semistructured interviews	Involved DBT experts, target end users, and their clinicians	Available for users on subscription	Yes
Unnamed (2019) [49]: *Delivered over 10-12 sessions*	Audio or video calls to deliver problem-solving CBT[j]	RCT	Aged 16-30 years, depression and self-harm—location unknown	Nonsuicidal self-harm	Urgency Perseverance Premeditation Sensation-seeking Impulsive Inventory urgency subscale	No data	Not widely available	No data (terminated)
MyPlan (2017) [50]: *Accessible anytime*	Smartphone app to store safety plan	RCT (protocol)	Experiencing self-harm—Denmark	Suicidal and nonsuicidal self-harm	Self-reported, no further information	Developed by Skovgaard Larsen et al [51]	Yes (in Denmark and Norway)	No data
SMS SOS (2019) [52]: *Staggered texts that start 48 h after discharge from A&E, declining to monthly for 12 months*	Supportive text messages	RCT (protocol)	In A&E after self-harm—Australia	Suicidal and nonsuicidal self-harm	Hospital presentations	Developed with people with lived experience of self-harm or mental health problems	Not widely available	No data
Crisis Text Line (2020) [53]: *Accessible anytime*	Texting with a crisis counselor	Trend analysis	No criteria—international	Suicidal and nonsuicidal self-harm	N/A	Crisis text line is a global organization	Available globally	N/A
SPARX (2020) [54]: *Must complete program within 6 weeks*	Game-style smartphone app	RCT (protocol)	Year 8 school students—Australia	Suicidal and nonsuicidal self-harm	Self-harm questionnaire	Developed with young people	Yes (in New Zealand)	No data
ERITA (2018) [55]: *12 weeks to complete 11 modules*	Website and companion app to receive treatment for emotion regulation	Pilot	Adolescents with self-harm—Sweden	Nonsuicidal self-harm	Deliberate self-harm inventory	Developed by the authors	Not widely available	Yes
CATCH-IT[k] (2009) [56]: *Recipients had 3 phone calls, no further information*	Therapy provided via website with motivational interviewing phone calls	RCT	Aged 14-21 years with subthreshold depression—United States	Nonsuicidal self-harm	PHQ-A	Developed by the authors	Not widely available	Yes
Unnamed (1999) [57]: *Accessible anytime for 6 months following discharge*	Telephone crisis consultation with on-call psychiatrist	RCT	Adult inpatients with self-harm—United Kingdom	Nonsuicidal self-harm	Hospital presentations	No data	Not widely available	No
Unnamed (2020) [58]: *A module per week for 6 weeks*	Web-based intervention to provide CBT-based modules	RCT	Turkish adults with suicidal ideation—United Kingdom and Netherlands	Suicidal self-harm	Self-harm questionnaire	Developed by van Spijker et al [59] adapted by the authors	Not yet, but will be if findings are positive	Yes
Unnamed (2018) [60]: *Encouraged to write 5mina day for 28 days*	Web-based diary	RCT	Adults on an online forum who self-harm—international	Suicidal and nonsuicidal self-harm	SITBI	Developed by the author	Technique can be adopted	Yes

Intervention details	Type of intervention	Design	Sample	Type of self-harm	Self-harm measure used	Development	Availability	Improvement in self-harm
Uncut (2014) [61]: *Ecological momentary assessment*	App to track mood and practice DBT skills, therapist can monitor progress	Development study	Self-harm experts—Austria, Germany, and United States	Nonsuicidal self-harm	N/A	Developed with international psychological experts	No data	No data
Living under control (2017) [62]: *A module per week for 6 weeks*	Web-based intervention providing 6 modules around managing thoughts and feelings	RCT (protocol)	Adults—Denmark	Suicidal and nonsuicidal self-harm	Hospital records and self-report questionnaire (no further information)	Developed by van Spijker et al [63] with mental health professionals who work with people experiencing suicidality	Not yet, but will be if findings are positive	No data
Crisis Care (2017) [64]: *Accessible anytime*	Smartphone app containing coping skills, distraction activities, and a help me now section	Pilot study	Adolescent psychiatry outpatients—United States	Suicidal self-harm	N/A	Developed by the author	Not widely available	Yes
Monsenso (2020) [65]: *Accessible anytime*	Smartphone app providing DBT skills and mood monitoring; links with clinicians database	RCT (protocol)	Adults with borderline personality disorder and self-harm—Denmark	Suicidal and nonsuicidal self-harm	Self-harm inventory	Developed by the author	Available for health care providers on subscription	No data
Unnamed (2019) [66]: *No details yet*	Supportive text messages	Development study	Adolescents with self-harm—China	Nonsuicidal self-harm	N/A	Developed with youth with lived experience of self-harm	No data	No data
Unnamed (2020) [67]: *Unlimited access every evening*	Peer-supported hotline	Trend analysis	Teenagers—United States	Suicidal and nonsuicidal self-harm	N/A	Founded by mental health professionals	Available to teenagers in the United States	N/A

[a]Italics refer to the intervention duration.

[b]RCT: randomized controlled trial.

[c]A&E: accident & emergency.

[d]EMA: ecological momentary assessment.

[e]PHQ: patient health questionnaire.

[f]CAMHS: Child and adolescent mental health services.

[g]N/A: not applicable.

[h]DBT: dialectical behavior therapy.

[i]SITBI: self-injurious thoughts and behaviors interview.

[j]CBT: cognitive behavioral therapy.

[k]CATCH-IT: Competent Adulthood Transition with Cognitive Behavioural Humanistic and Interpersonal Training.

The 36 papers related to 35 separate studies, 2 of which were published from the same study with one paper detailing the quantitative findings [36] and the other the qualitative findings [37]. They were therefore not removed as duplicates; however, as the study details were the same, the study characteristics were merged to not present the same information twice.

A total of 12 papers reported protocols. The corresponding authors were contacted to request any update or preliminary findings; a preprint paper was received from one, so the protocol was replaced with this and the data were extracted. The remaining 11 protocols had no further data at the time of this review (April-June 2020).

The most common study design was the RCT (20/35, 57%), with most papers published between 2014 and 2020 (33/35, 94%). Studies were most commonly conducted in Europe (14/35, 40%) and North America (7/35, 20%). Most studies measured self-harm using various self-report questionnaires (18/35, 51%) and looked at both suicidal and nonsuicidal self-harm (19/35, 54%).

Participants

The participant data are summarized in Table 2. Sample sizes ranged from 3 to 122,909, involving mostly clinical cohorts (24/35, 69%), and approximately half were adult samples (18/35, 51%), with one focusing specifically on young adults (aged 18-25 years). None of the studies included children aged <12 years and adults aged >65 years. Gender was reported in 18 papers, with the majority including more females than males (14/18, 78%); only 1 study included a nonbinary participant (1/18, 6%). Most papers provided no data on ethnicity (25/35, 71%) and those that did recruited mostly White people (8/10, 80%). Presenting problems screened for or required as inclusion criteria among the samples were primarily self-harm (26/35, 74%), depression (22/35, 63%), suicidal ideation (15/35, 43%), and suicide attempts (15/35, 43%).

Table 2. Participant characteristics (n=35).[a]

Characteristic	n (%)	Study
Age (years)		
Adults	18 (54)	[30-32,35,38-41,44-46,48,57,58,60-62,65]
Adolescents (up to 18 years)	8 (23)	[33,36,37,47,54,55,64,66,67]
Adolescents and adults	6 (17)	[29,34,43,49,52,56]
No age restriction	1 (3)	[50]
Unknown	2 (5)	[42,53]
Population		
Clinical	24 (69)	[29,30,32,33,36-41,43-50,52,55-57,64-66]
Community	9 (26)	[31,35,42,53,54,58,60,62,67]
Clinical and community	1 (3)	[34]
Mental health professionals	1 (3)	[61]
Presenting problems or inclusion criteria		
Self-harm	26 (74)	[29,30,34-37,40,41,43-50,52,53,55-58,60,62,65-67]
Depression	22 (63)	[29,31-34,36-38,40,41,43,45,48-50,53-56,60,62,66,67]
Anxiety	13 (37)	[29,31,32,34,36,37,41,48,49,53-55,66,67]
Suicidal ideation	15(43)	[29,30,32,34,35,40,43-45,50,53,58,60,62,64]
Suicide attempt	15 (43)	[29,30,32,34,35,44-46,48,50,53,57,60,64,65]
Borderline personality disorder	5 (14)	[31,32,48,55,65]
Psychosis	4 (11)	[32,39,40,66]
Substance or alcohol use	4 (11)	[32,41,43,57]
Eating disorders	2 (6)	[31,55]
Sleep disorders	2 (6)	[32,54]
Neurological disorders	1 (3)	[32]
Somatoform disorder	1 (3)	[32]

[a]Grist et al (2018) [36] and Stallard et al (2018) [37] are separate papers from the same study.

Understand the Types and Features of Interventions That Have Been Developed and Evaluated

Characteristics of mHealth Interventions

Intervention characteristics are summarized in Table 3 and relate to 34 interventions. As mentioned earlier, another 2 papers reported on different trials of the same intervention. Most studies described apps (15/34, 44%) or texting or calling services (13/34, 38%), and most interventions required a mobile phone (16/34, 47%) or a smartphone (11/34, 32%), whereas the rest required any internet-enabled device (4/34, 12%) or an iPod touch (1/34, 3%). Approximately half of the interventions did not include any face-to-face support (16/34, 47%); of these, 9 were standalone interventions (9/16, 56%), 7 included an element of human support provided digitally (ie, texts from a clinician; 7/16, 21%), and 2 exclusively provided digital support (ie, a hotline; 2/16, 13%). A range of underpinning therapeutic models informing the intervention content were reported, with 21 studies (21/34, 62%) citing at least one approach and cognitive behavioral therapy being the most common (10/34, 29%). Supportive messages or phone calls were the most common elements among the interventions (14/34, 41%). Only 3 (9%) papers specified that the interventions contained a safety plan.

Table 3. Intervention characteristics (n=34).[a]

Intervention characteristics	n (%)	Study
Intervention type		
Apps	15 (44)	[31,32,34,36,38,39,41,45,48,50,54,61,64,65]
Texting or calling services	13 (38)	[30,33,40,42-44,47,49,52,53,57,66,67]
Websites or web-based therapies	4 (12)	[55,56,58,62]
Web-based diary	1 (3)	[60]
Social network	1 (3)	[29]
Device		
Mobile phone	16 (47)	[30,33,39,40,42-45,47,49,52,53,57,65-67]
Smartphone	11 (32)	[31,32,34-38,41,46,50,54,55,60,61,64]
Any internet-enabled device	4 (12)	[29,56,58,62]
iPod touch (or iPhone)	1 (3)	[48]
Human support included		
Face-to-face provided	12 (35)	[31,32,34,36-38,41,43,46-48,50,52,65]
Digital support provided	7 (21)	[29,49,53,55,57,58,67]
No support	9 (27)	[30,33,35,44,54,60,62,64,66]
Both compared	2 (6)	[45,56]
Not specified	4 (12)	[39,40,42,61]
Underpinning therapeutic model		
Cognitive behavioral therapy	10 (29)	[31,32,36-38,47,49,54,56,58,62,64]
Dialectical behavior therapy	6 (18)	[31,32,36-38,48,61,65]
Cognitive	3 (9)	[29,34,50]
Behavioral	3 (9)	[35-37,56]
Mindfulness	2 (6)	[29,36,37]
Problem-solving therapy	2 (6)	[41,46]
Acceptance and commitment therapy	1 (3)	[56]
Interpersonal psychotherapy	1 (3)	[56]
Autobiographic self enhancement	1 (3)	[60]
Not specified	13 (38)	[30,33,39,40,42-45,52,53,57,66,67]
Features		
Supportive messages or phone calls	14 (41)	[30,33,34,40,42-45,47,52,53,57,66,67]
Coping skills	9 (27)	[40,48,54,56,58,61,62,64,65]
Mood diaries	8 (24)	[34,36-39,45,60,61,65]
Links to helplines, services, or caregivers	8 (24)	[33,40,41,44-46,53,64]
Problem-solving techniques	6 (18)	[29,41,43,46,54,56]
Alerts to clinicians	6 (18)	[33,39,41,47,61,65]
Mood lifting or physical activities	6 (18)	[34,36,37,44,54,56,64]
Information and psychoeducation	5 (15)	[29,41,44,54,65]
Relaxation and meditation	5 (15)	[31,32,36,38,45,64]
Mindfulness	5 (15)	[29,31,32,36,37,48,58]
Thought challenging	5 (15)	[36,37,54,56,58,62]
Photos, music, and other media	4 (12)	[31,32,34,36,37,41]
Medication and intervention reminders	4 (12)	[34,39,40,43]

Intervention characteristics	n (%)	Study
Social and peer support	3 (9)	[29,41,67]
Safety plan	3 (9)	[41,50,62]
Distraction methods	3 (9)	[31,32,38,64]
Games	3 (9)	[31,35,54]

[a]Bush et al (2015) [31] and Bush et al (2017) [32] relate to the same intervention; Grist et al (2018) [36] and Stallard et al (2018) [37] are separate papers from the same study.

Development and Availability

Half of the mHealth interventions were developed with multiple collaborators (17/34, 50%), including mental health professionals (10/34, 29%), target end users (11/34, 32%), and companies or organizations (3/34, 9%). Most interventions were not currently available to the public (20/34, 59%). Conversely, only 15% (5/34) are publicly available, whereas others are available for purchase by health care professionals or services (4/34, 12%) or users (1/34, 3%). Table 1 provides further details on the intervention development and availability.

Highlight Research Findings Around mHealth Interventions for Self-Harm

Study Findings

Of the 19 papers that reported outcomes, 14 (74%) reported positive findings postintervention. Of the 5 studies that did not report positive findings, 2 (26%) were terminated during recruitment due to a lack of feasibility. One study noted that this was due to high levels of depression and the reluctance of participants who self-harmed to engage with mental health services [49], whereas the other suggested it was a good intervention but Child and Adolescent Mental Health Services was the wrong setting due to clinicians' time constraints [47]. Furthermore, 3 studies reported no significant effect of the intervention on self-harm recovery [32,43,57], one of which cited past episodes of self-harm as a barrier to efficacy [57].

Clinician and parental attitudes were typically favorable toward the interventions [31,32,38,41,47,64], with only 2 papers reporting concerns [38,47]. The identified benefits include promoting self-efficacy [32,58], helping difficult disclosure [33], immediate access [43], time and cost benefits [55], encouraging help-seeking [40], being useful in crises [64,66], and having a positive influence on the therapeutic alliance in the face-to-face element within blended approaches [41,58]. Barriers were not commonly mentioned in the papers, although 2 papers did note that digital interventions that were administered by a mental health worker posed challenges due to the lack of engagement with mental health services among people who self-harm [47,49].

Determine What Outcomes Are Typically Used to Assess the Efficacy of Interventions

The study outcomes are summarized in Table 4. Most studies had multiple outcomes related to self-harm (21/35, 60%), suicide attempts (19/35, 54%), suicidal ideation (16/35, 46%), intervention experience (16/35, 46%), and engagement with the intervention (13/35, 37%). Other mental health issues such as depression (16/35, 46%) and anxiety (9/35, 26%) were also considered. Outcomes related to self-harm mostly focused on episode frequency (14/21, 67%), whereas others focused on repeated presentations to hospital (5/21, 24%) or self-harm thoughts or urges (4/21, 19%).

Table 4. Study outcomes (n=35).[a]

Outcome	n (%)	Study
Mental health		
SH[b] frequency	14 (40)	[29,34-37,39,40,42,44,45,54-60,62]
Presentations to hospital for SH	5 (14)	[40,45,46,52,57]
SH thoughts or urges	4 (11)	[34,48,56,60]
SH (specifics unclear)	2 (6)	[50,58]
Suicidal ideation	16 (46)	[29,30,32,34,35,40,43-45,49,50,54,58,60,62,65]
Suicide attempts	19 (54)	[29,30,32,34,35,40,42,43,45,46,48-50,52,57,58,60,62,65]
Depression	16 (46)	[29,30,34,36-38,43,45,49,50,53,54,56,58,60,62,66]
Anxiety	9 (26)	[29,30,34,36,37,49,53,54,58,62]
Eating disorders	3 (9)	[30,53,54]
Borderline personality disorder	2 (6)	[55,65]
Psychosis	1 (3)	[54]
Other well-being		
Other mental well-being	7 (20)	[29,48,53-55,60,62]
Interpersonal issues	7 (20)	[29,32,43,45,49,53,54]
Hopelessness	6 (17)	[44,45,49,50,58,62]
Sleep	3 (9)	[29,33,54]
Alcohol or substance use	3 (9)	[43,53,54]
Quality of life	3 (9)	[29,58,62]
Other self-destructive behaviors	2 (6)	[54,55]
Intervention feasibility and acceptability		
Experience	16 (46)	[29-34,36,37,44,47,48,50,54,58,60,62,64,66]
Engagement	13 (37)	[29,31-34,39,44,45,48,53,54,56,67]
Health care costs	1 (3)	[46]

[a]Grist et al (2018) [36] and Stallard et al (2018) [37] are separate papers from the same study.

[b]SH: self-harm.

Discussion

Principal Findings

This scoping review identified 36 papers relating to 34 separate mHealth tools for managing self-harm. Papers were primarily RCTs and protocols published in Europe or North America within the last 6 years. This recent increase in papers reflects the growing interest in developing evidence-based mHealth interventions to improve access to psychological therapies. The large number of protocols suggests that this trend will continue as more findings are published in the coming years. Participants were mostly White adult females recruited from clinical populations, with only one nonbinary participant across all the studies, and many did not report the ethnicity of their participants. This is concerning given the high prevalence of self-harm found in both ethnic [68] and gender minorities [69]. It is possible that other nonbinary or gender-diverse individuals participated in these studies, but they were not truly represented in the way that the researchers assessed or reported participant demographics.

Depression and anxiety were the 2 most highly studied comorbidities, which is consistent with research suggesting strong correlations between these disorders and self-harm [70]. The interventions most commonly studied were not blended with any face-to-face support and were mostly text or call-based services or apps used on mobile phones or smartphones. Considering the ubiquity of both mobile phones [17,71] and smartphones [71], this is a positive finding and suggests that there are indeed mHealth interventions that could be more widely accessed. Interestingly, no interventions designed for use on other devices, such as wearable technologies, have been identified. This is despite research suggesting that they are acceptable for treating mental health issues among those who do not typically engage with mental health services [72]. However, this is still a relatively new area, and wearable devices designed to treat or help manage mental health difficulties are predicted to increase over the next few years [73].

Several studies did not specify any underpinning therapeutic models informing the content of their intervention, which raises concerns given the unhelpful and even dangerous advice that

has been found in other freely available mental health apps [74]. Similarly, less than a third of the interventions were developed with individuals who have lived experience of self-harm, despite evidence suggesting that this could lead to more effective interventions being developed. Given the expertise of those with lived experience of self-harm, their input is essential [75,76].

Most studies testing interventions reported overall promising findings, suggesting that mHealth can be a viable tool for people struggling with self-harm. The study outcomes largely promoted a reduction in self-harm frequency over and above others, such as a reduction in self-harm urges or severity. Although reduced frequency is a common measure of self-harm recovery within research, it has been argued that it is not advisable to rely solely on this; a reduction in self-harm episodes may mean that each episode has become more severe or has been replaced with a different type of self-destructive behavior [77]. Overall, there needs to be a more consistent framework to assess outcomes from self-harm research, with further research looking into the maintenance of positive outcomes. Similarly, there was considerable variation in the tools used to measure self-harm across the studies. This makes comparisons between studies difficult; therefore, greater consistency may also be beneficial.

Another commonly observed finding was the clinicians' favorable attitudes toward the intervention. This reflects overall positive attitudes and an eagerness to incorporate technology into practice found in research on clinicians' attitudes toward technology in mental health care [78]. This is important considering that there is research highlighting that clinicians' attitudes are pivotal in intervention implementation [79].

A barrier to the implementation of blended interventions that required clinician involvement was the lack of engagement that people who self-harm have with mental health services. This is consistent with the body of literature that corroborates the lack of professional help-seeking among people who self-harm [8-11] and further emphasizes the need for mHealth interventions for self-harm that individuals can access easily and discretely. Despite many studies having positive outcomes, it seems that not many of the interventions studied are readily available to the general public or even to those attending mental health services. Moreover, few papers made this information apparent, so information on availability was sought from internet and app store searches and by contacting the authors of the papers.

Implications for Future Research

This review identified several limitations in the current literature on mHealth tools for managing self-harm. Notably, research thus far has been limited to White adult females from western societies, yet self-harm has been identified as a significant issue among minority groups [68,69], highlighting the need to diversify research by recruiting understudied groups such as males and minority populations who may also benefit from mHealth interventions. It is also important for future studies to report the ethnicity of participants. Similarly, participants have mostly been from clinical samples; however, given that people who self-harm do not often seek professional help, it is possible that findings from clinical samples may not necessarily be generalizable to wider populations who self-harm. Therefore, it is important to assess the efficacy of interventions in community samples. It is also essential for more intervention developers to collaborate with people who have lived experience of self-harm, given the limited instances thus far highlighted in this review.

Another notable point is that the measurements of self-harm recovery were typically restricted to a reduction in the frequency of episodes and, although this may be a useful assessment, it may also be worth considering measuring other elements as well, such as the severity of self-harming episodes or any other substitute self-destructive behaviors.

Another consideration to take forward from this review is the reliance on RCTs to evaluate digital interventions. Although RCTs have long been considered the gold standard method, the pace at which digital interventions develop and evolve means that the data can be outdated before the trial has been completed [80-82]. Following this, it may be prudent for future research to also consider different study designs, such as pre- and posttests that can keep up with the rapid development of digital interventions. Furthermore, RCTs can indicate whether an intervention is effective but may be limited in their ability to explore the reason [83]. Therefore, it may also be beneficial for future research to apply the NIMH's experimental therapeutics approach or use qualitative studies that can contribute further to the understanding of how and why certain interventions are effective as well as whether they are safe and do no harm. Similarly, the application of the experimental therapeutic approach may contribute further to the understanding of how and why certain interventions are effective [84]. Future work should also focus on the implementation and dissemination of effective interventions, following the lack of availability of interventions within the papers in this review.

Limitations

As the purpose of this review was to collate the data available on mHealth tools for managing self-harm, there was no scope to conduct a quality assessment of the included studies. This means that we cannot verify that the research included here is of sufficient quality to draw concrete inferences from. This review may also have been limited by the decision to not include any papers for which there was no English version available. Nonetheless, this review offers insight into the current evidence base for mHealth interventions for self-harm.

Conclusions

This review has synthesized the current evidence for mHealth tools for managing self-harm. Overall, there are useful interventions that have been developed to promote recovery from self-harm. However, certain limitations pose challenges in drawing firm conclusions from the included studies. Suggestions for how future research can improve upon this have been made, in the hope of developing a robust evidence base so that clinicians and users are better equipped to make informed decisions about which mHealth tools to use. This will hopefully help to overcome some of the barriers that people who self-harm face in accessing support.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Full search strategy.
[DOC File , 28 KB - jmir_v23i4e25140_app1.doc]

References

1. Self-Harm. National Institute for Health and Care Excellence. 2013. URL: https://www.nice.org.uk/guidance/qs34 [accessed 2020-08-17]
2. Lundh L, Wångby-Lundh M, Bjärehed J. Deliberate self-harm and psychological problems in young adolescents: evidence of a bidirectional relationship in girls. Scand J Psychol 2011 Oct;52(5):476-483. [doi: 10.1111/j.1467-9450.2011.00894.x] [Medline: 21585393]
3. Tsiachristas A, McDaid D, Casey D, Brand F, Leal J, Park A, et al. General hospital costs in England of medical and psychiatric care for patients who self-harm: a retrospective analysis. Lancet Psych 2017 Oct;4(10):759-767 [FREE Full text] [doi: 10.1016/S2215-0366(17)30367-X] [Medline: 28890321]
4. Mars B, Heron J, Klonsky ED, Moran P, O'Connor RC, Tilling K, et al. Predictors of future suicide attempt among adolescents with suicidal thoughts or non-suicidal self-harm: a population-based birth cohort study. Lancet Psychiatry 2019 Apr;6(4):327-337 [FREE Full text] [doi: 10.1016/S2215-0366(19)30030-6] [Medline: 30879972]
5. McManus S, Gunnell D, Cooper C, Bebbington PE, Howard LM, Brugha T, et al. Prevalence of non-suicidal self-harm and service contact in England, 2000-14: repeated cross-sectional surveys of the general population. Lancet Psychiatry 2019 Jul;6(7):573-581 [FREE Full text] [doi: 10.1016/S2215-0366(19)30188-9] [Medline: 31175059]
6. Gratz KL. Risk Factors for and Functions of Deliberate Self‐Harm: An Empirical and Conceptual Review. Clinical Psychology: Science and Practice 2006 May 11;10(2):192-205. [doi: 10.1093/clipsy.bpg022]
7. Whitlock J, Eckenrode J, Silverman D. Self-injurious behaviors in a college population. Pediatrics 2006 Jun;117(6):1939-1948. [doi: 10.1542/peds.2005-2543] [Medline: 16740834]
8. Tørmoen AJ, Rossow I, Mork E, Mehlum L. Contact with child and adolescent psychiatric services among self-harming and suicidal adolescents in the general population: a cross sectional study. Child Adolesc Psychiatry Ment Health 2014;8:13 [FREE Full text] [doi: 10.1186/1753-2000-8-13] [Medline: 24742154]
9. Michelmore L, Hindley P. Help-seeking for suicidal thoughts and self-harm in young people: a systematic review. Suicide Life Threat Behav 2012 Oct;42(5):507-524. [doi: 10.1111/j.1943-278X.2012.00108.x] [Medline: 22889130]
10. Nada-Raja S, Morrison D, Skegg K. A population-based study of help-seeking for self-harm in young adults. Aust N Z J Psychiatry 2003 Oct;37(5):600-605. [Medline: 14511089]
11. Fortune S, Sinclair J, Hawton K. Help-seeking before and after episodes of self-harm: a descriptive study in school pupils in England. BMC Public Health 2008 Oct 24;8:369 [FREE Full text] [doi: 10.1186/1471-2458-8-369] [Medline: 18947435]
12. NICE. Self-harm in over 8s: long-term management. National Institute for Health and Care Excellence. 2011 Nov 23. URL: https://www.nice.org.uk/guidance/cg133 [accessed 2020-05-15]
13. Hawton K, Witt KG, Taylor Salisbury TL, Arensman E, Gunnell D, Hazell P, et al. Pharmacological interventions for self-harm in adults. Cochrane Database Syst Rev 2015 Jul 06(7):CD011777. [doi: 10.1002/14651858.CD011777] [Medline: 26147958]
14. Hawton K, Witt KG, Taylor STL, Arensman E, Gunnell D, Hazell P, et al. Psychosocial interventions for self-harm in adults. Cochrane Database Syst Rev 2016 May 12(5):CD012189. [doi: 10.1002/14651858.CD012189] [Medline: 27168519]
15. Cummins I. The Impact of Austerity on Mental Health Service Provision: A UK Perspective. Int J Environ Res Public Health 2018 Jun 01;15(6):1145-1145 [FREE Full text] [doi: 10.3390/ijerph15061145] [Medline: 29865157]
16. World HO. New horizons for health through mobile technologies. Geneva: World Health Organization; 2011.
17. Ofcom. Children and parents: Media use and attitudes report 2018. Ofcom. 2019. URL: https://www.ofcom.org.uk/__data/assets/pdf_file/0024/134907/children-and-parents-media-use-and-attitudes-2018.pdf [accessed 2020-01-27]
18. Ofcom. Adults: Media use and attitudes report 2019. Ofcom. 2019. URL: https://www.ofcom.org.uk/__data/assets/pdf_file/0021/149124/adults-media-use-and-attitudes-report.pdf [accessed 2020-03-27]
19. Barak A, Klein B, Proudfoot JG. Defining internet-supported therapeutic interventions. Ann Behav Med 2009 Aug;38(1):4-17. [doi: 10.1007/s12160-009-9130-7] [Medline: 19787305]
20. Clough BA, Casey LM. The smart therapist: A look to the future of smartphones and mHealth technologies in psychotherapy. Professional Psychology: Research and Practice 2015;46(3):147-153. [doi: 10.1037/pro0000011]
21. Firth J, Torous J, Nicholas J, Carney R, Rosenbaum S, Sarris J. Can smartphone mental health interventions reduce symptoms of anxiety? A meta-analysis of randomized controlled trials. J Affect Disord 2017 Aug 15;218:15-22 [FREE Full text] [doi: 10.1016/j.jad.2017.04.046] [Medline: 28456072]

22. Firth J, Torous J. Smartphone Apps for Schizophrenia: A Systematic Review. JMIR Mhealth Uhealth 2015;3(4):e102 [FREE Full text] [doi: 10.2196/mhealth.4930] [Medline: 26546039]
23. Arean PA, Hallgren KA, Jordan JT, Gazzaley A, Atkins DC, Heagerty PJ, et al. The Use and Effectiveness of Mobile Apps for Depression: Results From a Fully Remote Clinical Trial. J Med Internet Res 2016 Dec 20;18(12):e330 [FREE Full text] [doi: 10.2196/jmir.6482] [Medline: 27998876]
24. Prada P, Zamberg I, Bouillault G, Jimenez N, Zimmermann J, Hasler R, et al. EMOTEO: A Smartphone Application for Monitoring and Reducing Aversive Tension in Borderline Personality Disorder Patients, a Pilot Study. Perspect Psychiatr Care 2017 Oct;53(4):289-298. [doi: 10.1111/ppc.12178] [Medline: 27439663]
25. Witt K, Spittal MJ, Carter G, Pirkis J, Hetrick S, Currier D, et al. Effectiveness of online and mobile telephone applications ('apps') for the self-management of suicidal ideation and self-harm: a systematic review and meta-analysis. BMC Psychiatry 2017 Aug 15;17(1):297 [FREE Full text] [doi: 10.1186/s12888-017-1458-0] [Medline: 28810841]
26. Melia R, Francis K, Hickey E, Bogue J, Duggan J, O'Sullivan M, et al. Mobile Health Technology Interventions for Suicide Prevention: Systematic Review. JMIR Mhealth Uhealth 2020 Jan 15;8(1):e12516. [doi: 10.2196/12516]
27. Arshad U, Farhat-Ul-Ain, Gauntlett J, Husain N, Chaudhry N, Taylor PJ. A Systematic Review of the Evidence Supporting Mobile- and Internet-Based Psychological Interventions For Self-Harm. Suicide Life Threat Behav 2020 Feb;50(1):151-179 [FREE Full text] [doi: 10.1111/sltb.12583] [Medline: 31448847]
28. Cliffe B. What mHealth interventions are available for individuals who are struggling with self-harm? Open Science Framework. 2020. URL: https://osf.io/ay5pf/
29. Cochrane. Preventing relapse of major depressive disorder in youth: randomised Controlled Trial of a novel mindfulness-based cognitive online social therapy. Cochrane. 2019. URL: https://www.cochranelibrary.com/central/doi/10.1002/central/CN-02064859/full [accessed 2020-04-23]
30. Berrouiguet S, Alavi Z, Vaiva G, Courtet P, Baca-García E, Vidailhet P, et al. SIAM (Suicide intervention assisted by messages): the development of a post-acute crisis text messaging outreach for suicide prevention. BMC Psychiatry 2014;14:294 [FREE Full text] [doi: 10.1186/s12888-014-0294-8] [Medline: 25404215]
31. Bush NE, Dobscha SK, Crumpton R, Denneson LM, Hoffman JE, Crain A, et al. A Virtual Hope Box smartphone app as an accessory to therapy: proof-of-concept in a clinical sample of veterans. Suicide Life Threat Behav 2015 Feb;45(1):1-9. [doi: 10.1111/sltb.12103] [Medline: 24828126]
32. Bush NE, Smolenski DJ, Denneson LM, Williams HB, Thomas EK, Dobscha SK. A Virtual Hope Box: Randomized Controlled Trial of a Smartphone App for Emotional Regulation and Coping With Distress. Psychiatr Serv 2017 Apr 01;68(4):330-336. [doi: 10.1176/appi.ps.201600283] [Medline: 27842473]
33. Chen RY, Feltes JR, Tzeng WS, Lu ZY, Pan M, Zhao N, et al. Phone-Based Interventions in Adolescent Psychiatry: A Perspective and Proof of Concept Pilot Study With a Focus on Depression and Autism. JMIR Res Protoc 2017 Jun 16;6(6):e114 [FREE Full text] [doi: 10.2196/resprot.7245] [Medline: 28623183]
34. Di Simplicio M, Appiah-Kusi E, Wilkinson P, Watson P, Meiser-Stedman C, Kavanagh DJ, et al. Imaginator: A Proof-of-Concept Feasibility Trial of a Brief Imagery-Based Psychological Intervention for Young People Who Self-Harm. Suicide Life Threat Behav 2020 Jun;50(3):724-740. [doi: 10.1111/sltb.12620] [Medline: 32057131]
35. Franklin JC, Fox KR, Franklin CR, Kleiman EM, Ribeiro JD, Jaroszewski AC, et al. A brief mobile app reduces nonsuicidal and suicidal self-injury: Evidence from three randomized controlled trials. J Consult Clin Psychol 2016 Jun;84(6):544-557. [doi: 10.1037/ccp0000093] [Medline: 27018530]
36. Grist R, Porter J, Stallard P. Acceptability, Use, and Safety of a Mobile Phone App (BlueIce) for Young People Who Self-Harm: Qualitative Study of Service Users' Experience. JMIR Ment Health 2018 Feb 23;5(1):e16 [FREE Full text] [doi: 10.2196/mental.8779] [Medline: 29475823]
37. Stallard P, Porter J, Grist R. A Smartphone App (BlueIce) for Young People Who Self-Harm: Open Phase 1 Pre-Post Trial. JMIR Mhealth Uhealth 2018 Jan 30;6(1):e32. [doi: 10.2196/mhealth.8917]
38. Hetrick SE, Robinson J, Burge E, Blandon R, Mobilio B, Rice SM, et al. Youth Codesign of a Mobile Phone App to Facilitate Self-Monitoring and Management of Mood Symptoms in Young People With Major Depression, Suicidal Ideation, and Self-Harm. JMIR Ment Health 2018 Jan 23;5(1):e9 [FREE Full text] [doi: 10.2196/mental.9041] [Medline: 29362208]
39. Cochrane. Feasibility trial of ClinTouch-CareLoop Enhanced Management. Cochrane. 2014. URL: https://www.cochranelibrary.com/central/doi/10.1002/central/CN-01857649/full [accessed 2020-04-23]
40. Kodama T, Syouji H, Takaki S, Fujimoto H, Ishikawa S, Fukutake M, et al. Text Messaging for Psychiatric Outpatients: Effect on Help-Seeking and Self-Harming Behaviors. J Psychosoc Nurs Ment Health Serv 2016 Apr;54(4):31-37. [doi: 10.3928/02793695-20160121-01] [Medline: 27042926]
41. Mackie C, Dunn N, MacLean S, Testa V, Heisel M, Hatcher S. A qualitative study of a blended therapy using problem solving therapy with a customised smartphone app in men who present to hospital with intentional self-harm. Evid Based Ment Health 2017 Nov;20(4):118-122. [doi: 10.1136/eb-2017-102764] [Medline: 29030503]
42. Madsen LB, Eddleston M, Hansen KS, Pearson M, Agampodi S, Jayamanne S, et al. Cost-effectiveness analyses of self-harm strategies aimed at reducing the mortality of pesticide self-poisonings in Sri Lanka: a study protocol. BMJ Open 2015 Feb 27;5(2):e007333 [FREE Full text] [doi: 10.1136/bmjopen-2014-007333] [Medline: 25724984]

43. Marasinghe RB, Edirippulige S, Kavanagh D, Smith A, Jiffry MTM. Effect of mobile phone-based psychotherapy in suicide prevention: a randomized controlled trial in Sri Lanka. J Telemed Telecare 2012 Apr;18(3):151-155. [doi: 10.1258/jtt.2012.SFT107] [Medline: 22362830]
44. Cochrane. Text Message Intervention to Reduce Repeat Self-harm. Cochrane. 2013. URL: https://www.cochranelibrary.com/central/doi/10.1002/central/CN-01541374/full [accessed 2020-04-23]
45. Cochrane. Effects of Community-based Caring Contact on Post-discharge Young Adults With Self-harm. Cochrane. 2017. URL: https://www.cochranelibrary.com/central/doi/10.1002/central/CN-01562629/full [accessed 2020-04-23]
46. Cochrane. The BEACON Study: smartphone-Assisted Problem-Solving Therapy in Men Presenting to the ED With Self-Harm (Protocol A). Cochrane. 2018. URL: https://www.cochranelibrary.com/central/doi/10.1002/central/CN-01567843/full [accessed 2020-04-23]
47. Owens C, Charles N. Implementation of a text-messaging intervention for adolescents who self-harm (TeenTEXT): a feasibility study using normalisation process theory. Child Adolesc Psychiatry Ment Health 2016;10:14 [FREE Full text] [doi: 10.1186/s13034-016-0101-z] [Medline: 27354855]
48. Rizvi SL, Hughes CD, Thomas MC. The DBT Coach mobile application as an adjunct to treatment for suicidal and self-injuring individuals with borderline personality disorder: A preliminary evaluation and challenges to client utilization. Psychol Serv 2016 Dec;13(4):380-388. [doi: 10.1037/ser0000100] [Medline: 27797571]
49. Sayal K, Roe J, Ball H, Atha C, Kaylor-Hughes C, Guo B, et al. Feasibility of a randomised controlled trial of remotely delivered problem-solving cognitive behaviour therapy versus usual care for young people with depression and repeat self-harm: lessons learnt (e-DASH). BMC Psychiatry 2019 Jan 24;19(1):42. [doi: 10.1186/s12888-018-2005-3] [Medline: 30678674]
50. Andreasson K, Krogh J, Bech P, Frandsen H, Buus N, Stanley B, et al. MYPLAN -mobile phone application to manage crisis of persons at risk of suicide: study protocol for a randomized controlled trial. Trials 2017 Dec 11;18(1):171 [FREE Full text] [doi: 10.1186/s13063-017-1876-9] [Medline: 28399909]
51. Skovgaard LJL, Frandsen H, Erlangsen A. MYPLAN - A Mobile Phone Application for Supporting People at Risk of Suicide. Crisis 2016 May;37(3):236-240. [doi: 10.1027/0227-5910/a000371] [Medline: 26831213]
52. Stevens GJ, Hammond TE, Brownhill S, Anand M, de la Riva A, Hawkins J, et al. SMS SOS: a randomized controlled trial to reduce self-harm and suicide attempts using SMS text messaging. BMC Psychiatry 2019 Apr 18;19(1):117 [FREE Full text] [doi: 10.1186/s12888-019-2104-9] [Medline: 30999952]
53. Szlyk HS, Roth KB, García-Perdomo V. Engagement With Crisis Text Line Among Subgroups of Users Who Reported Suicidality. Psychiatr Serv 2020 Apr 01;71(4):319-327. [doi: 10.1176/appi.ps.201900149] [Medline: 31795857]
54. Werner-Seidler A, Huckvale K, Larsen ME, Calear AL, Maston K, Johnston L, et al. A trial protocol for the effectiveness of digital interventions for preventing depression in adolescents: The Future Proofing Study. Trials 2020 Jan 02;21(1):2 [FREE Full text] [doi: 10.1186/s13063-019-3901-7] [Medline: 31898512]
55. Bjureberg J, Sahlin H, Hedman-Lagerlöf E, Gratz KL, Tull MT, Jokinen J, et al. Extending research on Emotion Regulation Individual Therapy for Adolescents (ERITA) with nonsuicidal self-injury disorder: open pilot trial and mediation analysis of a novel online version. BMC Psychiatry 2018 Oct 11;18(1):326 [FREE Full text] [doi: 10.1186/s12888-018-1885-6] [Medline: 30305103]
56. Van Voorhees BW, Fogel J, Reinecke MA, Gladstone T, Stuart S, Gollan J, et al. Randomized clinical trial of an Internet-based depression prevention program for adolescents (Project CATCH-IT) in primary care: 12-week outcomes. J Dev Behav Pediatr 2009 Feb;30(1):23-37 [FREE Full text] [doi: 10.1097/DBP.0b013e3181966c2a] [Medline: 19194326]
57. Evans MO, Morgan HG, Hayward A, Gunnell DJ. Crisis telephone consultation for deliberate self-harm patients: effects on repetition. Br J Psychiatry 1999 Jul;175:23-27. [doi: 10.1192/bjp.175.1.23] [Medline: 10621764]
58. Eylem O, van SA, Bhui K, Kerkhof AJFM. Protocol: Reducing suicidal ideation among Turkish migrants in the Netherlands and in the UK: effectiveness of an online intervention. Int Rev Psychiatry 2015 Feb;27(1):72-81. [doi: 10.3109/09540261.2014.996121] [Medline: 25747026]
59. van Spijker BAJ, van Straten A, Kerkhof AJFM. The effectiveness of a web-based self-help intervention to reduce suicidal thoughts: a randomized controlled trial. Trials 2010;11:25 [FREE Full text] [doi: 10.1186/1745-6215-11-25] [Medline: 20214777]
60. Hooley JM, Fox KR, Wang SB, Kwashie AND. Novel online daily diary interventions for nonsuicidal self-injury: a randomized controlled trial. BMC Psychiatry 2018 Aug 22;18(1):264 [FREE Full text] [doi: 10.1186/s12888-018-1840-6] [Medline: 30134866]
61. Lederer N, Grechenig T, Baranyi R. unCUT: bridging the gap from paper diary cards towards mobile electronic monitoring solutions in borderline and self-injury. : SeGAH; 2014 Presented at: 2014 IEEE 3nd International Conference on Serious Games and Applications for Health (SeGAH); 14-16 May 2014; Brazil. [doi: 10.1109/segah.2014.7067092]
62. Mühlmann C, Madsen T, Hjorthøj C, Kerkhof A, Nordentoft M, Erlangsen A. The Self-help Online against Suicidal thoughts (SOS) trial: study protocol for a randomized controlled trial. Trials 2017 Jan 28;18(1):45 [FREE Full text] [doi: 10.1186/s13063-017-1794-x] [Medline: 28129783]

63. van Spijker BAJ, van Straten A, Kerkhof AJFM. Effectiveness of online self-help for suicidal thoughts: results of a randomised controlled trial. PLoS One 2014;9(2):e90118 [FREE Full text] [doi: 10.1371/journal.pone.0090118] [Medline: 24587233]
64. McManama OKH, LeCloux M, Ross A, Gironda C, Wharff EA. A Pilot Study of the Acceptability and Usability of a Smartphone Application Intervention for Suicidal Adolescents and Their Parents. Arch Suicide Res 2016 May 2:1-11. [doi: 10.1080/13811118.2016.1182094] [Medline: 27136115]
65. Helweg-Jørgensen S, Beck Lichtenstein M, Fruzzetti AE, Møller Dahl C, Pedersen SS. Daily Self-Monitoring of Symptoms and Skills Learning in Patients With Borderline Personality Disorder Through a Mobile Phone App: Protocol for a Pragmatic Randomized Controlled Trial. JMIR Res Protoc 2020 May 25;9(5):e17737 [FREE Full text] [doi: 10.2196/17737] [Medline: 32449690]
66. Duan S, Wang H, Wilson A, Qiu J, Chen G, He Y, et al. Developing a Text Messaging Intervention to Reduce Deliberate Self-Harm in Chinese Adolescents: Qualitative Study. JMIR Mhealth Uhealth 2020 Jun 11;8(6):e16963 [FREE Full text] [doi: 10.2196/16963] [Medline: 32392173]
67. Kerner B, Carlson M, Eskin CK, Tseng C, Ho JG, Zima B, et al. Trends in the utilization of a peer-supported youth hotline. Child Adolesc Ment Health 2020 May 25:65-72. [doi: 10.1111/camh.12394] [Medline: 32449589]
68. Al-Sharifi A, Krynicki CR, Upthegrove R. Self-harm and ethnicity: A systematic review. Int J Soc Psychiatry 2015 Sep;61(6):600-612. [doi: 10.1177/0020764015573085] [Medline: 25783961]
69. Rimes K, Goodship N, Ussher G, Baker D, West E. Non-binary and binary transgender youth: Comparison of mental health, self-harm, suicidality, substance use and victimization experiences. Int J Transgenderism. (2-3) 2019;20:230-240. [doi: 10.4324/9781003015888-11] [Medline: 32999609]
70. Glenn CR, Klonsky ED. One-year test-retest reliability of the Inventory of Statements about Self-Injury (ISAS). Assessment 2011 Sep;18(3):375-378. [doi: 10.1177/1073191111411669] [Medline: 21665881]
71. Ofcom. Communications Market Report. Ofcom. 2018. URL: https://www.ofcom.org.uk/__data/assets/pdf_file/0022/117256/CMR-2018-narrative-report.pdf [accessed 2020-08-11]
72. Hunkin H, King DL, Zajac IT. Perceived acceptability of wearable devices for the treatment of mental health problems. J Clin Psychol 2020 Jun;76(6):987-1003. [doi: 10.1002/jclp.22934] [Medline: 32022908]
73. Gartner. Gartner Says Worldwide Wearable Device Sales to Grow 26 Percent in 2019. Gartner. 2018 Nov 29. URL: https://www.gartner.com/en/newsroom/press-releases/2018-11-29-gartner-says-worldwide-wearable-device-sales-to-grow-#:~:text=Gartner%2C%20Inc.,billion%20will%20be%20on%20smartwatches [accessed 2020-08-11]
74. Nicholas J, Larsen ME, Proudfoot J, Christensen H. Mobile Apps for Bipolar Disorder: A Systematic Review of Features and Content Quality. J Med Internet Res 2015;17(8):e198 [FREE Full text] [doi: 10.2196/jmir.4581] [Medline: 26283290]
75. Willig C. Introducing qualitative research in psychology: Adventures in theory and method. Buckingham & Philadelphia: Open University Press; 2008.
76. Mason J. Qualitative Researching. In: Qualitative Researching. London: SAGE Publications; 2012.
77. Owens C, Fox F, Redwood S, Davies R, Foote L, Salisbury N, et al. Measuring outcomes in trials of interventions for people who self-harm: qualitative study of service users' views. BJPsych Open 2020 Feb 12;6(2):e22 [FREE Full text] [doi: 10.1192/bjo.2019.93] [Medline: 32046813]
78. Cliffe B, Croker A, Denne M, Stallard P. Clinicians' use of and attitudes towards technology to provide and support interventions in child and adolescent mental health services. Child Adolesc Ment Health 2020 May;25(2):95-101. [doi: 10.1111/camh.12362] [Medline: 32307835]
79. Hollis C, Falconer CJ, Martin JL, Whittington C, Stockton S, Glazebrook C, et al. Annual Research Review: Digital health interventions for children and young people with mental health problems: a systematic and meta-review. J Child Psychol Psychiatry 2016 Dec 10:474-503. [doi: 10.1111/jcpp.12663] [Medline: 27943285]
80. Aguilera A. Digital Technology and Mental Health Interventions: Opportunities and Challenges. Arbor 2015 May 12;191(771):a210. [doi: 10.3989/arbor.2015.771n1012]
81. Greaves F, Joshi I, Campbell M, Roberts S, Patel N, Powell J. What is an appropriate level of evidence for a digital health intervention? Lancet 2019 Dec 22;392(10165):2665-2667. [doi: 10.1016/S0140-6736(18)33129-5] [Medline: 30545779]
82. Pagoto S, Bennett GG. How behavioral science can advance digital health. Transl Behav Med 2013 Sep;3(3):271-276 [FREE Full text] [doi: 10.1007/s13142-013-0234-z] [Medline: 24073178]
83. Morrison K. Randomised Controlled Trials for Evidence-based Education: Some Problems in Judging 'What Works'. Evaluation & Research in Education 2001 Oct;15(2):69-83. [doi: 10.1080/09500790108666984]
84. Raghavan R, Munson MR, Le C. Toward an Experimental Therapeutics Approach in Human Services Research. Psychiatr Serv 2019 Dec 01;70(12):1130-1137 [FREE Full text] [doi: 10.1176/appi.ps.201800577] [Medline: 31500543]

Abbreviations

mHealth: mobile health
RCT: randomized controlled trial

Edited by R Kukafka; submitted 19.10.20; peer-reviewed by E Bailey, E Kleiman, B Zima; comments to author 05.12.20; revised version received 19.01.21; accepted 16.03.21; published 30.04.21.

Please cite as:
Cliffe B, Tingley J, Greenhalgh I, Stallard P
mHealth Interventions for Self-Harm: Scoping Review
J Med Internet Res 2021;23(4):e25140
URL: https://www.jmir.org/2021/4/e25140
doi:10.2196/25140
PMID:33929329

Review

Long-term Effectiveness of mHealth Physical Activity Interventions: Systematic Review and Meta-analysis of Randomized Controlled Trials

Annette Mönninghoff[1,2], MA; Jan Niklas Kramer[3,4], PhD; Alexander Jan Hess[3,5], MSc; Kamila Ismailova[3], MSc; Gisbert W Teepe[6], MSc; Lorainne Tudor Car[7,8], MD; Falk Müller-Riemenschneider[9], MD; Tobias Kowatsch[3,6,9,10], PhD

[1]Institute for Customer Insight, University of St. Gallen, St. Gallen, Switzerland

[2]Institute for Mobility, University of St. Gallen, St. Gallen, Switzerland

[3]Centre for Digital Health Interventions, Institute of Technology Management, University of St. Gallen, St. Gallen, Switzerland

[4]CSS Insurance, Lucerne, Switzerland

[5]Translational Neuromodeling Unit, Institute for Biomedical Engineering, University of Zurich and ETH Zurich, Zurich, Switzerland

[6]Centre for Digital Health Interventions, Department of Management, Technology, and Economics, ETH Zurich, Zurich, Switzerland

[7]Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore, Singapore

[8]School of Public Health, Department of Primary Care and Public Health, Imperial College London, London, United Kingdom

[9]Saw Swee Hock School of Public Health, National University of Singapore, Singapore, Singapore

[10]Future Health Technologies Programme, Campus for Research Excellence and Technological Enterprise (CREATE), Singapore-ETH Centre, Singapore, Singapore

Corresponding Author:
Annette Mönninghoff, MA
Institute for Customer Insight
University of St. Gallen
Bahnhofstrasse 8
St. Gallen, 9000
Switzerland
Phone: 41 76 229 3150
Email: Annette.Moenninghoff@unisg.ch

Abstract

Background: Mobile health (mHealth) interventions can increase physical activity (PA); however, their long-term impact is not well understood.

Objective: The primary aim of this study is to understand the immediate and long-term effects of mHealth interventions on PA. The secondary aim is to explore potential effect moderators.

Methods: We performed this study according to the Cochrane and PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines. We searched PubMed, the Cochrane Library, SCOPUS, and PsycINFO in July 2020. Eligible studies included randomized controlled trials of mHealth interventions targeting PA as a primary outcome in adults. Eligible outcome measures were walking, moderate-to-vigorous physical activity (MVPA), total physical activity (TPA), and energy expenditure. Where reported, we extracted data for 3 time points (ie, end of intervention, follow-up ≤6 months, and follow-up >6 months). To explore effect moderators, we performed subgroup analyses by population, intervention design, and control group type. Results were summarized using random effects meta-analysis. Risk of bias was assessed using the Cochrane Collaboration tool.

Results: Of the 2828 identified studies, 117 were included. These studies reported on 21,118 participants with a mean age of 52.03 (SD 14.14) years, of whom 58.99% (n=12,459) were female. mHealth interventions significantly increased PA across all the 4 outcome measures at the end of intervention (walking standardized mean difference [SMD] 0.46, 95% CI 0.36-0.55; *P*<.001; MVPA SMD 0.28, 95% CI 0.21-0.35; *P*<.001; TPA SMD 0.34, 95% CI 0.20-0.47; *P*<.001; energy expenditure SMD 0.44, 95% CI 0.13-0.75; *P*=.01). Only 33 studies reported short-term follow-up measurements, and 8 studies reported long-term follow-up measurements in addition to end-of-intervention results. In the short term, effects were sustained for walking (SMD 0.26, 95%

CI 0.09-0.42; *P*=.002), MVPA (SMD 0.20, 95% CI 0.05-0.35; *P*=.008), and TPA (SMD 0.53, 95% CI 0.13-0.93; *P*=.009). In the long term, effects were also sustained for walking (SMD 0.25, 95% CI 0.10-0.39; *P*=.001) and MVPA (SMD 0.19, 95% CI 0.11-0.27; *P*<.001). We found the study population to be an effect moderator, with higher effect scores in sick and at-risk populations. PA was increased both in scalable and nonscalable mHealth intervention designs and regardless of the control group type. The risk of bias was rated high in 80.3% (94/117) of the studies. Heterogeneity was significant, resulting in low to very low quality of evidence.

Conclusions: mHealth interventions can foster small to moderate increases in PA. The effects are maintained long term; however, the effect size decreases over time. The results encourage using mHealth interventions in at-risk and sick populations and support the use of scalable mHealth intervention designs to affordably reach large populations. However, given the low evidence quality, further methodologically rigorous studies are warranted to evaluate the long-term effects.

(J Med Internet Res 2021;23(4):e26699) doi:10.2196/26699

KEYWORDS

mHealth; physical activity; systematic review; meta-analysis; mobile phone

Introduction

Background

In recent decades, populations have become increasingly sedentary. The World Health Organization (WHO) recommends 150 minutes of moderate-intensity physical activity (PA) or 75 minutes of vigorous-intensity PA per week for adults and 60 minutes of moderate-to-vigorous physical activity (MVPA) for adolescents per day [1]. An estimated 28% of adults worldwide do not meet these guidelines [2]. The prevalence of inactivity is high in Latin America and many high-income countries, with approximately every second adult inactive in Brazil or Saudi Arabia, and 40% of adults insufficiently active in the United States [2].

According to the WHO, physical inactivity is 1 of the 4 core modifiable risk factors for noncommunicable diseases (NCDs). As such, it is as important to be addressed as tobacco use or obesity and proven to increase the risk of cancer, cardiovascular diseases, diabetes, dementia, and depression [3-6].

In response to the high prevalence and substantial risk posed by physical inactivity, the WHO has formulated a target to reduce physical inactivity by 10% by 2025 as part of its strategy against NCDs [7]. Scaling up PA interventions is key to achieving the WHO target. However, there are various barriers, including cost, resource restrictions, and poorly scalable intervention designs [8,9]. Owing to the increasing dissemination and ubiquity of mobile technology, mobile technology–based interventions, that is, mobile health (mHealth), have been discussed as a solution for overcoming scalability challenges [10,11]. There are only a few examples of nationwide mHealth programs such as the *NHS Diabetes Prevention Program* [12] in the United Kingdom, the 10,000 steps program in Australia [13], and the *National Steps Challenge* and *Live Healthy SG* in Singapore [14,15]. Most governments and health organizations are still hesitant about rolling out mHealth PA programs, as clear evidence for the effectiveness of mHealth interventions for sustainable behavior change is lacking [16,17].

Previous evidence for the effectiveness of mHealth interventions on PA is mixed (Multimedia Appendix 1 [18-31]). Most existing meta-analyses found significant positive effects on PA in sick and at-risk populations, with effect sizes ranging from small to large [18-28]. However, some studies did not find significant effects or reported conflicting results [29-31]. There is limited evidence for the sustainability of increased PA levels beyond the end of intervention. Only 2 studies quantitatively analyzed long-term effects: one review found that PA increases are maintained up to 3-4 years after the intervention [20] and the other did not find significant long-term results [31]. Kirk et al [25] and Romeo et al [30] found that shorter mHealth PA interventions (<16 weeks and <12 weeks, respectively) are more effective than longer ones, indicating that effects might not be maintained in the long term.

We also lack clarity on how population types, intervention design, and control group type moderate the impact of mHealth interventions on PA. Only 3 studies performed subgroup analyses according to population type with mixed results. A total of 2 studies found interventions to be equally effective in sick and healthy populations [23,30], and 1 review found mHealth interventions to be more effective in sick populations; however, the results were not statistically significant [27]. Most studies focused exclusively on sick or at-risk populations [21,22,24-26,28,31], making it difficult to draw clear conclusions.

The design of mHealth interventions influences the degree to which they are scalable. The promise of mHealth is that the technology itself (ie, without costly and limited human resources) promotes active lifestyles. However, these highly scalable interventions miss the element of human-to-human interaction, which is a potentially important active ingredient in behavior change interventions. Current evidence draws an inconclusive picture: existing studies have found no effects on PA when interventions are scalable [24,30] (ie, mHealth interventions without human-to-human interactions), stronger effects when interventions are nonscalable (ie, mHealth interventions with human-to-human interactions) [27,32], stronger effects in scalable interventions [20], or no moderating effects [22]. Thus, we need a comprehensive evaluation of scalable versus nonscalable designs to judge the potential of mHealth technologies in reaching large populations at low costs.

Furthermore, our current understanding of the effects of mHealth PA interventions is limited by the inclusion of different control groups in previous studies. Most previous studies included both

minimal or no intervention control groups and control groups receiving an alternative intervention [18-20,22-29,31]. This makes it impossible to distinguish between the absolute effect of mHealth PA interventions on behavior and the degree to which mHealth interventions are superior to alternative nonmobile designs or the standard of care.

Objectives

Accordingly, we sought to comprehensively collate and analyze trials evaluating mHealth interventions that promote PA in adult populations. Our primary aim is to understand the long-term impact of these interventions on PA. Our secondary aim is to explore potential effect moderators (population type, intervention design, and control group type), to understand which populations can benefit from mHealth interventions, to understand if scalable mHealth intervention designs are effective, and to understand if mHealth interventions produce superior results to nonmobile interventions.

Methods

Overview

This study was performed according to the Cochrane methodology, and the results were reported following the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines. We searched PubMed, the Cochrane Library, SCOPUS, and PsycINFO for randomized controlled trials (RCTs) on mHealth interventions targeting PA increases (all search strategies are given in Multimedia Appendices 2 and 3) published from database inception to July 3, 2020. We also searched the reference lists of the relevant existing systematic reviews for eligible studies. This study was registered with PROSPERO (CRD 42019124716).

Eligibility Criteria

Studies were eligible if they assessed the impact of mHealth interventions on PA as a primary study outcome in individuals aged 18 years or more and were published in English. Eligible study designs were RCTs or cluster RCTs. Eligible comparators comprised no or minimal interventions and alternative interventions that did not include mobile technologies.

Types of Interventions

mHealth interventions were defined as programs that fully or partly deliver interventions using mobile technology such as pedometers or accelerometers with displays, activity trackers, smartphones, or tablets. We excluded interventions where the use of a mobile device was unclear (eg, telephone or website interventions) or where increasing PA was not the primary outcome. This was to ensure that interventions genuinely aimed to increase PA and to avoid including studies measuring PA only as a supplemental outcome.

Types of Outcomes

Eligible outcome measures were walking, MVPA, total physical activity (TPA), and energy expenditure (EE), as these outcomes are most commonly reported. Multiple outcome units were eligible per outcome measure (eg, walking in minutes and walking in steps). Studies reporting objectively measured or self-reported outcome data were eligible.

Data Collection Process

Abstracts of all identified papers were exported and uploaded into Covidence Systematic Review software (Veritas Health Innovation Ltd, version accessed July 2020) for screening. Two reviewers independently screened the abstracts for eligibility (AM, JNK, or KI). If reviewers doubted whether an article was potentially relevant, it was included for full inspection. Next, full texts of potentially eligible papers were uploaded into Covidence and screened by 2 independent reviewers (AM, JNK, or KI). Conflicts were resolved by discussion or where required by a third reviewer. We contacted authors of potentially relevant articles for further information when needed. All reviewers were trained during a full-day workshop on eligibility criteria and software before screening.

Data Extraction and Management

Data for each study were extracted independently by 2 reviewers (AM, JNK, KI, AJH, or GWT) using standardized extraction forms. Conflicts were resolved by discussion between 2 primary reviewers or with a third, independent reviewer. All reviewers were trained to use the extraction form and Cochrane risk of bias criteria during a full-day workshop. Where reported, data were extracted for all 4 eligible outcome measures (walking, MVPA, TPA, and EE) and time points (end of intervention, short-term follow-up [≤6 months after the end of intervention], and long-term follow-up [>6 months after the end of intervention]). If studies reported both objectively measured and self-reported outcome data, the former were used for meta-analysis. If studies only reported self-reported outcome data, these were extracted, and the quality of evidence was rated as high risk of detection bias. Data were extracted as means and SDs per outcome measure and time point. If SDs were not reported, they were calculated using the RevMan calculator and following the Cochrane Handbook [33]. Respective authors were contacted for any missing data.

Assessment of Risk of Bias in Included Studies

Two reviewers (AM, JNK, KI, AJH, or GWT) independently assessed the risk of bias for each study using the Cochrane Collaboration tool [34]. Additional criteria for cluster RCTs were assessed [35] and documented within the *other bias* domains of the Cochrane Collaboration tool. Discrepancies were resolved by consensus between reviewers or where needed by a third reviewer. We classified studies as overall high risk of bias if they scored high in any bias domain other than *performance bias*, as blinding of participants and personnel is almost impossible in mHealth intervention studies [23]. Blinding of outcome assessors was rated as high risk if outcomes were self-reported.

Statistical Methods

We summarized the intervention and sample characteristics of all the included studies. We quantitatively analyzed the data using RevMan software (Cochrane, version 5.4) and a DerSimonian and Laird random effects model for our meta-analysis [36]. We reported all 4 outcome measures using standardized mean differences (SMDs) and 95% CI. Where

appropriate (eg, if one mHealth intervention was compared with a minimal and alternative nonmobile intervention), we combined means and SDs of control or intervention groups following the Cochrane Handbook [33]. We classified populations into 3 groups based on the reported recruitment criteria: sick, at-risk, and healthy. The sick group included populations experiencing illnesses such as diabetes, cancer, chronic obstructive pulmonary disease, and coronary heart disease. The at-risk group included inactive or sedentary, older, overweight, and obese populations. We classified mHealth interventions into 2 designs: scalable and nonscalable. Scalability is defined as the ability to scale up an intervention without requiring human resources. Consequently, scalable mHealth interventions were defined as interventions that only leveraged automated components without any human-to-human interactions. Nonscalable mHealth interventions included human-to-human interactions, such as coaching, in-person feedback, or group activity sessions. We classified control groups into no or minimal interventions (no intervention or information material only) and alternative (nonmobile) interventions.

Following the recommendations of Richardson et al [37] and the Cochrane Handbook [33], a subgroup analysis was performed based on end-of-intervention values for all outcomes where at least 10 studies were available. We a priori defined 3 subgroup analyses following the population, intervention, comparison, and outcome framework [33] to identify possible effect moderators. Our aim is to understand the impact of population type (sick, at-risk, and healthy), intervention design (scalable and nonscalable), and control group type (no or minimal and alternative).

We present the primary results using forest plots for each outcome and time point. Subgroup analyses were displayed in forest plots using end-of-intervention data. We quantified inconsistencies between studies using the I^2 statistics (ie, the varying effect estimates towing to heterogeneity rather than chance) [33]. We classified I^2>50% as having substantial heterogeneity [38]. We examined the significance of heterogeneity using chi-square tests (P≤.05). We assessed subgroup differences following the guidelines given by Richardson et al [37], which recommend testing for significant subgroup differences (P≤.10) and covariate distribution and comparing heterogeneity and effect sizes between subgroups. Funnel plot analysis was used to detect sampling bias. We used end-of-intervention effect values in our funnel plot analyses, as all studies reported this time point.

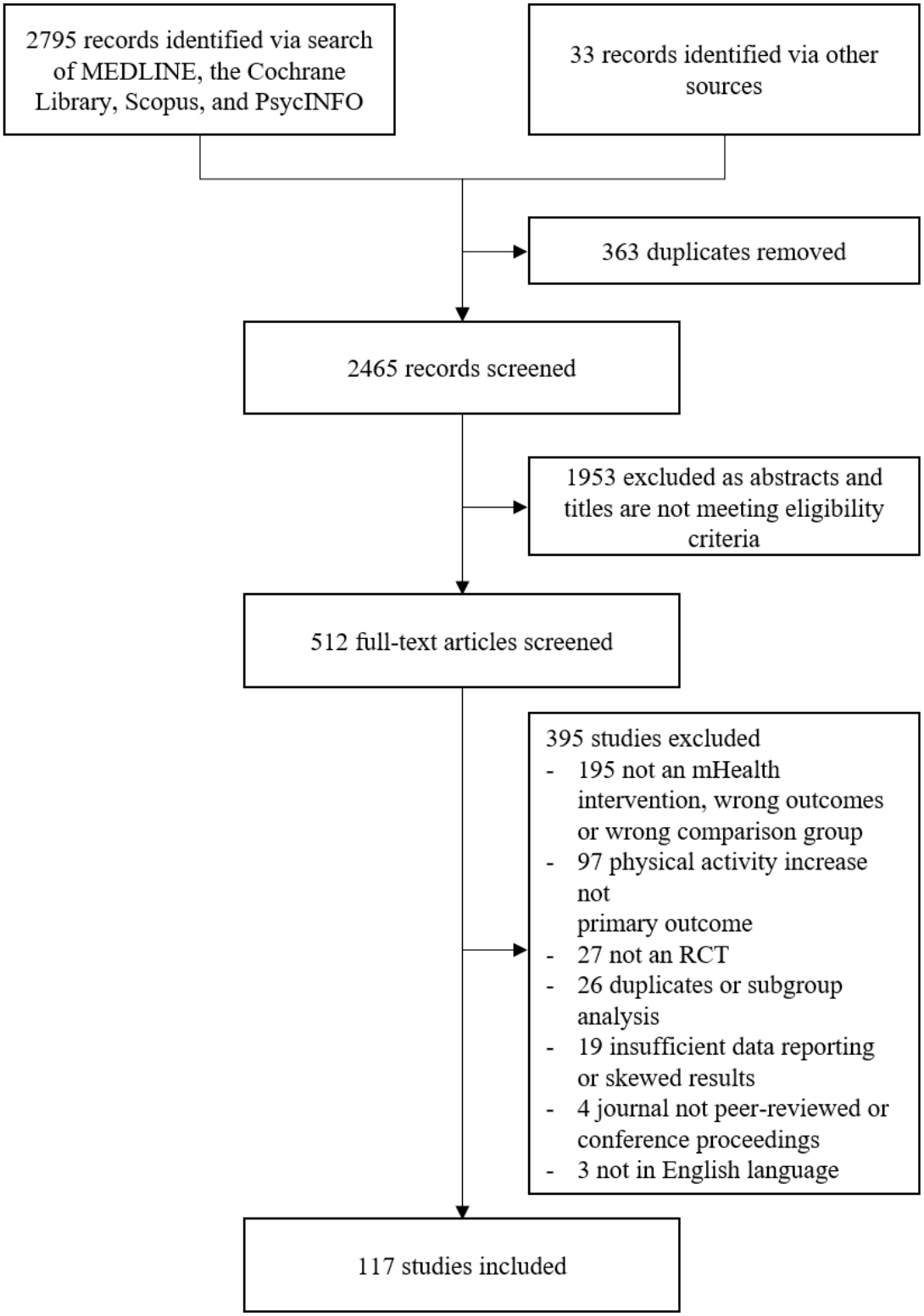
We conducted 3 sensitivity analyses to evaluate the robustness of our primary results: first, we excluded outlier studies; second, we excluded studies with high risk of bias; and third, we excluded studies not reporting long-term follow-up measurements to keep the study sample consistent across all time points. We used the grading of recommendations, assessment, development, and evaluation (GRADE) framework to assess the quality of evidence at the outcome measure level for the end-of-intervention time point and to report the standardized quality of evidence profiles, following the study by Guyatt et al [39].

Results

Overview

Of the 2828 identified studies, 512 full-text articles were screened, and 117 studies were included in the meta-analysis (Figure 1).

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) chart. mHealth: mobile health; RCT: randomized controlled trial.



Study Characteristics

Multimedia Appendix 4 [40-156] contains the included studies and their characteristics. The 117 trials represented 21,188 participants with a mean age of 52.03 years (SD 14.14), of whom 58.99% (12,459/21,118) were female. Most studies were conducted in high-income, developed regions such as North America (43/117, 36.8%), Europe (39/117, 33.3%), and Australia and New Zealand (24/117, 20.5%). Very few studies were conducted in Asia (7/117, 6.0%), Latin America (3/117, 2.6%), or Africa (1/117, 0.9%), limiting the representativeness of the evidence for low-income countries. Sample sizes ranged widely (from 15 to 1442), and the intervention duration ranged from 1 week to 2 years. All but one study [40] reported end-of-intervention results, 33 studies reported short-term follow-up results [40-72], and only 8 studies reported long-term

follow-up results [61,72-78]. The mean time point for the short-term follow-up was 4.14 months (SD 2.08) after the end of intervention. The long-term follow-up measurement was taken on average after 13.96 months (SD 11.91).

Walking was the most reported outcome measure (77/117, 65.8%) [41,43,44,48,50-54,58-60,63-65,67-69,72-130], followed by MVPA (62/117, 53.0%) [42,44,46,47,49,51-55,57,59,61-64,68,70-76,78-80,82,84-87,89-91,94,98,104,107,109,112,114,117,125,126,131-147], TPA (33/117, 28.2%) [44,45,50,52,54,56,64,66,72,74,76-78,84,85,89,96,110-112,114,118,131,135,146,148-154,157], and EE (5/117, 4.3%) [61,103,137,155,156]. Most RCTs were conducted in at-risk (48/117, 41.0%) or sick populations (46/117, 39.3%). Only 19.7% (23/117) of the studies tested mHealth interventions in healthy populations within a preventative setting. In most interventions, mHealth technologies were leveraged in nonscalable intervention designs (71/117, 60.7%). Human-to-human interactions included individual coaching, group coaching, PA classes, and physical education classes. Of the 117 interventions, 45 (38.5%) used mHealth technologies without any human-to-human interactions and were thus classified as scalable. In 1 study [75], 2 mHealth interventions (scalable and nonscalable) were combined. Most mHealth interventions only leveraged basic technologies such as pedometers or accelerometers (86/117, 73.5%), text messages (20/117, 17.1%), or websites (20/117, 17.1%). Although some recent studies pioneered innovative mHealth technologies [49,96], overall, only a few studies used advanced mHealth technologies such as automated individualized feedback (19/117, 16.2%), mobile phone apps (15/117, 12.8%), social comparison (10/117, 8.5%), and automated coaching or virtual advisors (5/117, 4.3%). Most studies had no or minimal intervention control groups (83/117, 70.9%). Only a few trials had alternative intervention control groups (22/117, 18.8%). Different control group types were combined into one control group in 12 cases.

Meta-analysis of mHealth Interventions on PA

Overall, mHealth interventions significantly increased PA across all 4 outcome measures at the end of intervention: walking SMD 0.46 (95% CI 0.36-0.55; $P<.001$; $I^2=83\%$, $P<.001$); MVPA SMD 0.28 (95% CI 0.21-0.35; $P<.001$; $I^2=62\%$, $P<.001$); TPA SMD 0.34 (95% CI 0.20-0.47; $P<.001$; $I^2=77\%$, $P<.001$); and EE SMD 0.44 (95% CI 0.13-0.75; $P=.005$; $I^2=60\%$, $P=.04$; Figures 2-9). Short-term effects were sustained (≤6 months after the end of intervention) for 3 of 4 outcome measures: walking SMD 0.26 (95% CI 0.09-0.42; $P=.002$; $I^2=73\%$, $P<.001$); MVPA SMD 0.20 (95% CI 0.05-0.35; $P=.008$; $I^2=72\%$, $P<.001$); and TPA SMD 0.53 (95% CI 0.13-0.93; $P=.009$; $I^2=87\%$, $P<.001$). Only one study [61] reported short-term follow-up measurements for EE, and the results were not statistically significant. In addition, long-term (>6 months after the end of intervention) effects were sustained for 2 of 4 outcome measures: walking SMD 0.25 (95% CI 0.10-0.39; $P=.001$; $I^2=68\%$, $P=.004$) and MVPA SMD 0.19 (95% CI 0.11-0.27; $P<.001$; $I^2=0\%$, $P=.44$). TPA results were sustained, but the effects were just below the significance threshold (SMD 0.19, 95% CI 0.00-0.38; $P=.05$; $I^2=72\%$, $P=.003$). Again, only one study [61] reported long-term follow-up effects for EE, which were not statistically significant. Effect sizes decreased over time, from almost moderate at the end of intervention to small during the long-term follow-up measurement. We found substantial and significant heterogeneity across all outcomes and most time points, with I^2 ranging from 60% to 83% for end-of-intervention measurements (Figures 2-9).

Figure 2. Primary outcome analysis for the outcome walking at timepoint end of intervention.

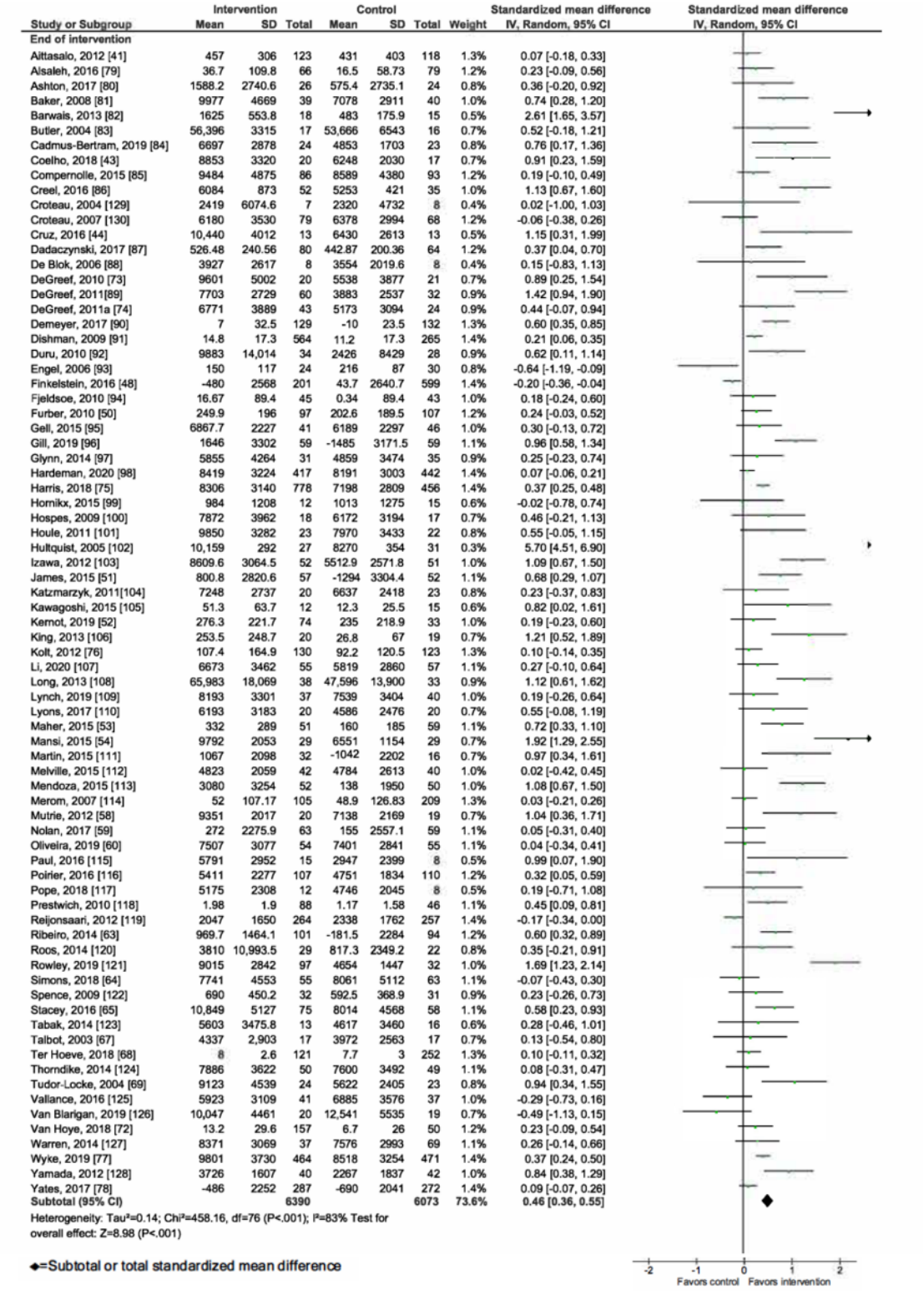


Figure 3. Primary outcome analysis for the outcome walking at timepoint short-term follow-up.

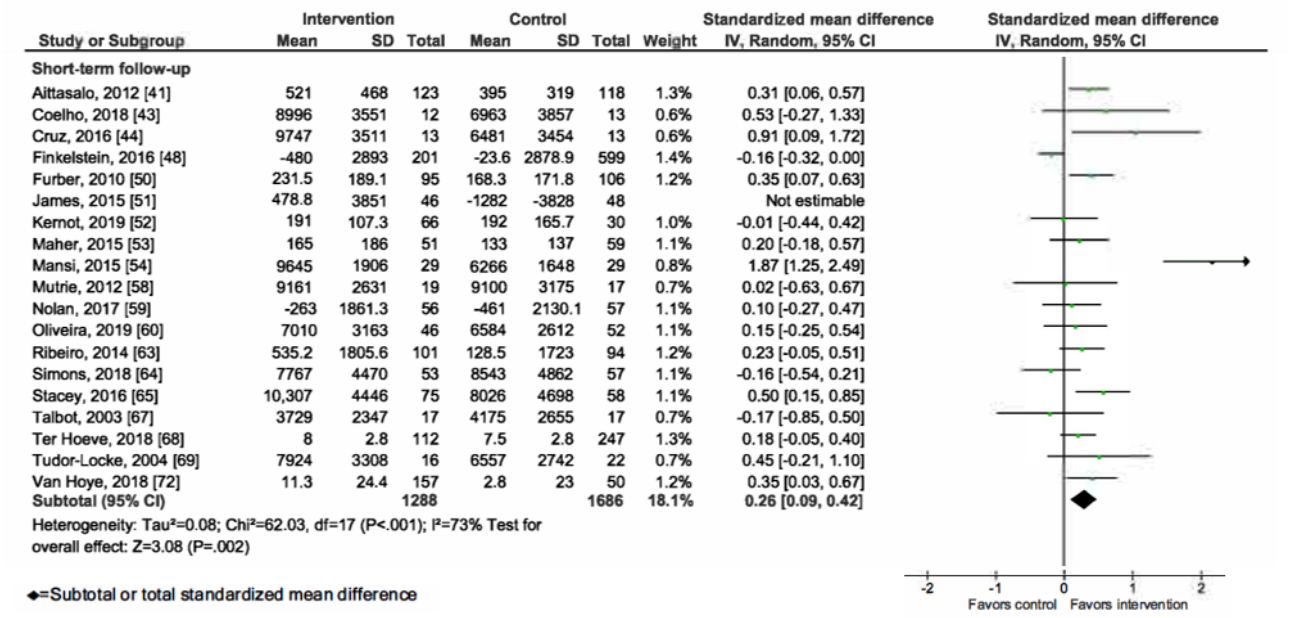


Figure 4. Primary outcome analysis for the outcome walking at timepoint long-term follow-up.

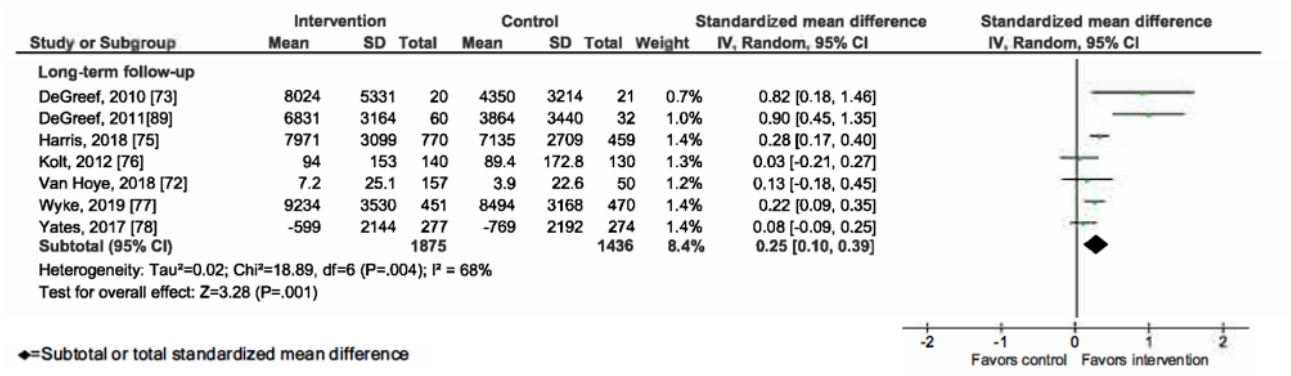
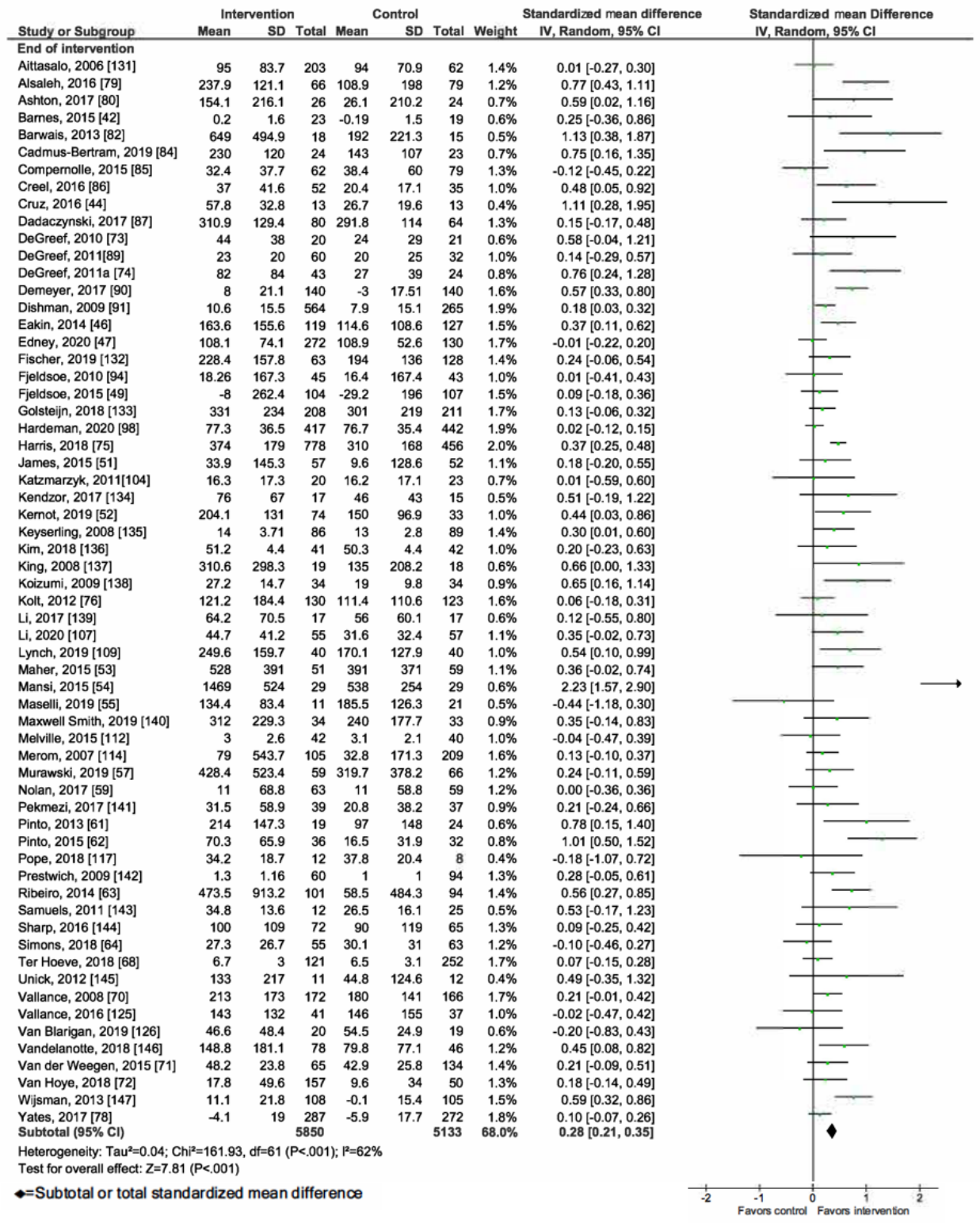


Figure 5. Primary outcome analysis for the outcome moderate-to-vigorous physical activity at timepoint end of intervention.



Study or Subgroup	Intervention Mean	SD	Total	Control Mean	SD	Total	Weight	Standardized mean difference IV, Random, 95% CI
End of intervention								
Aittasalo, 2006 [131]	95	83.7	203	94	70.9	62	1.4%	0.01 [-0.27, 0.30]
Alsaleh, 2016 [79]	237.9	121.1	66	108.9	198	79	1.2%	0.77 [0.43, 1.11]
Ashton, 2017 [80]	154.1	216.1	26	26.1	210.2	24	0.7%	0.59 [0.02, 1.16]
Barnes, 2015 [42]	0.2	1.6	23	-0.19	1.5	19	0.6%	0.25 [-0.36, 0.86]
Barwais, 2013 [82]	649	494.9	18	192	221.3	15	0.5%	1.13 [0.38, 1.87]
Cadmus-Bertram, 2019 [84]	230	120	24	143	107	23	0.7%	0.75 [0.16, 1.35]
Compernolle, 2015 [85]	32.4	37.7	62	38.4	60	79	1.3%	-0.12 [-0.45, 0.22]
Creel, 2016 [86]	37	41.6	52	20.4	17.1	35	1.0%	0.48 [0.05, 0.92]
Cruz, 2016 [44]	57.8	32.8	13	26.7	19.6	13	0.4%	1.11 [0.28, 1.95]
Dadaczynski, 2017 [87]	310.9	129.4	80	291.8	114	64	1.3%	0.15 [-0.17, 0.48]
DeGreef, 2010 [73]	44	38	20	24	29	21	0.6%	0.58 [-0.04, 1.21]
DeGreef, 2011[89]	23	20	60	20	25	32	1.0%	0.14 [-0.29, 0.57]
DeGreef, 2011a [74]	82	84	43	27	39	24	0.8%	0.76 [0.24, 1.28]
Demeyer, 2017 [90]	8	21.1	140	-3	17.51	140	1.6%	0.57 [0.33, 0.80]
Dishman, 2009 [91]	10.6	15.5	564	7.9	15.1	265	1.9%	0.18 [0.03, 0.32]
Eakin, 2014 [46]	163.6	155.6	119	114.6	108.6	127	1.5%	0.37 [0.11, 0.62]
Edney, 2020 [47]	108.1	74.1	272	108.9	52.6	130	1.7%	-0.01 [-0.22, 0.20]
Fischer, 2019 [132]	228.4	157.8	63	194	136	128	1.4%	0.24 [-0.06, 0.54]
Fjeldsoe, 2010 [94]	18.26	167.3	45	16.4	167.4	43	1.0%	0.01 [-0.41, 0.43]
Fjeldsoe, 2015 [49]	-8	262.4	104	-29.2	196	107	1.5%	0.09 [-0.18, 0.36]
Golsteijn, 2018 [133]	331	234	208	301	219	211	1.7%	0.13 [-0.06, 0.32]
Hardeman, 2020 [98]	77.3	36.5	417	76.7	35.4	442	1.9%	0.02 [-0.12, 0.15]
Harris, 2018 [75]	374	179	778	310	168	456	2.0%	0.37 [0.25, 0.48]
James, 2015 [51]	33.9	145.3	57	9.6	128.6	52	1.1%	0.18 [-0.20, 0.55]
Katzmarzyk, 2011[104]	16.3	17.3	20	16.2	17.1	23	0.7%	0.01 [-0.59, 0.60]
Kendzor, 2017 [134]	76	67	17	46	43	15	0.5%	0.51 [-0.19, 1.22]
Kernot, 2019 [52]	204.1	131	74	150	96.9	33	1.0%	0.44 [0.03, 0.86]
Keyserling, 2008 [135]	14	3.71	86	13	2.8	89	1.4%	0.30 [0.01, 0.60]
Kim, 2018 [136]	51.2	4.4	41	50.3	4.4	42	1.0%	0.20 [-0.23, 0.63]
King, 2008 [137]	310.6	298.3	19	135	208.2	18	0.6%	0.66 [0.00, 1.33]
Koizumi, 2009 [138]	27.2	14.7	34	19	9.8	34	0.9%	0.65 [0.16, 1.14]
Kolt, 2012 [76]	121.2	184.4	130	111.4	110.6	123	1.6%	0.06 [-0.18, 0.31]
Li, 2017 [139]	64.2	70.5	17	56	60.1	17	0.6%	0.12 [-0.55, 0.80]
Li, 2020 [107]	44.7	41.2	55	31.6	32.4	57	1.1%	0.35 [-0.02, 0.73]
Lynch, 2019 [109]	249.6	159.7	40	170.1	127.9	40	1.0%	0.54 [0.10, 0.99]
Maher, 2015 [53]	528	391	51	391	371	59	1.1%	0.36 [-0.02, 0.74]
Mansi, 2015 [54]	1469	524	29	538	254	29	0.6%	2.23 [1.57, 2.90]
Maselli, 2019 [55]	134.4	83.4	11	185.5	126.3	21	0.5%	-0.44 [-1.18, 0.30]
Maxwell Smith, 2019 [140]	312	229.3	34	240	177.7	33	0.9%	0.35 [-0.14, 0.83]
Melville, 2015 [112]	3	2.6	42	3.1	2.1	40	1.0%	-0.04 [-0.47, 0.39]
Merom, 2007 [114]	79	543.7	105	32.8	171.3	209	1.6%	0.13 [-0.10, 0.37]
Murawski, 2019 [57]	428.4	523.4	59	319.7	378.2	66	1.2%	0.24 [-0.11, 0.59]
Nolan, 2017 [59]	11	68.8	63	11	58.8	59	1.2%	0.00 [-0.36, 0.36]
Pekmezi, 2017 [141]	31.5	58.9	39	20.8	38.2	37	0.9%	0.21 [-0.24, 0.66]
Pinto, 2013 [61]	214	147.3	19	97	148	24	0.6%	0.78 [0.15, 1.40]
Pinto, 2015 [62]	70.3	65.9	36	16.5	31.9	32	0.8%	1.01 [0.50, 1.52]
Pope, 2018 [117]	34.2	18.7	12	37.8	20.4	8	0.4%	-0.18 [-1.07, 0.72]
Prestwich, 2009 [142]	1.3	1.16	60	1	1	94	1.3%	0.28 [-0.05, 0.61]
Ribeiro, 2014 [63]	473.5	913.2	101	58.5	484.3	94	1.4%	0.56 [0.27, 0.85]
Samuels, 2011 [143]	34.8	13.6	12	26.5	16.1	25	0.5%	0.53 [-0.17, 1.23]
Sharp, 2016 [144]	100	109	72	90	119	65	1.3%	0.09 [-0.25, 0.42]
Simons, 2018 [64]	27.3	26.7	55	30.1	31	63	1.2%	-0.10 [-0.46, 0.27]
Ter Hoeve, 2018 [68]	6.7	3	121	6.5	3.1	252	1.7%	0.07 [-0.15, 0.28]
Unick, 2012 [145]	133	217	11	44.8	124.6	12	0.4%	0.49 [-0.35, 1.32]
Vallance, 2008 [70]	213	173	172	180	141	166	1.7%	0.21 [-0.01, 0.42]
Vallance, 2016 [125]	143	132	41	146	155	37	1.0%	-0.02 [-0.47, 0.42]
Van Blarigan, 2019 [126]	46.6	48.4	20	54.5	24.9	19	0.6%	-0.20 [-0.83, 0.43]
Vandelanotte, 2018 [146]	148.8	181.1	78	79.8	77.1	46	1.2%	0.45 [0.08, 0.82]
Van der Weegen, 2015 [71]	48.2	23.8	65	42.9	25.8	134	1.4%	0.21 [-0.09, 0.51]
Van Hoye, 2018 [72]	17.8	49.6	157	9.6	34	50	1.3%	0.18 [-0.14, 0.49]
Wijsman, 2013 [147]	11.1	21.8	108	-0.1	15.4	105	1.5%	0.59 [0.32, 0.86]
Yates, 2017 [78]	-4.1	19	287	-5.9	17.7	272	1.8%	0.10 [-0.07, 0.26]
Subtotal (95% CI)			**5850**			**5133**	**68.0%**	**0.28 [0.21, 0.35]**

Heterogeneity: Tau²=0.04; Chi²=161.93, df=61 (P<.001); I²=62%
Test for overall effect: Z=7.81 (P<.001)

◆=Subtotal or total standardized mean difference

Figure 6. Primary outcome analysis for the outcome moderate-to-vigorous physical activity at timepoint short-term follow-up.

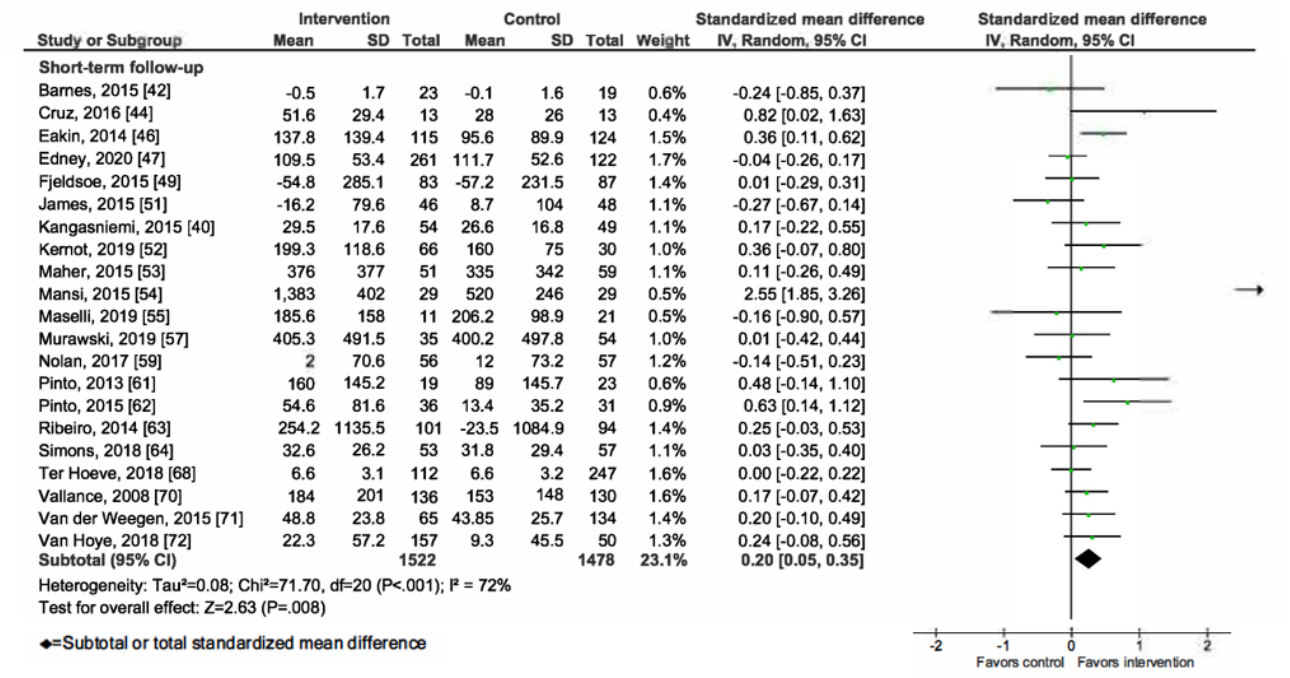


Figure 7. Primary outcome analysis for the outcome moderate-to-vigorous physical activity at timepoint long-term follow-up.

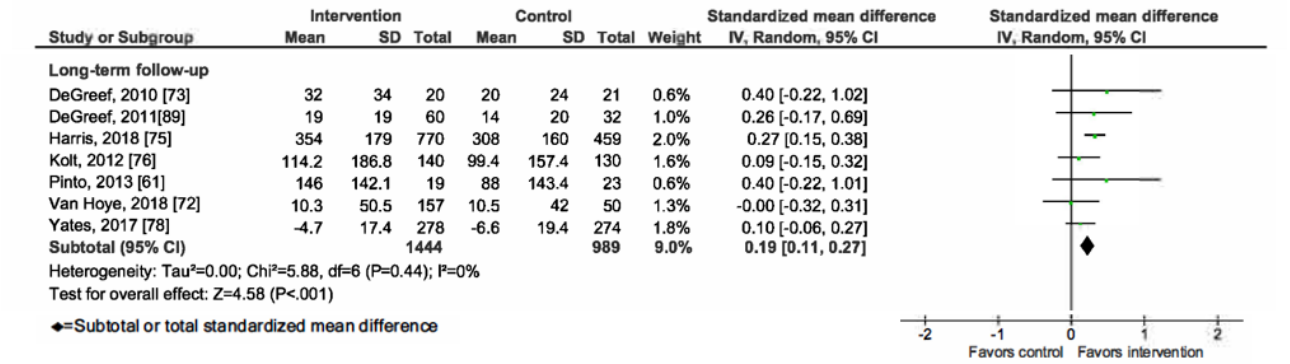


Figure 8. Primary outcome analysis by measurement time point for the outcome total physical activity.

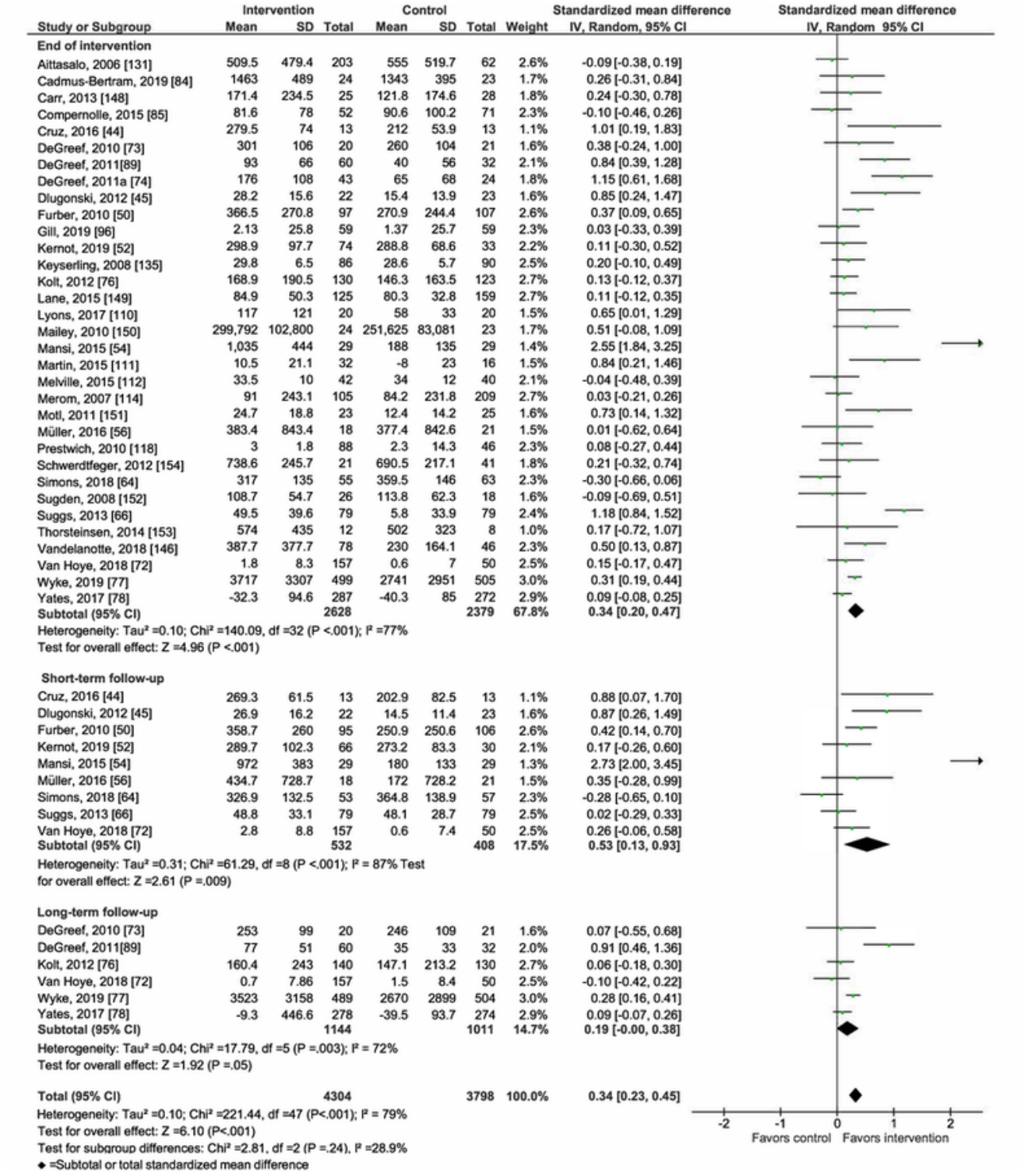
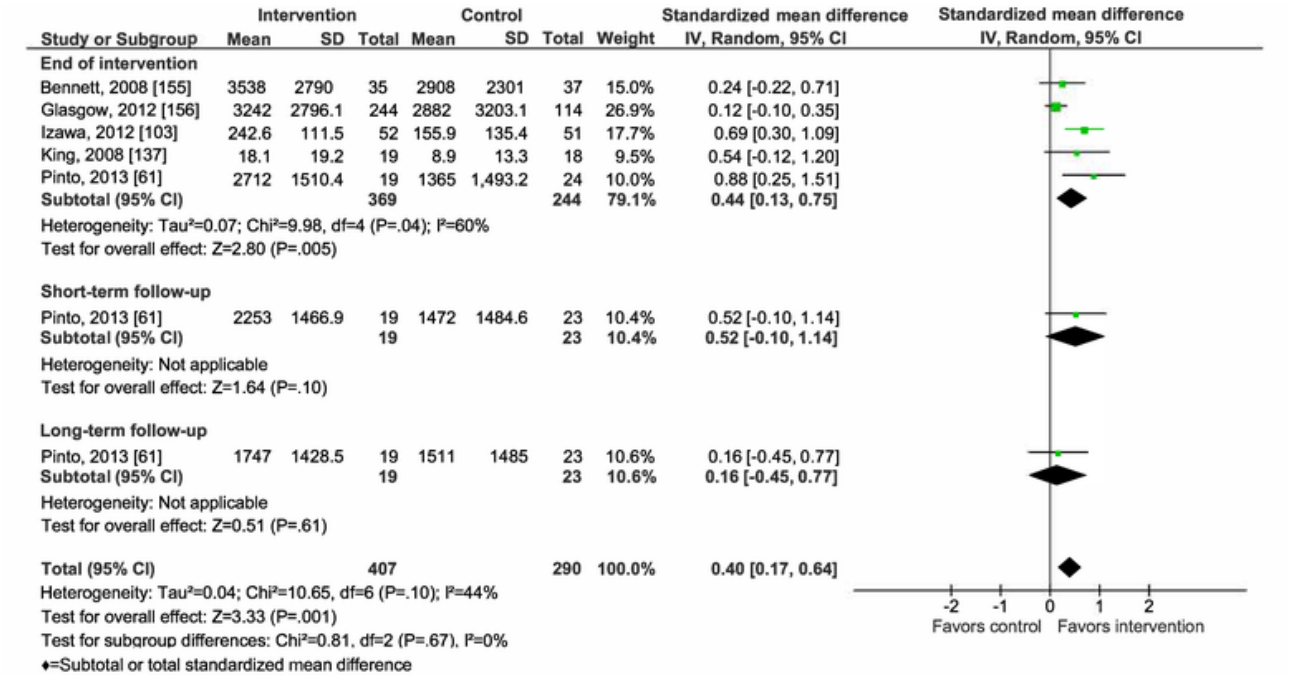


Figure 9. Primary outcome analysis by measurement time point for the outcome energy expenditure.



Publication Bias Assessment

Publication bias was assessed using funnel plot analysis for end-of-intervention measurements, as all but one study [40] reported this time point (Multimedia Appendix 5 [54,82,102]). No systematic publication bias was observed. However, the funnel plot analysis revealed 3 outlier studies [54,82,102]. We identified unusually high adherence rates [54], possibly because the research team and the study participants were based on the same campus, and a short intervention duration (only 4 weeks) [82,102] as potential reasons for the high effect scores in the outlier studies. We conducted a sensitivity analysis and excluded these studies across all outcome measures. All effects were found to be stable, and heterogeneity was substantially reduced (Multimedia Appendix 6).

Sensitivity Analyses

Sensitivity analysis by risk of bias was conducted for all outcome measures but for EE, as only one study [155] measuring EE classified as low risk of bias. Results at the end of intervention were found to be robust across outcome measures (Multimedia Appendix 6), with effect sizes substantially increasing for walking, MVPA, and TPA to moderate effect sizes. Short- and long-term follow-up effects were not statistically significant when only studies with low risk of bias were included. Heterogeneity increased and remained substantial. Sensitivity analysis of studies reporting long-term follow-up measurements was conducted for all outcome measures but for EE, as only one study [61] measuring EE reported a long-term follow-up measurement. The results across all time points were robust for all outcome measures.

Subgroup Analysis by Population Type

We used subgroup analysis to evaluate the effect moderators. Table 1 summarizes all results, and Multimedia Appendices 7 [41-154], 8 [41-154], and 9 [41-154] provide detailed forest plots for each analysis. We found that population type moderates the effect of mHealth interventions on PA. The intervention design and control group type were not found to be significant effect moderators. Subgroup analysis by population type revealed statistically significant ($P \leq .10$) quantitative subgroup effects for all outcome measures. The treatment effect at the end of intervention was greater in sick populations (walking SMD 0.44, 95% CI 0.29-0.60, $P<.001$, $I^2=71\%$, $P<.001$; MVPA SMD 0.33, 95% CI 0.21-0.45, $P<.001$, $I^2=55\%$, $P<.001$; TPA SMD 0.59, 95% CI 0.36-0.81, $P<.001$, $I^2=49\%$, $P=.03$; Multimedia Appendix 7) than in healthy populations (walking SMD 0.20, 95% CI 0.04-0.35, $P=.01$, $I^2=78\%$, $P<.001$; MVPA SMD 0.14, 95% CI 0.06-0.23, $P=.001$, $I^2=15\%$, $P=.29$; TPA SMD 0.29, 95% CI −0.10 to 0.67, $P=.14$, $I^2=85\%$, $P<.001$; Multimedia Appendix 7). Within the healthy subgroup, summary effects were only statistically significant for the outcome measures walking and MVPA. The results for at-risk populations were mixed. The outcome measures walking and MVPA exhibited high effect scores, similar to the high effect scores of sick populations (walking SMD 0.59, 95% CI 0.42-0.76, $P<.001$, $I^2=87\%$, $P<.001$; MVPA SMD 0.30, 95% CI 0.18-0.43, $P<.001$, $I^2=72\%$, $P<.001$; Multimedia Appendix 7), whereas effect scores for at-risk populations were lower for TPA (SMD 0.21, 95% CI 0.04-0.38; $P=.02$; $I^2=78\%$, $P<.001$). Although heterogeneity was somewhat reduced within most subgroups compared with the overall outcome heterogeneity, it remained high and significant. The covariate distribution between sick, at-risk, and healthy population subgroups was uneven, as fewer studies investigated preventative mHealth PA interventions in healthy populations.

Table 1. Summary of subgroup analyses results.

Outcome measure and time point	Studies, n (%)	SMD[a] (95% CI)	*P* value	Heterogeneity		Test for subgroup differences (*P* value)
				I^2 (%)	*P* value	
Population type						
Walking (n=77)						.003
Healthy	14 (18)	0.20 (0.04 to 0.35)	.01	78	<.001	
At-risk	30 (39)	0.59 (0.42 to 0.76)	<.001	87	<.001	
Sick	33 (42)	0.44 (0.29 to 0.60)	<.001	71	<.001	
MVPA[b] (n=62)						.02
Healthy	12 (19)	0.14 (0.06 to 0.23)	.001	15	.29	
At-risk	25 (40)	0.30 (0.18 to 0.43)	<.001	72	<.001	
Sick	25 (40)	0.33 (0.21 to 0.45)	<.001	55	<.001	
TPA[c] (n=33)						.03
Healthy	6 (18)	0.29 (–0.10 to 0.67)	.14	85	<.001	
At-risk	16 (48)	0.21 (0.04 to 0.38)	.02	78	<.001	
Sick	11 (33)	0.59 (0.36 to 0.81)	<.001	49	.03	
Intervention design						
Walking (n=77)						.35
Scalable	31 (40)	0.54 (0.34 to 0.74)	<.001	89	<.001	
Nonscalable	45 (58)	0.42 (0.31 to 0.54)	<.001	77	<.001	
Combined	1 (1)	0.37 (0.25 to 0.48)	<.001	N/A[d,e]	N/A	
MVPA (n=62)						.12
Scalable	23 (37)	0.20 (0.08 to 0.32)	.001	67	<.001	
Nonscalable	38 (61)	0.33 (0.24 to 0.43)	<.001	57	<.001	
Combined	1 (2)	0.37 (0.25 to 0.48)	<.001	N/A	N/A	
TPA (n=33)						.60
Scalable	12 (36)	0.39 (0.06 to 0.73)	.02	89	<.001	
Nonscalable	21 (64)	0.30 (0.18 to 0.42)	<.001	53	.002	
Combined	__[f]	—	—	—	—	
Control group type						
Walking (n=77)						.15
No or minimal intervention	62 (81)	0.47 (0.36 to 0.59)	<.001	85	<.001	
Alternative intervention	10 (13)	0.48 (0.12 to 0.83)	.009	80	<.001	
Combined	5 (6)	0.23 (0.01 to 0.45)	.04	62	.03	
MVPA (n=62)						.26
No or minimal intervention	43 (69)	0.29 (0.21 to 0.38)	<.001	67	<.001	
Alternative intervention	9 (15)	0.39 (0.14 to 0.65)	.002	62	.007	
Combined	10 (16)	0.20 (0.08 to 0.32)	<.001	34	.13	
TPA (n=33)						.006
No or minimal intervention	24 (73)	0.34 (0.19 to 0.50)	<.001	76	<.001	
Alternative intervention	6 (18)	0.48 (0.06 to 0.91)	.03	83	<.001	
Combined	3 (9)	0.00 (−0.17 to 0.17)	.97	0	.59	

[a]SMD: standardized mean difference.

[b]MVPA: moderate-to-vigorous physical activity.

[c]TPA: total physical activity.

[d]N/A: not applicable.

[e]In subgroups where n=1, heterogeneity cannot be calculated.

[f]No studies with combined subgroups. Thus, no numbers reported.

Subgroup Analysis by Intervention Design

Subgroup analysis by intervention design revealed no significant subgroup differences across 3 outcome measures (walking, *P*=.35; MVPA, *P*=.12; TPA, *P*=.60; Table 1; Multimedia Appendix 8) and did not identify mHealth intervention design as a significant effect moderator. Heterogeneity within subgroups was substantial and significant across all outcome measures (Table 1). Both scalable and nonscalable mHealth intervention designs significantly increased PA at similar levels (scalable walking SMD 0.54, 95% CI 0.34-0.74, *P*<.001, I^2=89%, *P*<.001; scalable MVPA SMD 0.20, 95% CI 0.08-0.32, *P*=.001, I^2=67%, *P*<.001; scalable TPA SMD 0.39, 95% CI 0.06-0.73, *P*=.02, I^2=89%, *P*<.001; nonscalable walking SMD 0.42, 95% CI 0.31-0.54, *P*<.001, I^2=77%, *P*<.001; nonscalable MVPA SMD 0.33, 95% CI 0.24-0.43, *P*<.001, I^2=57%, *P*<.001; nonscalable TPA SMD 0.30, 95% CI 0.18-0.42, *P*<.001, I^2=53%, *P*=.002; Multimedia Appendix 8).

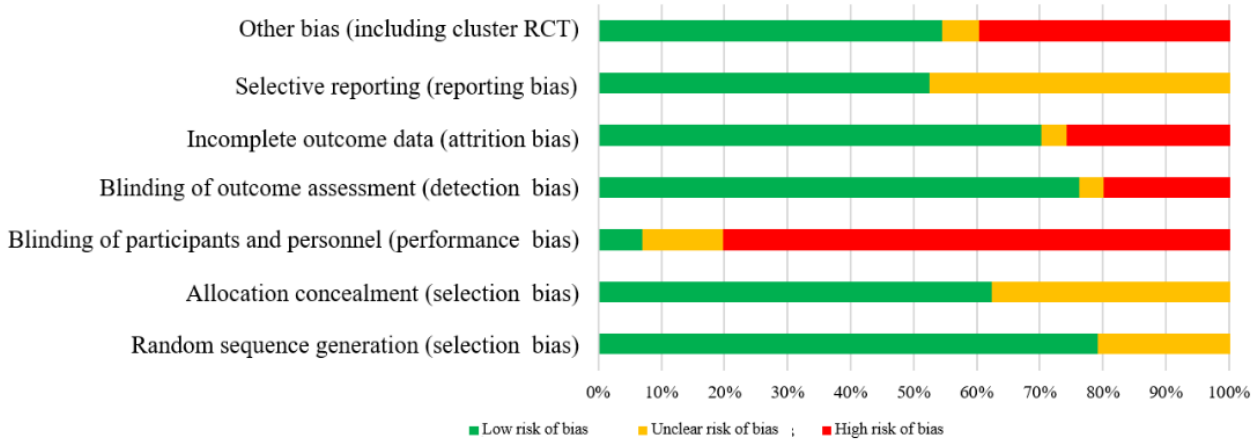
Subgroup Analysis by Control Group Type

Subgroup analysis by control group type found no statistically significant subgroup effect for the outcome measures walking (*P*=.15) and MVPA (*P*=.26). Subgroup differences were only significant for TPA (*P*=.006), where mHealth interventions led to larger effects in studies compared with alternative control groups (SMD 0.48, 95% CI 0.06-0.91; *P*=.03; I^2=83%, *P*<.001; Multimedia Appendix 9) than in studies with no or minimal control groups (SMD 0.34, 95% CI 0.19-0.50; *P*<.001; I^2=76%, *P*<.001; Multimedia Appendix 9). Subgroup analysis did not significantly reduce heterogeneity, and the covariate distribution between the no or minimal intervention subgroup and the alternative intervention subgroup was extremely uneven).

Risk of Bias in Included Studies

Figure 10 shows the overall risk of bias assessment across all included studies. Overall, 94 studies were classified as high risk because of selection bias (14/117, 11.9%), detection bias (37/117, 31.6%), attrition bias (42/117, 35.9%), reporting bias (13/117, 11.1%), and other biases (56/117, 47.9%). These mostly included baseline group indifferences or biases resulting from the respective study design (including potential cluster RCT biases) [35]. Multimedia Appendix 10 [40-156] displays the individual risk of bias assessment by study. GRADE analysis of all 4 outcomes (Multimedia Appendix 11) revealed no evidence of publication bias but evidence of inconsistency for the outcome measure EE. Thus, the overall quality of evidence rating ranged from low (walking, MPVA, and TPA) to very low (EE).

Figure 10. Summary of the overall risk of bias assessment for included studies. RCT: randomized controlled trial.



Discussion

Principal Findings

This systematic review is the most comprehensive study to date of mHealth PA interventions in adult populations. The aims of this study are to understand the long-term impacts of mHealth interventions on PA and to identify important effect moderators.

Overall, our analysis confirms the potential of mHealth interventions to increase PA at the end of intervention. We found small to moderate positive effects (SMD 0.28-0.46), which concur with previous research that reported small to large effect sizes [19,21-24,27,28]. Transforming our results into mean differences based on a representative low risk of bias study [109], we found mHealth interventions to result in 1566 incremental steps per day and an additional 36 minutes of MVPA per week. Previous research found that 1000 incremental steps per day can result in a 10% lower risk of having metabolic syndrome (MetS) and a 6% risk reduction of all-cause mortality, substantiating that mHealth interventions can result in significant

health benefits [158,159]. This study is among the first to find that activity increases are sustained beyond the end of intervention. Increased PA levels remained significant in short-term follow-ups taken on average 4.14 months after the end of intervention for the outcome measures walking, MVPA, and TPA. Long-term follow-up measurements, taken on average after 13.96 months, confirmed these results. However, effect sizes decreased over time and ranged from 0.19 to 0.25 at the long-term follow-up time point, which is equivalent to an incremental 851 steps per day and 24 minutes per week of MVPA. Our results concur with the recent review by Chaudhry et al [20], who also found maintained but decreasing effects of step-count monitoring interventions on PA; however, as Chaudhry et al [20] defined time frames from the start of intervention and this study looks at follow-up measurements after the end of intervention, absolute effect scores cannot be compared. Given the inverse relationship of PA with the prevalence of MetS [158], it can be assumed that mHealth interventions still yield health benefits in the long term. These observations are encouraging and provide initial evidence that mHealth interventions can support sustainable behavioral changes. However, our follow-up effects were not robust when only low risk of bias studies were analyzed because of the limited number of high-quality studies with longitudinal designs. Thus, as the current evidence base for studies with long-term follow-up measurements is very limited, further primary research is needed to confirm the sustained effects of mHealth on PA beyond the end of intervention.

Our analysis of effect moderators found that population type moderates the effect of mHealth on PA, whereas intervention design and control group type were not found to be effect moderators. Our evidence suggests that mHealth interventions might be most effective when targeting sick or at-risk populations, thereby supporting the indicative results by Smith et al [27]—effect sizes in sick and at-risk populations were about twice as high as in healthy populations. However, we still found mHealth interventions to be effective in all population types. These results challenge previous findings by Gal et al [23] and Romeo et al [30], who found no differences in effectiveness by population type, likely owing to the small number of studies reviewed. Previous studies found that baseline activity levels are negatively correlated with activity increases in mHealth interventions [144,160,161]. An underlying driver for the higher effectiveness of mHealth interventions in sick and at-risk populations could thus be lower baseline activity levels usually seen within these populations. However, there could also be further underlying factors, such as higher expectations that increases in PA lead to improved health outcomes (outcome expectancy). Further research is thus needed to understand the variety of underlying factors driving higher effectiveness in sick and at-risk populations. Our results provide helpful guidance to policy makers developing scaled-up mHealth intervention programs. Our results suggest that technology-enabled preventative, population-wide programs (eg, *The National Steps Challenge* [14]) might maximize their public health impact if they specifically target at-risk populations (eg, older or overweight groups). Focusing on at-risk groups should also increase the cost-effectiveness of large-scale mHealth programs.

mHealth technologies are cost-effective and scalable. However, this holds true only if technologies are effective without additional nonscalable intervention components (eg, face-to-face coaching). Previous research has found no effects in scalable mHealth intervention designs [24,30], stronger effects in nonscalable designs that combined technology with human-to-human interactions [27,32], and stronger effects when technology was used stand-alone [20]. We found preliminary evidence that mHealth interventions could be effective in scalable intervention designs. Our analysis found no significant subgroup differences between scalable and nonscalable intervention designs, suggesting that both designs can be equally effective in increasing PA. These results are promising and encourage the development of scalable mHealth intervention designs to efficiently increase PA in large population groups. Within our sample, most scalable mHealth interventions leveraged basic technologies (eg, texting, pedometers, or accelerometers), without taking advantage of more advanced mobile technologies (eg, automated individualized coaching, social comparison, and mobile apps), which could have further increased intervention effectiveness [162,163].

Our analysis is among the first to explore whether mHealth PA interventions produce results superior to alternative nonmobile interventions. We found that across the outcome measures walking, MPVA, and TPA, mHealth interventions led to increased levels of PA compared with alternative nonmobile interventions and no or minimal control groups, which accords with previous findings [21]. These results encourage the addition of mHealth technology to nonmobile PA interventions to increase their effectiveness.

Strengths and Limitations

The strengths of this study are the large number of mHealth interventions analyzed and its rigorous methodology. However, this study has several limitations. First, in line with other studies [164], we encountered large and significant heterogeneity in our results, despite performing several subgroup analyses. Our wide inclusion criteria led us to expect high heterogeneity because of the diverse multicomponent interventions, settings, and intervention durations. In addition, the uneven covariate distribution between subgroups limits the validity of our findings on effect moderators. Second, most of the included studies were classified as having a high risk of bias, and the overall quality of evidence was graded low to very low. The quality of evidence could be improved if future research agreed on standardized reporting of PA outcomes (eg, MVPA in minutes per day) and objective outcome measurement [21]. When replicating our primary results with low risk of bias studies, we could not confirm the effectiveness of mHealth interventions to increase PA beyond the end of interventions, as the available high-quality evidence was limited. Third, we did not attempt to identify unpublished reports or gray literature. Previous research has shown that excluding gray literature might exaggerate the results of a meta-analysis [165]. We tried to mitigate this limitation by conducting a funnel plot analysis to detect potential publication bias. Furthermore, we performed sensitivity analyses to assess the robustness of our results. We detected no systematic publication bias and sensitivity analyses that excluded outlier studies, confirming that our results were robust. Fourth, some

studies included in this study allowed intervention participants to keep mHealth devices after the end of intervention. This might have positively skewed the follow-up effects of our review. Finally, although this study provides initial evidence on the long-term effects of mHealth interventions, it only presents results for follow-up measurements taken on average 13.96 months after intervention, and our analysis included only 8 studies. We found that summary effects decrease over time, thus raising the question about the sustainability of positive effects. Further research is required to evaluate whether behavior change—toward a more active lifestyle—is truly sustainable in long term.

Conclusions

We conclude that mHealth interventions can moderately increase PA in adults at the end of intervention, both compared with alternative nonmobile control groups and no or minimal control groups. PA increases are maintained in follow-up measurements taken after intervention but decrease over time. Population type seems to moderate the effect of mHealth intervention on PA, with higher effectiveness in sick and at-risk populations compared with healthy population samples. mHealth interventions with scalable and nonscalable intervention designs seem to be equivalent in terms of effectiveness. Further high-quality studies investigating scalable mHealth interventions with long-term follow-up measurements are needed to confirm our results. This study concludes that mHealth technologies might not only support sustainable behavior change toward more active lifestyles but also contribute to preventing and controlling chronic disease risk.

Acknowledgments

This study was supported by the National Research Foundation, Prime Minister's Office, Singapore, under its Campus for Research Excellence and Technological Enterprise Programme and by the CSS Insurance (Switzerland). We thank Iva Milhaylova for her support during screening and extraction.

Authors' Contributions

JNK wrote the initial study protocol with inputs from AM and KI. AM wrote the adjusted study protocol. JNK designed and implemented the search strategy. AM, KI, GWT, and AJH screened and coded the primary studies and extracted data. AM analyzed the data and drafted the initial manuscript, supervised by TK. FM and LTC provided methodological guidance and feedback on the manuscript. All authors reviewed and approved the final manuscript.

Conflicts of Interest

JNK, AJH, KI, GWT and TK are affiliated with the Centre for Digital Health Interventions, a joint initiative of the Department of Management, Technology and Economics at ETH Zurich and the Institute of Technology Management at the University of St. Gallen, which was funded in part by CSS Insurance, Switzerland. TK is also a cofounder of Pathmate Technologies, a university spin-off company that creates and delivers digital clinical pathways. However, Pathmate Technologies was not involved in this research. Since January 2021, JNK is associated with CSS Insurance, Switzerland.

Multimedia Appendix 1
Overview of existing meta-analyses on the effect of mobile health interventions on physical activity.
[PDF File (Adobe PDF File), 319 KB - jmir_v23i4e26699_app1.pdf]

Multimedia Appendix 2
Overview of the search strategy and keywords.
[PDF File (Adobe PDF File), 238 KB - jmir_v23i4e26699_app2.pdf]

Multimedia Appendix 3
Search algorithms.
[PDF File (Adobe PDF File), 247 KB - jmir_v23i4e26699_app3.pdf]

Multimedia Appendix 4
Overview of the study characteristics.
[PDF File (Adobe PDF File), 431 KB - jmir_v23i4e26699_app4.pdf]

Multimedia Appendix 5
Funnel plot analysis to detect publication bias.
[PDF File (Adobe PDF File), 333 KB - jmir_v23i4e26699_app5.pdf]

Multimedia Appendix 6
Sensitivity analysis.
[PDF File (Adobe PDF File), 402 KB - jmir_v23i4e26699_app6.pdf]

Multimedia Appendix 7
Subgroup analysis by population type.
[PDF File (Adobe PDF File), 1357 KB - jmir_v23i4e26699_app7.pdf]

Multimedia Appendix 8
Subgroup analysis by intervention design.
[PDF File (Adobe PDF File), 1407 KB - jmir_v23i4e26699_app8.pdf]

Multimedia Appendix 9
Subgroup analysis by control group type.
[PDF File (Adobe PDF File), 1362 KB - jmir_v23i4e26699_app9.pdf]

Multimedia Appendix 10
Study-specific risk of bias judgments.
[PDF File (Adobe PDF File), 356 KB - jmir_v23i4e26699_app10.pdf]

Multimedia Appendix 11
Grading of recommendations, assessment, development, and evaluation quality of evidence profile.
[PDF File (Adobe PDF File), 277 KB - jmir_v23i4e26699_app11.pdf]

References

1. Mendis S, Armstrong T, Bettcher D, Branca F, Lauer J, Mace C, et al. Global Status Report on Noncommunicable Diseases 2014. Geneva: World Health Organization; 2014.
2. Guthold R, Stevens GA, Riley LM, Bull FC. Worldwide trends in insufficient physical activity from 2001 to 2016: a pooled analysis of 358 population-based surveys with 1·9 million participants. Lancet Glob Health 2018 Oct;6(10):1077-1086 [FREE Full text] [doi: 10.1016/S2214-109X(18)30357-7] [Medline: 30193830]
3. Lee I, Shiroma EJ, Lobelo F, Puska P, Blair SN, Katzmarzyk PT, Lancet Physical Activity Series Working Group. Effect of physical inactivity on major non-communicable diseases worldwide: an analysis of burden of disease and life expectancy. Lancet 2012 Jul 21;380(9838):219-229 [FREE Full text] [doi: 10.1016/S0140-6736(12)61031-9] [Medline: 22818936]
4. Sallis JF, Bull F, Guthold R, Heath GW, Inoue S, Kelly P, Lancet Physical Activity Series 2 Executive Committee. Progress in physical activity over the Olympic quadrennium. Lancet 2016 Sep 24;388(10051):1325-1336. [doi: 10.1016/S0140-6736(16)30581-5] [Medline: 27475270]
5. Bianchini F, Kaaks R, Vainio H. Weight control and physical activity in cancer prevention. Obes Rev 2002 Feb;3(1):5-8. [doi: 10.1046/j.1467-789x.2002.00046.x] [Medline: 12119660]
6. Choi KW, Chen C, Stein MB, Klimentidis YC, Wang M, Koenen KC, Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium. Assessment of bidirectional relationships between physical activity and depression among adults: a 2-sample Mendelian randomization study. JAMA Psychiatry 2019 Apr 01;76(4):399-408 [FREE Full text] [doi: 10.1001/jamapsychiatry.2018.4175] [Medline: 30673066]
7. World Health Organization. Global Action Plan for the Prevention and Control of NCDs 2013-2020. Geneva: World Health Organization; 2013:1-103.
8. Kowatsch T, Otto L, Harperink S, Cotti A, Schlieter H. A design and evaluation framework for digital health interventions. Inf Technol 2019;61(5-6):253-263 [FREE Full text] [doi: 10.1515/itit-2019-0019]
9. Reis RS, Salvo D, Ogilvie D, Lambert EV, Goenka S, Brownson RC, Lancet Physical Activity Series 2 Executive Committee. Scaling up physical activity interventions worldwide: stepping up to larger and smarter approaches to get people moving. Lancet 2016 Sep 24;388(10051):1337-1348 [FREE Full text] [doi: 10.1016/S0140-6736(16)30728-0] [Medline: 27475273]
10. Richard E, van Charante EP, Hoevenaar-Blom MP, Coley N, Barbera M, van der Groep A, et al. Healthy ageing through internet counselling in the elderly (HATICE): a multinational, randomised controlled trial. Lancet Digit Health 2019 Dec;1(8):424-434. [doi: 10.1016/s2589-7500(19)30153-0]
11. Fairburn CG, Patel V. The impact of digital technology on psychological treatments and their dissemination. Behav Res Ther 2017 Jan;88:19-25 [FREE Full text] [doi: 10.1016/j.brat.2016.08.012] [Medline: 28110672]
12. NHS Diabetes Prevention Programme – digital stream. NHS England. 2016. URL: https://tinyurl.com/u58h67ry [accessed 2020-06-03]

13. Guertler D, Vandelanotte C, Kirwan M, Duncan MJ. Engagement and nonusage attrition with a free physical activity promotion program: the case of 10,000 steps Australia. J Med Internet Res 2015 Jul 15;17(7):e176 [FREE Full text] [doi: 10.2196/jmir.4339] [Medline: 26180040]
14. Yao J, Tan CS, Chen C, Tan J, Lim N, Müller-Riemenschneider F. Bright spots, physical activity investments that work: National Steps Challenge, Singapore: a nationwide mHealth physical activity programme. Br J Sports Med 2020 Sep;54(17):1047-1048. [doi: 10.1136/bjsports-2019-101662] [Medline: 31857340]
15. Fitbit collaborates with Singapore's Health Promotion Board on population-based public health initiative in Singapore. Fitbit Press Release. 2019. URL: https://investor.fitbit.com/press/press-releases/press-release-details/2019/Fitbit-Collaborates-with-Singapores-Health-Promotion-Board-on-Population-Based-Public-Health-Initiative-in-Singapore/default.aspx [accessed 2020-07-01]
16. The Lancet Digital Health. Wearable technology and lifestyle management: the fight against obesity and diabetes. Lancet Digit Health 2019 Oct;1(6):e243 [FREE Full text] [doi: 10.1016/S2589-7500(19)30135-9] [Medline: 33323244]
17. Rowland SP, Fitzgerald JE, Holme T, Powell J, McGregor A. What is the clinical value of mHealth for patients? NPJ Digit Med 2020;3:4 [FREE Full text] [doi: 10.1038/s41746-019-0206-x] [Medline: 31970289]
18. Bravata DM, Smith-Spangler C, Sundaram V, Gienger AL, Lin N, Lewis R, et al. Using pedometers to increase physical activity and improve health: a systematic review. J Am Med Assoc 2007 Nov 21;298(19):2296-2304. [doi: 10.1001/jama.298.19.2296] [Medline: 18029834]
19. Brickwood KJ, Watson G, O'Brien J, Williams AD. Consumer-based wearable activity trackers increase physical activity participation: systematic review and meta-analysis. JMIR Mhealth Uhealth 2019 Apr 12;7(4):11819 [FREE Full text] [doi: 10.2196/11819] [Medline: 30977740]
20. Chaudhry UA, Wahlich C, Fortescue R, Cook DG, Knightly R, Harris T. The effects of step-count monitoring interventions on physical activity: systematic review and meta-analysis of community-based randomised controlled trials in adults. Int J Behav Nutr Phys Act 2020 Oct 09;17(1):129 [FREE Full text] [doi: 10.1186/s12966-020-01020-8] [Medline: 33036635]
21. De Vries HJ, Kooiman TJ, Van Ittersum MW, Van Brussel M, De Groot M. Do activity monitors increase physical activity in adults with overweight or obesity? A systematic review and meta-analysis. Obesity (Silver Spring) 2016 Oct;24(10):2078-2091 [FREE Full text] [doi: 10.1002/oby.21619] [Medline: 27670401]
22. Franssen WM, Franssen GH, Spaas J, Solmi F, Eijnde BO. Can consumer wearable activity tracker-based interventions improve physical activity and cardiometabolic health in patients with chronic diseases? A systematic review and meta-analysis of randomised controlled trials. Int J Behav Nutr Phys Act 2020 May 11;17(1):57 [FREE Full text] [doi: 10.1186/s12966-020-00955-2] [Medline: 32393357]
23. Gal R, May AM, van Overmeeren EJ, Simons M, Monninkhof EM. The effect of physical activity interventions comprising wearables and smartphone applications on physical activity: a systematic review and meta-analysis. Sports Med Open 2018 Sep 03;4(1):42 [FREE Full text] [doi: 10.1186/s40798-018-0157-9] [Medline: 30178072]
24. Hodkinson A, Kontopantelis E, Adeniji C, van Marwijk H, McMillan B, Bower P, et al. Accelerometer- and pedometer-based physical activity interventions among adults with cardiometabolic conditions: a systematic review and meta-analysis. JAMA Netw Open 2019 Oct 02;2(10):e1912895 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.12895] [Medline: 31596494]
25. Kirk MA, Amiri M, Pirbaglou M, Ritvo P. Wearable technology and physical activity behavior change in adults with chronic cardiometabolic disease: a systematic review and meta-analysis. Am J Health Promot 2019 Jun;33(5):778-791. [doi: 10.1177/0890117118816278] [Medline: 30586996]
26. Qiu S, Cai X, Chen X, Yang B, Sun Z. Step counter use in type 2 diabetes: a meta-analysis of randomized controlled trials. BMC Med 2014 Feb 27;12:36 [FREE Full text] [doi: 10.1186/1741-7015-12-36] [Medline: 24571580]
27. Smith DM, Duque L, Huffman JC, Healy BC, Celano CM. Text message interventions for physical activity: a systematic review and meta-analysis. Am J Prev Med 2020 Jan;58(1):142-151 [FREE Full text] [doi: 10.1016/j.amepre.2019.08.014] [Medline: 31759805]
28. Vaes AW, Cheung A, Atakhorrami M, Groenen MT, Amft O, Franssen FM, et al. Effect of 'activity monitor-based' counseling on physical activity and health-related outcomes in patients with chronic diseases: a systematic review and meta-analysis. Ann Med 2013 Sep;45(5-6):397-412. [doi: 10.3109/07853890.2013.810891] [Medline: 23952917]
29. Direito A, Carraça E, Rawstorn J, Whittaker R, Maddison R. mHealth technologies to influence physical activity and sedentary behaviors: behavior change techniques, systematic review and meta-analysis of randomized controlled trials. Ann Behav Med 2017 Apr;51(2):226-239. [doi: 10.1007/s12160-016-9846-0] [Medline: 27757789]
30. Romeo A, Edney S, Plotnikoff R, Curtis R, Ryan J, Sanders I, et al. Can smartphone apps increase physical activity? Systematic review and meta-analysis. J Med Internet Res 2019 Mar 19;21(3):12053 [FREE Full text] [doi: 10.2196/12053] [Medline: 30888321]
31. Yerrakalva D, Yerrakalva D, Hajna S, Griffin S. Effects of mobile health app interventions on sedentary time, physical activity, and fitness in older adults: systematic review and meta-analysis. J Med Internet Res 2019 Nov 28;21(11):14343 [FREE Full text] [doi: 10.2196/14343] [Medline: 31778121]
32. Schoeppe S, Alley S, Van Lippevelde W, Bray NA, Williams SL, Duncan MJ, et al. Efficacy of interventions that use apps to improve diet, physical activity and sedentary behaviour: a systematic review. Int J Behav Nutr Phys Act 2016 Dec 07;13(1):127 [FREE Full text] [doi: 10.1186/s12966-016-0454-y] [Medline: 27927218]

33. Higgins JP, Green S. Cochrane Handbook for Systematic Reviews of Interventions: Cochrane Book Series. Hoboken, New Jersey, United States: John Wiley & Sons Inc; 2008:1-649.
34. Higgins JP, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, Cochrane Bias Methods Group, Cochrane Statistical Methods Group. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. Br Med J 2011 Oct 18;343:d5928 [FREE Full text] [doi: 10.1136/bmj.d5928] [Medline: 22008217]
35. Eldridge S, Campbell M, Campbell M, Dahota A, Giraudeau B, Higgins J, et al. Revised Cochrane risk of bias tool for randomized trials (RoB 2.0), additional considerations for cluster-randomized trials. 2016. URL: https://sites.google.com/site/riskofbiastool/welcome/rob-2-0-tool/archive-rob-2-0-cluster-randomized-trials-2016 [accessed 2020-07-01]
36. DerSimonian R, Laird N. Meta-analysis in clinical trials. Control Clin Trials 1986 Sep;7(3):177-188. [doi: 10.1016/0197-2456(86)90046-2] [Medline: 3802833]
37. Richardson M, Garner P, Donegan S. Interpretation of subgroup analyses in systematic reviews: a tutorial. Clin Epidemiol Glob Health 2019 Jun;7(2):192-198. [doi: 10.1016/j.cegh.2018.05.005]
38. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. Br Med J 2003 Sep 06;327(7414):557-560 [FREE Full text] [doi: 10.1136/bmj.327.7414.557] [Medline: 12958120]
39. Guyatt G, Oxman AD, Akl EA, Kunz R, Vist G, Brozek J, et al. GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables. J Clin Epidemiol 2011 Apr;64(4):383-394. [doi: 10.1016/j.jclinepi.2010.04.026] [Medline: 21195583]
40. Kangasniemi AM, Lappalainen R, Kankaanpää A, Tolvanen A, Tammelin T. Towards a physically more active lifestyle based on one's own values: the results of a randomized controlled trial among physically inactive adults. BMC Public Health 2015 Mar 18;15:260 [FREE Full text] [doi: 10.1186/s12889-015-1604-x] [Medline: 25848812]
41. Aittasalo M, Rinne M, Pasanen M, Kukkonen-Harjula K, Vasankari T. Promoting walking among office employees - evaluation of a randomized controlled intervention with pedometers and e-mail messages. BMC Public Health 2012 Jun 06;12:403 [FREE Full text] [doi: 10.1186/1471-2458-12-403] [Medline: 22672576]
42. Barnes AT, Plotnikoff RC, Collins CE, Morgan PJ. Feasibility and preliminary efficacy of the MADE4Life Program: a pilot randomized controlled trial. J Phys Act Health 2015 Oct;12(10):1378-1393. [doi: 10.1123/jpah.2014-0331] [Medline: 25599119]
43. Coelho CM, Reboredo MM, Valle FM, Malaguti C, Campos LA, Nascimento LM, et al. Effects of an unsupervised pedometer-based physical activity program on daily steps of adults with moderate to severe asthma: a randomized controlled trial. J Sports Sci 2018 May;36(10):1186-1193. [doi: 10.1080/02640414.2017.1364402] [Medline: 28799458]
44. Cruz J, Brooks D, Marques A. Walk2Bactive: a randomised controlled trial of a physical activity-focused behavioural intervention beyond pulmonary rehabilitation in chronic obstructive pulmonary disease. Chron Respir Dis 2016 Feb;13(1):57-66 [FREE Full text] [doi: 10.1177/1479972315619574] [Medline: 26703921]
45. Dlugonski D, Motl RW, Mohr DC, Sandroff BM. Internet-delivered behavioral intervention to increase physical activity in persons with multiple sclerosis: sustainability and secondary outcomes. Psychol Health Med 2012;17(6):636-651. [doi: 10.1080/13548506.2011.652640] [Medline: 22313192]
46. Eakin EG, Winkler EA, Dunstan DW, Healy GN, Owen N, Marshall AM, et al. Living well with diabetes: 24-month outcomes from a randomized trial of telephone-delivered weight loss and physical activity intervention to improve glycemic control. Diabetes Care 2014 Aug;37(8):2177-2185. [doi: 10.2337/dc13-2427] [Medline: 24658390]
47. Edney SM, Olds TS, Ryan JC, Vandelanotte C, Plotnikoff RC, Curtis RG, et al. A social networking and gamified app to increase physical activity: cluster RCT. Am J Prev Med 2020 Feb;58(2):51-62. [doi: 10.1016/j.amepre.2019.09.009] [Medline: 31959326]
48. Finkelstein EA, Haaland BA, Bilger M, Sahasranaman A, Sloan RA, Nang EE, et al. Effectiveness of activity trackers with and without incentives to increase physical activity (TRIPPA): a randomised controlled trial. Lancet Diabetes Endocrinol 2016 Dec;4(12):983-995. [doi: 10.1016/S2213-8587(16)30284-4] [Medline: 27717766]
49. Fjeldsoe BS, Miller YD, Graves N, Barnett AG, Marshall AL. Randomized controlled trial of an improved version of mobilemums, an intervention for increasing physical activity in women with young children. Ann Behav Med 2015 Aug;49(4):487-499. [doi: 10.1007/s12160-014-9675-y] [Medline: 25582987]
50. Furber S, Butler L, Phongsavan P, Mark A, Bauman A. Randomised controlled trial of a pedometer-based telephone intervention to increase physical activity among cardiac patients not attending cardiac rehabilitation. Patient Educ Couns 2010 Aug;80(2):212-218. [doi: 10.1016/j.pec.2009.11.012] [Medline: 20022201]
51. James EL, Stacey FG, Chapman K, Boyes AW, Burrows T, Girgis A, et al. Impact of a nutrition and physical activity intervention (ENRICH: Exercise and Nutrition Routine Improving Cancer Health) on health behaviors of cancer survivors and carers: a pragmatic randomized controlled trial. BMC Cancer 2015 Oct 15;15:710 [FREE Full text] [doi: 10.1186/s12885-015-1775-y] [Medline: 26471791]
52. Kernot J, Lewis L, Olds T, Maher C. Effectiveness of a Facebook-delivered physical activity intervention for postpartum women: a randomized controlled trial. J Phys Act Health 2019 Feb 01;16(2):125-133. [doi: 10.1123/jpah.2017-0573] [Medline: 30541378]

53. Maher C, Ferguson M, Vandelanotte C, Plotnikoff R, De Bourdeaudhuij I, Thomas S, et al. A web-based, social networking physical activity intervention for insufficiently active adults delivered via Facebook app: randomized controlled trial. J Med Internet Res 2015 Jul 13;17(7):e174 [FREE Full text] [doi: 10.2196/jmir.4086] [Medline: 26169067]
54. Mansi S, Milosavljevic S, Tumilty S, Hendrick P, Higgs C, Baxter DG. Investigating the effect of a 3-month workplace-based pedometer-driven walking programme on health-related quality of life in meat processing workers: a feasibility study within a randomized controlled trial. BMC Public Health 2015 Apr 22;15:410 [FREE Full text] [doi: 10.1186/s12889-015-1736-z] [Medline: 25895747]
55. Maselli M, Gobbi E, Carraro A. Effectiveness of individual counseling and activity monitors to promote physical activity among university students. J Sports Med Phys Fitness 2019 Jan;59(1):132-140. [doi: 10.23736/S0022-4707.17.07981-6] [Medline: 29199784]
56. Müller AM, Khoo S, Morris T. Text messaging for exercise promotion in older adults from an upper-middle-income country: randomized controlled trial. J Med Internet Res 2016 Jan 07;18(1):e5 [FREE Full text] [doi: 10.2196/jmir.5235] [Medline: 26742999]
57. Murawski B, Plotnikoff RC, Rayward AT, Oldmeadow C, Vandelanotte C, Brown WJ, et al. Efficacy of an m-Health physical activity and sleep health intervention for adults: a randomized waitlist-controlled trial. Am J Prev Med 2019 Oct;57(4):503-514. [doi: 10.1016/j.amepre.2019.05.009] [Medline: 31542128]
58. Mutrie N, Doolin O, Fitzsimons CF, Grant PM, Granat M, Grealy M, et al. Increasing older adults' walking through primary care: results of a pilot randomized controlled trial. Fam Pract 2012 Dec;29(6):633-642 [FREE Full text] [doi: 10.1093/fampra/cms038] [Medline: 22843637]
59. Nolan CM, Maddocks M, Canavan JL, Jones SE, Delogu V, Kaliaraju D, et al. Pedometer step count targets during pulmonary rehabilitation in chronic obstructive pulmonary disease. A randomized controlled trial. Am J Respir Crit Care Med 2017 May 15;195(10):1344-1352. [doi: 10.1164/rccm.201607-1372oc]
60. Oliveira JS, Sherrington C, Paul SS, Ramsay E, Chamberlain K, Kirkham C, et al. A combined physical activity and fall prevention intervention improved mobility-related goal attainment but not physical activity in older adults: a randomised trial. J Physiother 2019 Jan;65(1):16-22 [FREE Full text] [doi: 10.1016/j.jphys.2018.11.005] [Medline: 30581138]
61. Pinto BM, Papandonatos GD, Goldstein MG, Marcus BH, Farrell N. Home-based physical activity intervention for colorectal cancer survivors. Psychooncology 2013 Jan;22(1):54-64. [doi: 10.1002/pon.2047] [Medline: 21905158]
62. Pinto BM, Stein K, Dunsiger S. Peers promoting physical activity among breast cancer survivors: a randomized controlled trial. Health Psychol 2015 May;34(5):463-472 [FREE Full text] [doi: 10.1037/hea0000120] [Medline: 25110844]
63. Ribeiro MA, Martins MA, Carvalho CRF. Interventions to increase physical activity in middle-age women at the workplace: a randomized controlled trial. Med Sci Sports Exerc 2014;46(5):1008-1015. [doi: 10.1249/MSS.0000000000000190] [Medline: 24126967]
64. Simons D, De Bourdeaudhuij I, Clarys P, De Cocker K, Vandelanotte C, Deforche B. Effect and process evaluation of a smartphone app to promote an active lifestyle in lower educated working young adults: cluster randomized controlled trial. JMIR Mhealth Uhealth 2018 Aug 24;6(8):e10003 [FREE Full text] [doi: 10.2196/10003] [Medline: 30143477]
65. Stacey FG, James EL, Chapman K, Lubans DR. Social cognitive theory mediators of physical activity in a lifestyle program for cancer survivors and carers: findings from the ENRICH randomized controlled trial. Int J Behav Nutr Phys Act 2016 Apr 14;13:49 [FREE Full text] [doi: 10.1186/s12966-016-0372-z] [Medline: 27075417]
66. Suggs S, Blake H, Bardus M, Lloyd S. Effects of text messaging in addition to emails on physical activity among university and college employees in the UK. J Health Serv Res Policy 2013 Apr;18(1 Suppl):56-64. [doi: 10.1177/1355819613478001] [Medline: 27552780]
67. Talbot LA, Gaines JM, Huynh TN, Metter EJ. A home-based pedometer-driven walking program to increase physical activity in older adults with osteoarthritis of the knee: a preliminary study. J Am Geriatr Soc 2003 Mar;51(3):387-392. [doi: 10.1046/j.1532-5415.2003.51113.x] [Medline: 12588583]
68. Hoeve NT, Sunamura M, Stam HJ, Boersma E, Geleijnse ML, van Domburg RT, et al. Effects of two behavioral cardiac rehabilitation interventions on physical activity: a randomized controlled trial. Int J Cardiol 2018 Mar 15;255:221-228. [doi: 10.1016/j.ijcard.2017.12.015] [Medline: 29425564]
69. Tudor-Locke C, Bell RC, Myers AM, Harris SB, Ecclestone NA, Lauzon N, et al. Controlled outcome evaluation of the First Step Program: a daily physical activity intervention for individuals with type II diabetes. Int J Obes Relat Metab Disord 2004 Jan;28(1):113-119. [doi: 10.1038/sj.ijo.0802485] [Medline: 14569279]
70. Vallance JK, Courneya KS, Plotnikoff RC, Dinu I, Mackey JR. Maintenance of physical activity in breast cancer survivors after a randomized trial. Med Sci Sports Exerc 2008 Jan;40(1):173-180. [doi: 10.1249/mss.0b013e3181586b41] [Medline: 18091007]
71. van der Weegen S, Verwey R, Spreeuwenberg M, Tange H, van der Weijden T, de Witte L. It's LiFe! Mobile and web-based monitoring and feedback tool embedded in primary care increases physical activity: a cluster randomized controlled trial. J Med Internet Res 2015 Jul 24;17(7):e184 [FREE Full text] [doi: 10.2196/jmir.4579] [Medline: 26209025]
72. Van Hoye K, Wijtzes AI, Lefevre J, De Baere S, Boen F. Year-round effects of a four-week randomized controlled trial using different types of feedback on employees' physical activity. BMC Public Health 2018 Apr 12;18(1):492 [FREE Full text] [doi: 10.1186/s12889-018-5402-0] [Medline: 29649997]

73. De Greef K, Deforche B, Tudor-Locke C, De Bourdeaudhuij I. A cognitive-behavioural pedometer-based group intervention on physical activity and sedentary behaviour in individuals with type 2 diabetes. Health Educ Res 2010 Oct;25(5):724-736 [FREE Full text] [doi: 10.1093/her/cyq017] [Medline: 20338978]
74. De Greef KP, Deforche BI, Ruige JB, Bouckaert JJ, Tudor-Locke CE, Kaufman J, et al. The effects of a pedometer-based behavioral modification program with telephone support on physical activity and sedentary behavior in type 2 diabetes patients. Patient Educ Couns 2011 Aug;84(2):275-279. [doi: 10.1016/j.pec.2010.07.010] [Medline: 20732776]
75. Harris T, Kerry SM, Limb ES, Furness C, Wahlich C, Victor CR, et al. Physical activity levels in adults and older adults 3-4 years after pedometer-based walking interventions: Long-term follow-up of participants from two randomised controlled trials in UK primary care. PLoS Med 2018 Mar;15(3):e1002526 [FREE Full text] [doi: 10.1371/journal.pmed.1002526] [Medline: 29522529]
76. Kolt GS, Schofield GM, Kerse N, Garrett N, Ashton T, Patel A. Healthy Steps trial: pedometer-based advice and physical activity for low-active older adults. Ann Fam Med 2012;10(3):206-212 [FREE Full text] [doi: 10.1370/afm.1345] [Medline: 22585884]
77. Wyke S, Bunn C, Andersen E, Silva MN, Van Nassau F, McSkimming P, et al. Correction: the effect of a programme to improve men's sedentary time and physical activity: the European Fans in Training (EuroFIT) randomised controlled trial. PLoS Med 2019 Mar;16(3):e1002772 [FREE Full text] [doi: 10.1371/journal.pmed.1002772] [Medline: 30870433]
78. Yates T, Edwardson CL, Henson J, Gray LJ, Ashra NB, Troughton J, et al. Walking away from type 2 diabetes: a cluster randomized controlled trial. Diabet Med 2017 May;34(5):698-707. [doi: 10.1111/dme.13254] [Medline: 27589017]
79. Alsaleh E, Windle R, Blake H. Behavioural intervention to increase physical activity in adults with coronary heart disease in Jordan. BMC Public Health 2016 Jul 26;16:643 [FREE Full text] [doi: 10.1186/s12889-016-3313-5] [Medline: 27460675]
80. Ashton LM, Morgan PJ, Hutchesson MJ, Rollo ME, Collins CE. Feasibility and preliminary efficacy of the 'HEYMAN' healthy lifestyle program for young men: a pilot randomised controlled trial. Nutr J 2017 Jan 13;16(1):2 [FREE Full text] [doi: 10.1186/s12937-017-0227-8] [Medline: 28086890]
81. Baker G, Gray SR, Wright A, Fitzsimons C, Nimmo M, Lowry R, Scottish Physical Activity Research Collaboration (SPARColl). The effect of a pedometer-based community walking intervention "Walking for Wellbeing in the West" on physical activity levels and health outcomes: a 12-week randomized controlled trial. Int J Behav Nutr Phys Act 2008 Sep 05;5:44 [FREE Full text] [doi: 10.1186/1479-5868-5-44] [Medline: 18775062]
82. Barwais FA, Cuddihy TF, Tomson LM. Physical activity, sedentary behavior and total wellness changes among sedentary adults: a 4-week randomized controlled trial. Health Qual Life Outcomes 2013 Oct 29;11:183 [FREE Full text] [doi: 10.1186/1477-7525-11-183] [Medline: 24168638]
83. Butler L, Dwyer D. Pedometers may not provide a positive effect on walking activity. Health Promot J Aust 2004 Aug 01;15(2):134-136. [doi: 10.1071/he04134]
84. Cadmus-Bertram L, Tevaarwerk AJ, Sesto ME, Gangnon R, Van Remortel B, Date P. Building a physical activity intervention into clinical care for breast and colorectal cancer survivors in Wisconsin: a randomized controlled pilot trial. J Cancer Surviv 2019 Aug;13(4):593-602 [FREE Full text] [doi: 10.1007/s11764-019-00778-6] [Medline: 31264183]
85. Compernolle S, Vandelanotte C, Cardon G, De Bourdeaudhuij I, De Cocker K. Effectiveness of a web-based, computer-tailored, pedometer-based physical activity intervention for adults: a cluster randomized controlled trial. J Med Internet Res 2015 Feb 09;17(2):e38 [FREE Full text] [doi: 10.2196/jmir.3402] [Medline: 25665498]
86. Creel DB, Schuh LM, Reed CA, Gomez AR, Hurst LA, Stote J, et al. A randomized trial comparing two interventions to increase physical activity among patients undergoing bariatric surgery. Obesity (Silver Spring) 2016 Aug;24(8):1660-1668 [FREE Full text] [doi: 10.1002/oby.21548] [Medline: 27367821]
87. Dadaczynski K, Schiemann S, Backhaus O. Promoting physical activity in worksite settings: results of a German pilot study of the online intervention Healingo fit. BMC Public Health 2017 Sep 08;17(1):696 [FREE Full text] [doi: 10.1186/s12889-017-4697-6] [Medline: 28886734]
88. de Blok BM, de Greef MH, Hacken NH, Sprenger SR, Postema K, Wempe JB. The effects of a lifestyle physical activity counseling program with feedback of a pedometer during pulmonary rehabilitation in patients with COPD: a pilot study. Patient Educ Couns 2006 Apr;61(1):48-55. [doi: 10.1016/j.pec.2005.02.005] [Medline: 16455222]
89. De Greef K, Deforche B, Tudor-Locke C, De Bourdeaudhuij I. Increasing physical activity in Belgian type 2 diabetes patients: a three-arm randomized controlled trial. Int J Behav Med 2011 Sep;18(3):188-198. [doi: 10.1007/s12529-010-9124-7] [Medline: 21052886]
90. Demeyer H, Louvaris Z, Frei A, Rabinovich RA, de Jong C, Gimeno-Santos E, Mr Papp PROactive study groupthe PROactive consortium. Physical activity is increased by a 12-week semiautomated telecoaching programme in patients with COPD: a multicentre randomised controlled trial. Thorax 2017 May;72(5):415-423 [FREE Full text] [doi: 10.1136/thoraxjnl-2016-209026] [Medline: 28137918]
91. Dishman RK, DeJoy DM, Wilson MG, Vandenberg RJ. Move to improve: a randomized workplace trial to increase physical activity. Am J Prev Med 2009 Feb;36(2):133-141. [doi: 10.1016/j.amepre.2008.09.038] [Medline: 19135905]
92. Duru OK, Sarkisian CA, Leng M, Mangione CM. Sisters in motion: a randomized controlled trial of a faith-based physical activity intervention. J Am Geriatr Soc 2010 Oct;58(10):1863-1869 [FREE Full text] [doi: 10.1111/j.1532-5415.2010.03082.x] [Medline: 20929464]

93. Engel L, Lindner H. Impact of using a pedometer on time spent walking in older adults with type 2 diabetes. Diabetes Educ 2006;32(1):98-107. [doi: 10.1177/0145721705284373] [Medline: 16439498]
94. Fjeldsoe BS, Miller YD, Marshall AL. MobileMums: a randomized controlled trial of an SMS-based physical activity intervention. Ann Behav Med 2010 May;39(2):101-111. [doi: 10.1007/s12160-010-9170-z] [Medline: 20174902]
95. Gell NM, Wadsworth DD. The use of text messaging to promote physical activity in working women: a randomized controlled trial. J Phys Act Health 2015 Jun;12(6):756-763 [FREE Full text] [doi: 10.1123/jpah.2013-0144] [Medline: 25110303]
96. Gill DP, Blunt W, Silva NC, Stiller-Moldovan C, Zou GY, Petrella RJ. The HealtheSteps™ lifestyle prescription program to improve physical activity and modifiable risk factors for chronic disease: a pragmatic randomized controlled trial. BMC Public Health 2019 Jun 28;19(1):841 [FREE Full text] [doi: 10.1186/s12889-019-7141-2] [Medline: 31253112]
97. Glynn LG, Hayes PS, Casey M, Glynn F, Alvarez-Iglesias A, Newell J, et al. Effectiveness of a smartphone application to promote physical activity in primary care: the SMART MOVE randomised controlled trial. Br J Gen Pract 2014 Jul;64(624):384-391 [FREE Full text] [doi: 10.3399/bjgp14X680461] [Medline: 24982490]
98. Hardeman W, Mitchell J, Pears S, Van Emmenis M, Theil F, Gc VS, VBI Research Team. Evaluation of a very brief pedometer-based physical activity intervention delivered in NHS Health Checks in England: the VBI randomised controlled trial. PLoS Med 2020 Mar;17(3):e1003046 [FREE Full text] [doi: 10.1371/journal.pmed.1003046] [Medline: 32142507]
99. Hornikx M, Demeyer H, Camillo CA, Janssens W, Troosters T. The effects of a physical activity counseling program after an exacerbation in patients with Chronic Obstructive Pulmonary Disease: a randomized controlled pilot study. BMC Pulm Med 2015 Nov 04;15:136 [FREE Full text] [doi: 10.1186/s12890-015-0126-8] [Medline: 26530543]
100. Hospes G, Bossenbroek L, Hacken NH, Van Hengel P, De Greef MH. Enhancement of daily physical activity increases physical fitness of outclinic COPD patients: results of an exercise counseling program. Patient Educ Couns 2009 May;75(2):274-278. [doi: 10.1016/j.pec.2008.10.005] [Medline: 19036552]
101. Houle J, Doyon O, Vadeboncoeur N, Turbide G, Diaz A, Poirier P. Innovative program to increase physical activity following an acute coronary syndrome: randomized controlled trial. Patient Educ Couns 2011 Dec;85(3):237-244. [doi: 10.1016/j.pec.2011.03.018] [Medline: 21546203]
102. Hultquist CN, Albright C, Thompson DL. Comparison of walking recommendations in previously inactive women. Med Sci Sports Exerc 2005 Apr;37(4):676-683. [doi: 10.1249/01.mss.0000158993.39760.1b] [Medline: 15809569]
103. Izawa KP, Watanabe S, Hiraki K, Morio Y, Kasahara Y, Takeichi N, et al. Determination of the effectiveness of accelerometer use in the promotion of physical activity in cardiac patients: a randomized controlled trial. Arch Phys Med Rehabil 2012 Nov;93(11):1896-1902. [doi: 10.1016/j.apmr.2012.06.015] [Medline: 22750166]
104. Katzmarzyk PT, Champagne CM, Tudor-Locke C, Broyles ST, Harsha D, Kennedy BM, et al. A short-term physical activity randomized trial in the Lower Mississippi Delta. PLoS One 2011;6(10):e26667 [FREE Full text] [doi: 10.1371/journal.pone.0026667] [Medline: 22046325]
105. Kawagoshi A, Kiyokawa N, Sugawara K, Takahashi H, Sakata S, Satake M, et al. Effects of low-intensity exercise and home-based pulmonary rehabilitation with pedometer feedback on physical activity in elderly patients with chronic obstructive pulmonary disease. Respir Med 2015 Mar;109(3):364-371 [FREE Full text] [doi: 10.1016/j.rmed.2015.01.008] [Medline: 25682543]
106. King AC, Bickmore TW, Campero MI, Pruitt LA, Yin JL. Employing virtual advisors in preventive care for underserved communities: results from the COMPASS study. J Health Commun 2013;18(12):1449-1464 [FREE Full text] [doi: 10.1080/10810730.2013.798374] [Medline: 23941610]
107. Li LC, Feehan LM, Xie H, Lu N, Shaw C, Gromala D, et al. Efficacy of a physical activity counseling program with use of a wearable tracker in people with inflammatory arthritis: a randomized controlled trial. Arthritis Care Res (Hoboken) 2020 Dec;72(12):1755-1765. [doi: 10.1002/acr.24199] [Medline: 32248626]
108. Long JE, Ring C, Bosch JA, Eves F, Drayson MT, Calver R, et al. A life-style physical activity intervention and the antibody response to pneumococcal vaccination in women. Psychosom Med 2013 Oct;75(8):774-782. [doi: 10.1097/PSY.0b013e3182a0b664] [Medline: 23922400]
109. Lynch BM, Nguyen NH, Moore MM, Reeves MM, Rosenberg DE, Boyle T, et al. A randomized controlled trial of a wearable technology-based intervention for increasing moderate to vigorous physical activity and reducing sedentary behavior in breast cancer survivors: the ACTIVATE trial. Cancer 2019 Aug 15;125(16):2846-2855 [FREE Full text] [doi: 10.1002/cncr.32143] [Medline: 31012970]
110. Lyons EJ, Swartz MC, Lewis ZH, Martinez E, Jennings K. Feasibility and acceptability of a wearable technology physical activity intervention with telephone counseling for mid-aged and older adults: a randomized controlled pilot trial. JMIR Mhealth Uhealth 2017 Mar 06;5(3):e28 [FREE Full text] [doi: 10.2196/mhealth.6967] [Medline: 28264796]
111. Martin SS, Feldman DI, Blumenthal RS, Jones SR, Post WS, McKibben RA, et al. mActive: a randomized clinical trial of an automated mhealth intervention for physical activity promotion. J Am Heart Assoc 2015 Nov 09;4(11):1-9 [FREE Full text] [doi: 10.1161/JAHA.115.002239] [Medline: 26553211]
112. Melville CA, Mitchell F, Stalker K, Matthews L, McConnachie A, Murray HM, et al. Effectiveness of a walking programme to support adults with intellectual disabilities to increase physical activity: walk well cluster-randomised controlled trial. Int J Behav Nutr Phys Act 2015 Sep 29;12:125 [FREE Full text] [doi: 10.1186/s12966-015-0290-5] [Medline: 26416606]

113. Mendoza L, Horta P, Espinoza J, Aguilera M, Balmaceda N, Castro A, et al. Pedometers to enhance physical activity in COPD: a randomised controlled trial. Eur Respir J 2015 Feb;45(2):347-354 [FREE Full text] [doi: 10.1183/09031936.00084514] [Medline: 25261324]
114. Merom D, Rissel C, Phongsavan P, Smith BJ, Van Kemenade C, Brown WJ, et al. Promoting walking with pedometers in the community: the step-by-step trial. Am J Prev Med 2007 Apr;32(4):290-297. [doi: 10.1016/j.amepre.2006.12.007] [Medline: 17303369]
115. Paul L, Wyke S, Brewster S, Sattar N, Gill JMR, Alexander G, et al. Increasing physical activity in stroke survivors using STARFISH, an interactive mobile phone application: a pilot study. Top Stroke Rehabil 2016 Jun;23(3):170-177. [doi: 10.1080/10749357.2015.1122266] [Medline: 27077973]
116. Poirier J, Bennett WL, Jerome GJ, Shah NG, Lazo M, Yeh H, et al. Effectiveness of an activity tracker- and internet-based adaptive walking program for adults: a randomized controlled trial. J Med Internet Res 2016 Feb 09;18(2):e34 [FREE Full text] [doi: 10.2196/jmir.5295] [Medline: 26860434]
117. Pope ZC, Zeng N, Zhang R, Lee HY, Gao Z. Effectiveness of combined smartwatch and social media intervention on breast cancer survivor health outcomes: a 10-week pilot randomized trial. J Clin Med 2018 Jun 07;7(6):1-17 [FREE Full text] [doi: 10.3390/jcm7060140] [Medline: 29880779]
118. Prestwich A, Perugini M, Hurling R. Can implementation intentions and text messages promote brisk walking? A randomized trial. Health Psychol 2010 Jan;29(1):40-49. [doi: 10.1037/a0016993] [Medline: 20063934]
119. Reijonsaari K, Vehtari A, Kahilakoski O, Van Mechelen W, Aro T, Taimela S. The effectiveness of physical activity monitoring and distance counseling in an occupational setting - results from a randomized controlled trial (CoAct). BMC Public Health 2012 May 11;12:344 [FREE Full text] [doi: 10.1186/1471-2458-12-344] [Medline: 22578104]
120. Roos R, Myezwa H, Van Aswegen H, Musenge E. Effects of an education and home-based pedometer walking program on ischemic heart disease risk factors in people infected with HIV: a randomized trial. J Acquir Immune Defic Syndr 2014 Nov 01;67(3):268-276. [doi: 10.1097/QAI.0000000000000299] [Medline: 25140908]
121. Rowley TW, Lenz EK, Swartz AM, Miller NE, Maeda H, Strath SJ. Efficacy of an individually tailored, internet-mediated physical activity intervention in older adults: a randomized controlled trial. J Appl Gerontol 2019 Jul;38(7):1011-1022. [doi: 10.1177/0733464817735396] [Medline: 29165018]
122. Spence JC, Burgess J, Rodgers W, Murray T. Effect of pretesting on intentions and behaviour: a pedometer and walking intervention. Psychol Health 2009 Sep;24(7):777-789. [doi: 10.1080/08870440801989938] [Medline: 20205026]
123. Tabak M, Vollenbroek-Hutten MM, Van der Valk PD, Van der Palen J, Hermens HJ. A telerehabilitation intervention for patients with Chronic Obstructive Pulmonary Disease: a randomized controlled pilot trial. Clin Rehabil 2014 Jun;28(6):582-591. [doi: 10.1177/0269215513512495] [Medline: 24293120]
124. Thorndike AN, Mills S, Sonnenberg L, Palakshappa D, Gao T, Pau CT, et al. Activity monitor intervention to promote physical activity of physicians-in-training: randomized controlled trial. PLoS One 2014;9(6):e100251 [FREE Full text] [doi: 10.1371/journal.pone.0100251] [Medline: 24950218]
125. Vallance JK, Friedenreich CM, Lavallee CM, Culos-Reed N, Mackey JR, Walley B, et al. Exploring the feasibility of a broad-reach physical activity behavior change intervention for women receiving chemotherapy for breast cancer: a randomized trial. Cancer Epidemiol Biomarkers Prev 2015 Dec 16;25(2):391-398. [doi: 10.1158/1055-9965.epi-15-0812]
126. Van Blarigan EL, Chan H, Van Loon K, Kenfield SA, Chan JM, Mitchell E, et al. Self-monitoring and reminder text messages to increase physical activity in colorectal cancer survivors (Smart Pace): a pilot randomized controlled trial. BMC Cancer 2019 Mar 11;19(1):218 [FREE Full text] [doi: 10.1186/s12885-019-5427-5] [Medline: 30866859]
127. Warren FC, Stych K, Thorogood M, Sharp DJ, Murphy M, Turner KM, et al. Evaluation of different recruitment and randomisation methods in a trial of general practitioner-led interventions to increase physical activity: a randomised controlled feasibility study with factorial design. Trials 2014 Apr 21;15:134 [FREE Full text] [doi: 10.1186/1745-6215-15-134] [Medline: 24746263]
128. Yamada M, Mori S, Nishiguchi S, Kajiwara Y, Yoshimura K, Sonoda T, et al. Pedometer-based behavioral change program can improve dependency in sedentary older adults: a randomized controlled trial. J Frailty Aging 2012;1(1):39-44. [doi: 10.14283/jfa.2012.7] [Medline: 27092936]
129. Croteau KA, Richeson NE, Vines SW, Jones DB. Effects of a pedometer-based physical activity program on older adults' mobility-related self-efficacy and physical performance. Act Adapt Aging 2004 Apr 21;28(2):19-33. [doi: 10.1300/j016v28n02_02]
130. Croteau KA, Richeson NE, Farmer BC, Jones DB. Effect of a pedometer-based intervention on daily step counts of community-dwelling older adults. Res Q Exerc Sport 2007 Dec;78(5):401-406. [doi: 10.1080/02701367.2007.10599439] [Medline: 18274211]
131. Aittasalo M, Miilunpalo S, Kukkonen-Harjula K, Pasanen M. A randomized intervention of physical activity promotion and patient self-monitoring in primary health care. Prev Med 2006 Jan;42(1):40-46. [doi: 10.1016/j.ypmed.2005.10.003] [Medline: 16297442]
132. Fischer X, Kreppke J, Zahner L, Gerber M, Faude O, Donath L. Telephone-based coaching and prompting for physical activity: short- and long-term findings of a randomized controlled trial (Movingcall). Int J Environ Res Public Health 2019 Jul 23;16(14):1-16 [FREE Full text] [doi: 10.3390/ijerph16142626] [Medline: 31340528]

133. Golsteijn RH, Bolman C, Volders E, Peels DA, de Vries H, Lechner L. Short-term efficacy of a computer-tailored physical activity intervention for prostate and colorectal cancer patients and survivors: a randomized controlled trial. Int J Behav Nutr Phys Act 2018 Oct 30;15(1):106 [FREE Full text] [doi: 10.1186/s12966-018-0734-9] [Medline: 30376857]
134. Kendzor DE, Allicock M, Businelle MS, Sandon LF, Gabriel KP, Frank SG. Evaluation of a shelter-based diet and physical activity intervention for homeless adults. J Phys Act Health 2017 Feb;14(2):88-97. [doi: 10.1123/jpah.2016-0343] [Medline: 27775471]
135. Keyserling TC, Samuel Hodge CD, Jilcott SB, Johnston LF, Garcia BA, Gizlice Z, et al. Randomized trial of a clinic-based, community-supported, lifestyle intervention to improve physical activity and diet: the North Carolina enhanced WISEWOMAN project. Prev Med 2008 Jun;46(6):499-510. [doi: 10.1016/j.ypmed.2008.02.011] [Medline: 18394692]
136. Kim Y, Lumpkin A, Lochbaum M, Stegemeier S, Kitten K. Promoting physical activity using a wearable activity tracker in college students: a cluster randomized controlled trial. J Sports Sci 2018 Aug;36(16):1889-1896. [doi: 10.1080/02640414.2018.1423886] [Medline: 29318916]
137. King AC, Ahn DK, Oliveira BM, Atienza AA, Castro CM, Gardner CD. Promoting physical activity through hand-held computer technology. Am J Prev Med 2008 Feb;34(2):138-142 [FREE Full text] [doi: 10.1016/j.amepre.2007.09.025] [Medline: 18201644]
138. Koizumi D, Rogers NL, Rogers ME, Islam MM, Kusunoki M, Takeshima N. Efficacy of an accelerometer-guided physical activity intervention in community-dwelling older women. J Phys Act Health 2009 Jul;6(4):467-474. [doi: 10.1123/jpah.6.4.467] [Medline: 19842461]
139. Li LC, Sayre EC, Xie H, Clayton C, Feehan LM. A community-based physical activity counselling program for people with knee osteoarthritis: feasibility and preliminary efficacy of the Track-OA study. JMIR Mhealth Uhealth 2017 Jun 26;5(6):e86 [FREE Full text] [doi: 10.2196/mhealth.7863] [Medline: 28652228]
140. Maxwell-Smith C, Hince D, Cohen PA, Bulsara MK, Boyle T, Platell C, et al. A randomized controlled trial of WATAAP to promote physical activity in colorectal and endometrial cancer survivors. Psychooncology 2019 Jul;28(7):1420-1429. [doi: 10.1002/pon.5090] [Medline: 30980691]
141. Pekmezi D, Ainsworth C, Joseph RP, Williams V, Desmond R, Meneses K, et al. Pilot trial of a home-based physical activity program for African American women. Med Sci Sports Exerc 2017 Dec;49(12):2528-2536 [FREE Full text] [doi: 10.1249/MSS.0000000000001370] [Medline: 28704343]
142. Prestwich A, Perugini M, Hurling R. Can the effects of implementation intentions on exercise be enhanced using text messages? Psychol Health 2009 Jul;24(6):677-687. [doi: 10.1080/08870440802040715] [Medline: 20205020]
143. Samuels TY, Raedeke TD, Mahar MT, Karvinen KH, DuBose KD. A randomized controlled trial of continuous activity, short bouts, and a 10,000 step guideline in inactive adults. Prev Med 2011 Feb;52(2):120-125. [doi: 10.1016/j.ypmed.2010.12.001] [Medline: 21147156]
144. Sharp P, Caperchione C. The effects of a pedometer-based intervention on first-year university students: a randomized control trial. J Am Coll Health 2016;64(8):630-638. [doi: 10.1080/07448481.2016.1217538] [Medline: 27471879]
145. Unick JL, O'Leary KC, Bond DS, Wing RR. Physical activity enhancement to a behavioral weight loss program for severely obese individuals: a preliminary investigation. ISRN Obes 2012 Sep 05;2012:1-4 [FREE Full text] [doi: 10.5402/2012/465158] [Medline: 24379985]
146. Vandelanotte C, Duncan MJ, Maher CA, Schoeppe S, Rebar AL, Power DA, et al. The effectiveness of a web-based computer-tailored physical activity intervention using Fitbit activity trackers: randomized trial. J Med Internet Res 2018 Dec 18;20(12):e11321 [FREE Full text] [doi: 10.2196/11321] [Medline: 30563808]
147. Wijsman CA, Westendorp RG, Verhagen EA, Catt M, Slagboom PE, De Craen AJ, et al. Effects of a web-based intervention on physical activity and metabolism in older adults: randomized controlled trial. J Med Internet Res 2013 Nov 06;15(11):e233 [FREE Full text] [doi: 10.2196/jmir.2843] [Medline: 24195965]
148. Carr LJ, Dunsiger SI, Lewis B, Ciccolo JT, Hartman S, Bock B, et al. Randomized controlled trial testing an internet physical activity intervention for sedentary adults. Health Psychol 2013 Mar;32(3):328-336 [FREE Full text] [doi: 10.1037/a0028962] [Medline: 22823069]
149. Lane A, Murphy N, Bauman A. An effort to 'leverage' the effect of participation in a mass event on physical activity. Health Promot Int 2015 Sep 13;30(3):542-551. [doi: 10.1093/heapro/dat077] [Medline: 24226297]
150. Mailey EL, Wójcicki TR, Motl RW, Hu L, Strauser DR, Collins KD, et al. Internet-delivered physical activity intervention for college students with mental health disorders: a randomized pilot trial. Psychol Health Med 2010 Dec;15(6):646-659. [doi: 10.1080/13548506.2010.498894] [Medline: 21154018]
151. Motl RW, Dlugonski D, Wójcicki TR, McAuley E, Mohr DC. Internet intervention for increasing physical activity in persons with multiple sclerosis. Mult Scler 2011 Jan;17(1):116-128. [doi: 10.1177/1352458510383148] [Medline: 20921239]
152. Sugden JA, Sniehotta FF, Donnan PT, Boyle P, Johnston DW, McMurdo ME. The feasibility of using pedometers and brief advice to increase activity in sedentary older women-a pilot study. BMC Health Serv Res 2008 Aug 08;8:169 [FREE Full text] [doi: 10.1186/1472-6963-8-169] [Medline: 18691392]
153. Thorsteinsen K, Vittersø J, Svendsen GB. Increasing physical activity efficiently: an experimental pilot study of a website and mobile phone intervention. Int J Telemed Appl 2014;2014 [FREE Full text] [doi: 10.1155/2014/746232] [Medline: 24963290]

154. Schwerdtfeger AR, Schmitz C, Warken M. Using text messages to bridge the intention-behavior gap? A pilot study on the use of text message reminders to increase objectively assessed physical activity in daily life. Front Psychol 2012;3:270 [FREE Full text] [doi: 10.3389/fpsyg.2012.00270] [Medline: 22876237]
155. Bennett JA, Young HM, Nail LM, Winters-Stone K, Hanson G. A telephone-only motivational intervention to increase physical activity in rural adults: a randomized controlled trial. Nurs Res 2008;57(1):24-32. [doi: 10.1097/01.NNR.0000280661.34502.c1] [Medline: 18091289]
156. Glasgow RE, Kurz D, King D, Dickman JM, Faber AJ, Halterman E, et al. Twelve-month outcomes of an internet-based diabetes self-management support program. Patient Educ Couns 2012 Apr;87(1):81-92 [FREE Full text] [doi: 10.1016/j.pec.2011.07.024] [Medline: 21924576]
157. McNab F, Hillebrand A, Swithenby SJ, Rippon G. Combining temporal and spectral information with spatial mapping to identify differences between phonological and semantic networks: a magnetoencephalographic approach. Front Psychol 2012;3:273 [FREE Full text] [doi: 10.3389/fpsyg.2012.00273] [Medline: 22908001]
158. Sisson SB, Camhi SM, Church TS, Tudor-Locke C, Johnson WD, Katzmarzyk PT. Accelerometer-determined steps/day and metabolic syndrome. Am J Prev Med 2010 Jun;38(6):575-582. [doi: 10.1016/j.amepre.2010.02.015] [Medline: 20494233]
159. Oftedal S, Holliday EG, Attia J, Brown WJ, Collins CE, Ewald B, et al. Daily steps and diet, but not sleep, are related to mortality in older Australians. J Sci Med Sport 2020 Mar;23(3):276-282. [doi: 10.1016/j.jsams.2019.09.018] [Medline: 31615727]
160. Chan CB, Ryan DA, Tudor-Locke C. Health benefits of a pedometer-based physical activity intervention in sedentary workers. Prev Med 2004 Dec;39(6):1215-1222. [doi: 10.1016/j.ypmed.2004.04.053] [Medline: 15539058]
161. Althoff T, Sosič R, Hicks JL, King AC, Delp SL, Leskovec J. Large-scale physical activity data reveal worldwide activity inequality. Nature 2017 Jul 20;547(7663):336-339 [FREE Full text] [doi: 10.1038/nature23018] [Medline: 28693034]
162. Shcherbina A, Hershman SG, Lazzeroni L, King AC, O'Sullivan JW, Hekler E, et al. The effect of digital physical activity interventions on daily step count: a randomised controlled crossover substudy of the MyHeart Counts Cardiovascular Health Study. Lancet Digit Health 2019 Nov;1(7):344-352. [doi: 10.1016/s2589-7500(19)30129-3] [Medline: 33323209]
163. Kramer J, Künzler F, Mishra V, Smith SN, Kotz D, Scholz U, et al. Which components of a smartphone walking app help users to reach personalized step goals? Results from an optimization trial. Ann Behav Med 2020 Jun 12;54(7):518-528 [FREE Full text] [doi: 10.1093/abm/kaaa002] [Medline: 32182353]
164. Cook R, Davidson P, Martin R, NIHR Dissemination Centre. Pedometers can help people get more active as part of an exercise programme. Br Med J 2020 Apr 06;369. [doi: 10.1136/bmj.m877] [Medline: 32253188]
165. McAuley L, Pham B, Tugwell P, Moher D. Does the inclusion of grey literature influence estimates of intervention effectiveness reported in meta-analyses? Lancet 2000 Oct 07;356(9237):1228-1231. [doi: 10.1016/S0140-6736(00)02786-0] [Medline: 11072941]

Abbreviations

EE: energy expenditure
GRADE: grading of recommendations, assessment, development, and evaluation
MetS: metabolic syndrome
mHealth: mobile health
MVPA: moderate-to-vigorous physical activity
NCD: noncommunicable disease
PA: physical activity
RCT: randomized controlled trial
SMD: standardized mean difference
TPA: total physical activity
WHO: World Health Organization

Edited by G Eysenbach; submitted 22.12.20; peer-reviewed by P Hendrick, H Zihao; comments to author 15.01.21; revised version received 24.02.21; accepted 02.04.21; published 30.04.21.

Please cite as:
Mönninghoff A, Kramer JN, Hess AJ, Ismailova K, Teepe GW, Tudor Car L, Müller-Riemenschneider F, Kowatsch T
Long-term Effectiveness of mHealth Physical Activity Interventions: Systematic Review and Meta-analysis of Randomized Controlled Trials
J Med Internet Res 2021;23(4):e26699
URL: https://www.jmir.org/2021/4/e26699
doi:10.2196/26699
PMID:33811021

Viewpoint

Prescribing Phones to Address Health Equity Needs in the COVID-19 Era: The PHONE-CONNECT Program

Gill Kazevman[1], BHS; Marck Mercado[1], BSc; Jennifer Hulme[2,3], MDCM, MPH; Andrea Somers[2,3], MD, MPH

[1]Faculty of Medicine, University of Toronto, Toronto, ON, Canada

[2]Department of Emergency Medicine, University Health Network, Toronto, ON, Canada

[3]Department of Family and Community Medicine, University of Toronto, Toronto, ON, Canada

Corresponding Author:
Andrea Somers, MD, MPH
Department of Emergency Medicine
University Health Network
200 Elizabeth St
RFE-GS-480
Toronto, ON, M5G 2C4
Canada
Phone: 1 416 340 3856
Email: Andrea.Somers@uhn.ca

Abstract

Vulnerable populations have been identified as having higher infection rates and poorer COVID-19–related outcomes, likely due to their inability to readily access primary care, follow public health directives, and adhere to self-isolation guidelines. As a response to the COVID-19 pandemic, many health care services have adopted new digital solutions, which rely on phone and internet connectivity. However, persons who are digitally inaccessible, such as those experiencing poverty or homelessness, are often unable to use these services. In response to this newly highlighted social disparity known as "digital health inequity," emergency physicians at the University Health Network in Toronto, Canada, initiated a program called PHONE-CONNECT (Phones for Healthier Ontarians iN EDs – COvid NEeds met by Cellular Telephone). This novel approach attempts to improve patients' access to health care, information, and social services, as well as improve their ability to adhere to public health directives (social isolation and contact tracing). Although similar programs addressing the same emerging issues have been recently described in the media, this is the first time phones have been provided as a health care intervention in an emergency department. This innovative emergency department point-of-care intervention may have a significant impact on improving health outcomes for vulnerable people during the COVID-19 pandemic and beyond.

(J Med Internet Res 2021;23(4):e23914) doi:10.2196/23914

KEYWORDS

digital health equity; health inequity; digital determinants of health; emergency medicine; COVID-19; public health; health policy; primary care; cell phone

Introduction

The COVID-19 pandemic has highlighted many social disparities encountered by vulnerable populations [1]. Vulnerable populations, such as those experiencing poverty or homelessness, have a higher rate of infection and significantly worse COVID-19–related health outcomes [2]. They have been disproportionately affected during the pandemic by a lack of safe housing for physical distancing or self-isolation, difficulty adhering to public health directives, and significant barriers to accessing health care and social services [3].

Emerging digital solutions adopted by health care providers in response to the pandemic rely heavily on phone connectivity and access to the internet [4]. These include telemedicine, telepsychiatry, and COVID-19–related communications such as testing status updates and contact tracing. Most recently, the Canadian federal government launched a new mobile app for efficient and secure contact tracing, named COVID Alert [5]. Although these new solutions provide exciting opportunities for the future of health care delivery and public health, they may create yet another barrier for those who are unable to afford a phone and phone plan [6]. The inability to access digital

services constitutes a newly explored digital determinant of health: digital health equity [7].

A recent study suggests that targeted access to mobile devices can help mitigate the downstream consequences of this health disparity [8]. Given the urgent needs presented during the COVID-19 pandemic, the emergency department (ED) at the University Health Network (UHN), in Toronto, Ontario, responded with a pilot program to provide free prepaid cell phones to vulnerable patients.

What is PHONE-CONNECT?

Phones for Healthier Ontarians iN EDs – COvid NEeds met by Cellular Telephone, also known as PHONE-CONNECT, is a program founded by UHN emergency physicians to address the health and social needs of vulnerable patients through cellular and digital connectivity. This program was conceived after recognizing that many patients facing digital inequity are at risk for further morbidity despite adequate treatment received in the ED. These risks are compounded by the COVID-19 pandemic wherein contact tracing, physical distancing, and virtual care are heavily reliant on having access to a digital device. The program repurposes donated cell phones and distributes them to eligible patients discharged from the ED as an intervention to address digital health inequity. The aims of the program are to identify people experiencing digital health inequities and address these inequities by improving patients' access to health care services, information, and social services, as well as to improve social connectivity. Additionally, the program seeks to increase patients' capacity to follow public health directives (including self-isolation) and their availability for contact tracing during the pandemic.

To the best of our knowledge, this is the first time cell phones have been distributed from EDs to address the needs of vulnerable patients, and the first time cell phones have been distributed from an ED during a pandemic. Since the program's inception at UHN, PHONE-CONNECT has expanded to include additional Greater Toronto Area sites (ie, St. Michael's Hospital and Michael Garron Hospital) and academic medical centers (ie, McMaster University).

Implementation

The PHONE-CONNECT program was initiated with the donation of 50 smartphones and 200 SIM cards with unlimited talk and text features from a telecommunications company. Additional phone donations were sought by word of mouth and via media coverage of the initiative. Members of the public were invited to mail in their pre-used cell phones. Donated phones were checked for operability and compatibility with the donated SIM cards, cleaned of the previous user's personal information and disinfected. Phone distribution packages (Figure 1) include a cellular device loaded with a SIM card, a compatible phone charger, and a phone manual (when available). A handout explaining the importance of using the phone to connect with health care providers was included, as well as a list of phone numbers for various health and social services.

Figure 1. Phone distribution packages.

Population

Patients are deemed eligible for a phone distribution package if a phone number is not listed on their ED chart. This applies to all patients who do not have access to a telephone and require communication of test results and pending appointments, those who require a phone to facilitate self-isolation, or those who need a device to access follow-up care, virtual or in-person. Health care providers who were trained to administer the phones include physicians, nurse practitioners, physician assistants, registered nurses, social workers, and peer support workers. These providers were informed of the PHONE-CONNECT initiative through daily department huddles, team meetings, and emails. They were given instructions on patient eligibility, administration of phones to patients, and documentation of the patient's new phone number. The same health care providers were responsible for identifying eligible patients during the course of regular clinical care, orienting the patients to their new phone prior to discharge from the ED, and advising them that they would be contacted for consent for a follow-up interview. Once distributed, the new phone number is added to the patient's electronic medical record, and the patient is encouraged to keep the phone turned on, charged, and available should a health care provider need to contact them. Aside from these requests, the patient may use the phone however they choose.

Data Collection

The PHONE-CONNECT pilot project is being assessed through an ongoing research study exploring the impact of the phone on the patient's well-being and their utilization of health care services, as well as the perspectives of the health care providers participating in the PHONE-CONNECT initiative.

Specifically, 2-4 weeks following discharge, research assistants attempt to contact phone recipients through their new device with a maximum of three calls and one SMS text message. A semistructured interview (Multimedia Appendix 1) is then conducted with the phone recipient to explore the role of their new phone in addressing digital health equity. This includes the following: the role of the phone in their ability to self-isolate in the context of COVID-19, their communication with health care providers, their ability to maintain their social networks, and the recipients' self-reported well-being. Quantitative data collected include the number of phones distributed, the number of participants attending virtual and in-person follow-up appointments, nonattendance to prescheduled appointments, as well as the number of participants tested for COVID-19 and who received their results via the phone intervention.

The impact of PHONE-CONNECT on the distributing health care providers is also being explored. The health care providers are asked to complete a questionnaire using 5-point Likert scale questions assessing their perceptions on how giving a phone in the ED affects their ability to provide care to socially complex patients. Additional questions assess the impact of phone distribution on provider burnout.

Review of Similar Initiatives

To the best of our knowledge, prior to the COVID-19 pandemic only one program offered free cellular phones for vulnerable populations. In the United States, free cellular phones are offered to a subset of participants in the Federal Lifeline program [9]. A study of people experiencing homelessness in New York City who received free phones through this program found that beneficiaries viewed these free devices as an important tool to help navigate, engage, and ultimately improve their access to care [10]. More recently, and as a response to the digital disparities experienced by the homeless population during the COVID-19 pandemic, two new programs have been described in the media. In the United Kingdom, a program administered by a charity and a telecommunication company was designed to provide 2500 cell phones to eligible recipients living in England, Scotland, and Wales [11]. More locally, the Social Planning and Research Council (SPARC) of British Columbia distributed 3500 smartphones to persons experiencing homelessness at parks and shelters [12].

Impact

Thus far, 180 devices have been distributed to eligible patients at two sites in Toronto. The personal narratives emerging from our nascent evaluation are compelling and will be highlighted in the publication of a future study. Recipients have used their phones to address their health care needs by connecting with physicians, addictions counsellors, social workers, medical specialists, and Toronto Public Health representatives, and to attend telemedicine appointments. They have accessed suicide hotlines and emergency medical services. They have used their phones to meet their basic needs, by accessing groceries during isolation and securing shelter beds. Furthermore, recipients have strengthened their social networks by virtue of being in regular contact with friends and family.

Limitations

Despite the reported benefits of providing mobile phones to those experiencing homelessness [10], there are several issues that should be addressed as this program finds greater adoption. Not all patients are inherently "tech savvy" and may incur additional stress and anxiety from using a mobile phone for health-related communication. Appropriate matching of user to device by technical confidence level and familiarity should be considered in the initiation of similar programs. Furthermore, "phone-seeking behavior" could be an unfortunate side effect of this program that may increase ED visits, although this has not been our experience.

Finally, as this is an ED point-of-care intervention, clinical demands on ED providers may increase. Empowering other allied health professionals in the ED to provide phones has addressed this concern. Moreover, our experience suggests that any increased clinical burden may be offset by better ease of ED discharge planning for patients now able to access care beyond the ED.

Discussion

The need for phone connectivity will increase as the COVID-19 pandemic continues to impact access to health services. We are concerned that if digital equity needs remain unaddressed, vulnerable patients will continue to experience disparate health outcomes of increasing magnitude during the pandemic and beyond. Although further research into the different aspects of digital health equity is necessary for the creation of universal policies and strategies, this innovative initiative is an achievable first step in overcoming the challenge of accessibility. This project may demonstrate that identification of those who are facing this equity issue can be accomplished in the primary care setting. Moreover, a recent study found that improving primary care access via phone connectivity can reduce avoidable ED visits [13]. Therefore, by facilitating access to outpatient health care and social services, the PHONE-CONNECT program may be an innovative solution to reduce unnecessary ED visits or prolonged ED length of stays.

As we continue to pilot this novel ED point-of-care intervention, we hope to further explore the effects of our program on its recipients. We are particularly interested in understanding the effects of this program on access to and utilization of the health care system, social services, COVID-19–related public health interventions, as well as the impact on overall well-being. We are also interested in determining whether the ED is an appropriate setting for addressing these digital equity issues. Answers to such questions will be elucidated in the currently ongoing associated study. We hope to continue to address digital health inequities by expanding our program to more sites, and by ensuring the ongoing activation of SIM cards currently in circulation.

Acknowledgments

The authors thank Lanujan Kaneswaran and Alexandra McKnight for providing logistical support for the PHONE-CONNECT program. The authors would also like to acknowledge Humaira Nakhuda and Madeline McDonald for their contributions in data collection and patient interviews.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PHONE-CONNECT follow-up interview guide.
[DOCX File , 13 KB - jmir_v23i4e23914_app1.docx]

References

1. Dorn AV, Cooney RE, Sabin ML. COVID-19 exacerbating inequalities in the US. Lancet 2020 Apr 18;395(10232):1243-1244 [FREE Full text] [doi: 10.1016/S0140-6736(20)30893-X] [Medline: 32305087]
2. Ontario Agency for Health Protection and Promotion (Public Health Ontario). COVID-19 - What we know so far about... social determinants of health. Toronto, ON: Queen's Printer for Ontario; 2020 May 24. URL: https://www.publichealthontario.ca/-/media/documents/ncov/covid-wwksf/2020/05/what-we-know-social-determinants-health.pdf?la=en [accessed 2021-03-31]
3. Perri M, Dosani N, Hwang SW. COVID-19 and people experiencing homelessness: challenges and mitigation strategies. CMAJ 2020 Jun 29;192(26):E716-E719 [FREE Full text] [doi: 10.1503/cmaj.200834] [Medline: 32601252]
4. Keesara S, Jonas A, Schulman K. Covid-19 and Health Care's Digital Revolution. N Engl J Med 2020 Jun 04;382(23):e82. [doi: 10.1056/nejmp2005835]
5. Prime Minister of Canada. New mobile app to help notify Canadians of potential COVID-19 exposure now available. 2020 Jul 31. URL: https://pm.gc.ca/en/news/news-releases/2020/07/31/new-mobile-app-help-notify-canadians-potential-covid-19-exposure-now [accessed 2020-08-12]
6. Lee C, Rogers WA, Braunack-Mayer A. Social Justice and Pandemic Influenza Planning: The Role of Communication Strategies. Public Health Ethics 2008 Nov 01;1(3):223-234. [doi: 10.1093/phe/phn031]
7. Crawford A, Serhal E. Digital Health Equity and COVID-19: The Innovation Curve Cannot Reinforce the Social Gradient of Health. J Med Internet Res 2020 Jun 02;22(6):e19361 [FREE Full text] [doi: 10.2196/19361] [Medline: 32452816]
8. Beaunoyer E, Dupéré S, Guitton MJ. COVID-19 and digital inequalities: Reciprocal impacts and mitigation strategies. Comput Human Behav 2020 Oct;111:106424 [FREE Full text] [doi: 10.1016/j.chb.2020.106424] [Medline: 32398890]
9. Federal Communications Commission. Lifeline Program for Low-Income Consumers. URL: https://www.fcc.gov/general/lifeline-program-low-income-consumers [accessed 2020-08-12]
10. Asgary R, Sckell B, Alcabes A, Naderi R, Adongo P, Ogedegbe G. Perceptions, Attitudes, and Experience Regarding mHealth Among Homeless Persons in New York City Shelters. J Health Commun 2015 Aug 27;20(12):1473-1480 [FREE Full text] [doi: 10.1080/10810730.2015.1033117] [Medline: 26313765]

11. Homeless people in Britain to receive free phones in connectivity drive. The Guardian. 2020 Aug 18. URL: https://www.theguardian.com/society/2020/aug/18/homeless-people-uk-receive-free-phones-connectivity-drive-covid-19 [accessed 2020-08-25]
12. Smartphones to better connect vulnerable. BC Gov News. 2020 May 12. URL: https://news.gov.bc.ca/releases/2020SDPR0028-000858 [accessed 2020-08-12]
13. Tsai M, Xirasagar S, Carroll S, Bryan CS, Gallagher PJ, Davis K, et al. Reducing High-Users' Visits to the Emergency Department by a Primary Care Intervention for the Uninsured: A Retrospective Study. Inquiry 2018 Mar 28;55:46958018763917 [FREE Full text] [doi: 10.1177/0046958018763917] [Medline: 29591539]

Abbreviations

ED: emergency department
PHONE-CONNECT: Phones for Healthier Ontarians iN EDs – COvid NEeds met by Cellular Telephone
UHN: University Health Network

Edited by G Eysenbach, C Basch; submitted 27.08.20; peer-reviewed by E Serhal, E Schneider; comments to author 18.11.20; revised version received 08.12.20; accepted 10.03.21; published 06.04.21.

Please cite as:
Kazevman G, Mercado M, Hulme J, Somers A
Prescribing Phones to Address Health Equity Needs in the COVID-19 Era: The PHONE-CONNECT Program
J Med Internet Res 2021;23(4):e23914
URL: https://www.jmir.org/2021/4/e23914
doi:10.2196/23914
PMID:33760753

Viewpoint

Addressing the Digital Inverse Care Law in the Time of COVID-19: Potential for Digital Technology to Exacerbate or Mitigate Health Inequalities

Alisha R Davies[1*], MSc, PhD, FFPH; Matthew Honeyman[2*], BA, MSc; Bob Gann[3*], BA, DipLib

[1]Research and Evaluation Division, Public Health Wales, Cardiff, United Kingdom

[2]Kings Fund, London, United Kingdom

[3]Digital Communities Wales, Cardiff, United Kingdom

[*]all authors contributed equally

Corresponding Author:
Alisha R Davies, MSc, PhD, FFPH
Research and Evaluation Division
Public Health Wales
2 Capital Quarter, Tyndall St
Cardiff, CF10 4BZ
United Kingdom
Phone: 44 2920227744
Email: alisha.davies@wales.nhs.uk

Abstract

Digital technologies have been transforming methods of health care delivery and have been embraced within the health, social, and public response to the COVID-19 pandemic. However, this has directed attention to the "inverse information law" (also called "digital inverse care law") and digital inequalities, as people who are most in need of support (in particular, older people and those experiencing social deprivation) are often least likely to engage with digital platforms. The response to the COVID-19 pandemic represents a sustained shift to the adoption of digital approaches to working and engaging with populations, which will continue beyond the COVID-19 pandemic. Therefore, it is important to understand the underlying factors contributing to digital inequalities and act immediately to avoid digital inequality contributing to health inequalities in the future. The response to COVID-19 represents a sustained shift to adopting digital approaches to working and engaging with populations which will continue beyond this pandemic. Therefore it is important that we understand the underlying factors contributing to digital inequalities, and act now to protect against digital inequality contributing to health inequalities in the future.

(J Med Internet Res 2021;23(4):e21726) doi:10.2196/21726

KEYWORDS

COVID-19; digital divide; digital exclusion; digital health; health inequality; population health

Introduction

Background

Technology is transforming the way we live our lives. A digital revolution is underway in the public sector, including the health sector, to facilitate the use of digital solutions to drive efficient systems and improve population health both in the United Kingdom [1-4] and worldwide [5]. The expected benefits to individuals include more rapid access to information and personalized care, more control, and empowerment for their own health. The benefits to health and care systems include opportunities to deliver efficient and effective care closer to patients and target scarce resources in a better manner, using precision medicines, big data, and artificial intelligence.

However, while the application of digital innovation in health care—hereinafter referred to as "digital health"—has gained impetus, the wider social system has received limited consideration, and people most in need of care are also least likely to have access to, or engage with, technology. A failure to acknowledge and address this challenge would imply that the adoption of digital health has the potential to inadvertently widen health inequalities, thus integrating the inverse care law [6] into the digital era as a "digital inverse care law", previously called "inverse information law" [7].

The COVID-19 pandemic has markedly highlighted the vital role of access to the internet and digital technology in enabling the general public to live their everyday lives during the pandemic. The pace of digital innovation and its adoption in health and care among the general public has accelerated. This includes widespread rapid adoption of internet-based health consultations and the development of digital health tools and apps to protect health and well-being. There has also been a demonstrable increase in the proportion of people in the general population who use the internet to search for information; maintain contact with others; support the continuation of work, study, or home-schooling during the COVID-19 pandemic; and access basic needs including shopping and financial support [8,9]. During a period of enforced social isolation, a focus on internet-enabled social responses has implied that people without a presence on the internet are effectively excluded. As we progress towards recovery from the COVID-19 pandemic, there is a risk of leaving behind people who do not engage with digital technologies and are most likely in need of support.

Increasing awareness and improving our understanding of the factors contributing to the digital inverse care law, and how these challenges can be addressed, is of considerable importance in the application of digital health [10,11]. We can expect the increased reliance on digital technologies during the COVID-19 pandemic to be sustained in many ways even after the COVID-19 pandemic; hence, it is important to recognize these issues immediately.

Objective

To help inform action, here we describe the factors contributing to digital exclusion, its contribution to health and social inequalities, and the potential factors that need to be addressed to prevent it. This viewpoint has been informed by the academic and gray literature on technological advancements in the health sector and seeks to highlight the complexity of the drivers of the digital inverse care law and the actions needed across government, public, private, and nonprofit sector organizations to ensure capitalizing on the potential for digital technologies to address health issues and minimize the risk of the exacerbation of health inequalities.

Factors Contributing to Digital Exclusion

We describe digital exclusion as a complex challenge that consists of 3 interconnected components, with inequalities evident in each component: (1) access to digital connectivity and infrastructure, (2) digital skills and literacy, and (3) engagement with digital platforms. Here we consider each of these elements, illustrated within the context of the key opportunities for digital technologies in the National Health Service (NHS) in the United Kingdom [11], with particular reference to the COVID-19 pandemic.

Access to Digital Connectivity and Infrastructure

People worldwide use the internet to stay in touch with others and live their lives during the COVID-19 pandemic. For example, a survey by the Pew Research Center in April 2020 [12] revealed that 53% of people in the United States felt that the internet is essential during the COVID-19 pandemic, and a further 34% felt that it is important but not essential. Nonetheless, only 55% of households worldwide have an internet connection, ranging from 87% of households in high-income countries, 47% in transitional countries, and only 19% in the low-income countries [13].

Even in high-income countries, marked inequalities in internet access are evident. In 2019, 7% of households in the United Kingdom did not use the internet [14], and 10% of adults in the United Kingdom did not use the internet regularly [15]. The same proportion has been reported in the United States, increasing with age to 27% among people aged ≥65 years [16]. A population-wide survey across 17 European countries revealed that 51% of people aged ≥50 years do not use the internet [17].

Internet access can be facilitated by enabling access to internet-enabled technologies alongside a fixed (eg, household) or mobile broadband connection. While internet access through a fixed broadband connection has increased in recent years, the quality and speed of the connection remains poor for many people, and this is particularly the case in rural areas. Data from the European Union suggest that the increase in very-high-capacity network coverage in rural areas remains significantly lower than the total coverage, despite marked improvements in recent years (from 2011 to 2019, very-high-capacity network coverage increased from 2% to 20% in rural areas, compared to an increase from 10% to 44% overall) [18]. In the United Kingdom, residential premises in rural areas have lower coverage of fixed superfast broadband connections (79% of properties in rural areas compared to 97% in urban areas), download speeds exceeding 10 Mbps (83% in rural areas compared to 98% in urban areas), and lower access to high-quality mobile data (4G data services; 41% in rural areas compared to 85% in urban areas) [19]. Regional disparities in broadband access are not a problem exclusive to rural areas. The National Digital Inclusion Alliance's report on the worst connected cities revealed that among 221 large- and medium-sized US cities, at least 30% of households lacked a broadband connection [20].

A lack of rapid, reliable connectivity across fixed and mobile internet services can be a challenge for patients to receive remote care and for health care staff who are mobile and work in patients' homes. In a survey among the members of the Queen's Nursing Institute (n=534) [21], 454 (85%) nursing professionals reported that poor connectivity in patients' homes is the greatest challenge to effective mobile working in the community.

Alongside infrastructure, another factor contributing to internet access is the affordability of internet-enabled digital devices and data plans. Over 25,000,000 mobile phone users in the United Kingdom are pay-as-you-go customers, with the majority of users having a low income. Community organizations have reported examples of vulnerable groups spending up to half the family budget on incurring mobile phone costs [22]. Described as "data poverty," accessing the internet through mobile digital technologies can be an unaffordable essential need and individuals may be reluctant to devote scare data resources to digital health in light of competing demands [23].

The COVID-19 pandemic response has exposed digital poverty with the lack of ownership of digital devices and low

affordability of data plans [24]. In addressing these challenges, the UK Government supported the DevicesDotNow initiative, which asked businesses to donate devices (tablets, smartphones, and laptops) and connectivity (in the form of SIM cards, dongles, and mobile hotspots) to be distributed to households who would otherwise be digitally excluded [25]. In Wales, there has been a rapid roll out of the Attend Anywhere video consultation service, accompanied by the supply of 1000 tablet devices to hospitals, care homes, and hospice settings, to enable vulnerable people to access the service on the internet [26]. In addressing data affordability, telecommunication companies in the United Kingdom have removed the data cap for fixed broadband contracts during the COVID-19 pandemic [27]. However, this did not apply to pay-as-you-go mobile contract holders; therefore, those likely to be in greater need were not able to benefit from this initiative.

Evaluation of the impact of such initiatives would help understand the extent to which such programs reach people who are most in need and address the access and affordability drivers of digital inequalities across all groups.

Digital Skills and Literacy

In 2020, a survey in the United Kingdom revealed that 10,500,000 people (16% of the adult population of the United Kingdom) cannot perform basic activities with digital devices, such as turning on a device, connecting to the Wi-Fi, or opening an app by themselves. In total, 7% of the population of the United Kingdom (3,600,000 people) is almost completely offline [28]. Data from the Digital Economy and Society Index of the European Commission suggest that while the level of digital skills continues to increase across many countries in recent years, progress among different population groups is highly variable. In 2019, 82% of young people (aged 16-24 years) and 85% of those with high formal education have at least basic digital skills, compared to only 35% of people aged 55-74 years [29].

There are a number of specific programs in the United Kingdom [30-32] and worldwide [33,34], which seek to address gaps in digital skills among specific groups of people. Structured programs, such as those provided by the Good Things Foundation [35] and Digital Communities Wales [36], focus on overcoming the barriers to opportunity, access, knowledge, and skills for using technology (particularly the internet). During the COVID-19 pandemic, web-based digital skills programs for the general public, such as Learn My Way [37], have been made available free of charge, thus eliminating financial barriers to internet access. Many other initiatives supporting the best practices in digital skills development across Europe are highlighted in the European Commission annual Digital Skills Awards [38], including, for example, community navigators to help older people living with long-term conditions to access and use technologies [39]. Digital skills programs tend to be focused on the general public; however, there is also the need to develop the digital skills among the NHS workforce, such that they can support patients to engage with a digital health care system and direct them to effective digital solutions. In a study on effective mobile working in the community, 21% of nurses reported that limited or no training for the use of the devices was a key challenge to implementation [21]. Examples of nationwide programs supporting digital skills training among health professionals include the Digital Readiness program of Health Education England [40].

Globally, digital literacy and continuous skill development as technologies evolve has been identified as an important driver of health technology use [5]. Ensuring everyone is equipped with the digital skills needed to effectively engage with digital technology is crucial to address digital exclusion.

Engagement With Digital Platforms

It is important to address internet access and digital skill gaps, but these factors alone are unlikely to be sufficient to ensure all patients have the potential to benefit from services delivered on digital platforms. Many other factors potentially influence an individual's choice to engage with digital platforms; these include levels of awareness, trust, and perceived benefits, the combination of which would differ across age groups, genders, and socioeconomic and cultural backgrounds.

As evidenced by the latest issue of the Consumer Digital Index of the United Kingdom [28], a lack of motivation or interest is one of the key barriers to internet engagement. Over one-third of internet nonusers claim that the internet does not interest them, and 48% claim that "nothing" could motivate them to become internet users. Alongside motivation, there is also a need to consider that different groups of people prefer to engage with different digital platforms. A survey conducted in the United Kingdom reported that although over 50% of adults were willing to have web-based consultations with general practitioners, approximately 25% of people aged over 65 years and 40%-45% of people from households earning less than £25,000 would not opt for a web-based video consultation with their general practitioner [11]. The COVID-19 pandemic has forced many primary care consultations to be carried out remotely through the telephone and web-based or video platforms, leading to widespread implementation of these forms of consultation as practical solutions [41]. Some studies have suggested high satisfaction among people who engage with online care [42]; nonetheless, few studies have evaluated the levels of engagement across different populations, the underlying contributing factors, and the impact on timely access to health information, care, patient experience, and health outcomes. More comprehensive evaluation is needed to understand and inform the development of digital or mixed models of service in a better manner to ensure that digital innovation does not inadvertently contribute to the exacerbation of inequalities in outcomes.

Coproduction is essential to address disparities in the engagement with digitally delivered health care among diverse populations. Tools such as the "Culturally-Informed Design Framework" in the United States are useful guides to encourage developers and providers to consider cultural differences in engagement to optimize the choice of digital platform, functionality, content, and user interface [43].

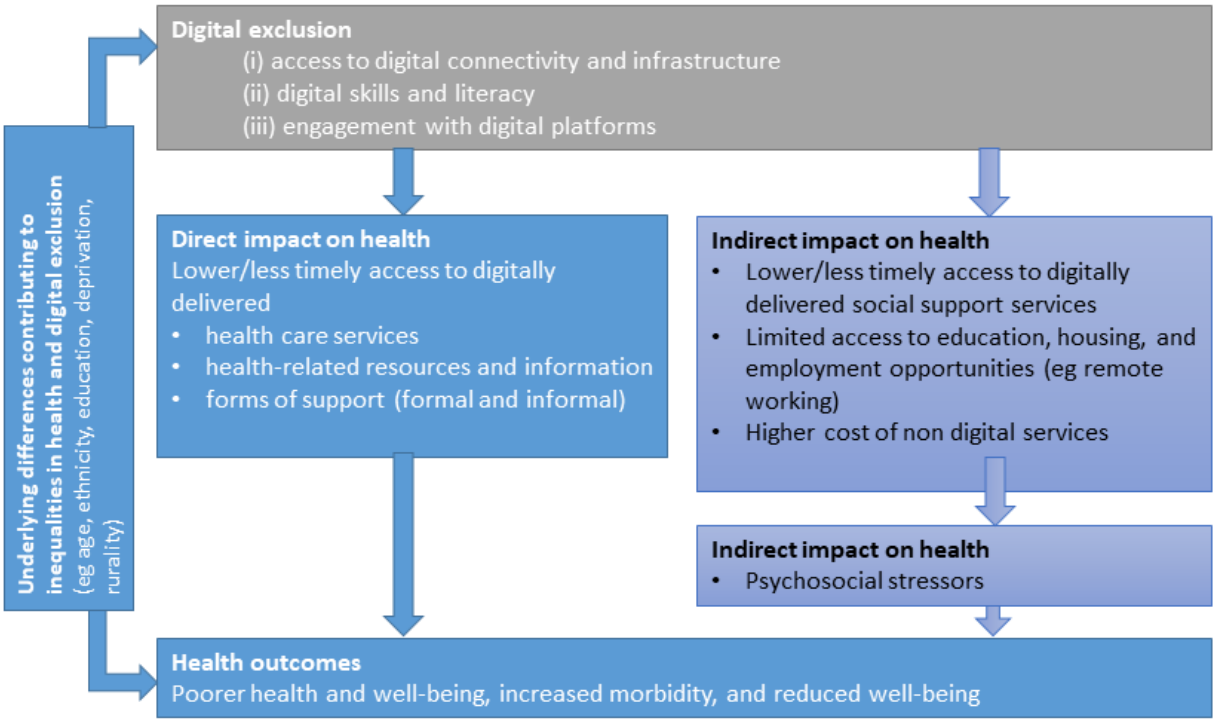
How Digital Exclusion May Exacerbate Health and Social Inequalities

The digital era has transformed and continues to transform society rapidly, influencing our access to information, employment, working conditions, health and care services, and social connections, and in turn the conditions for good health. However, alongside advancements in digital innovation, digital exclusion is increasingly being recognized as an important factor potentially contributing to both health and social inequalities [44].

The evidence summarized here describes how digital exclusion is a complex factor that reflects underlying social and economic inequalities (such as economic barriers to accessing data and devices or variations in regional internet infrastructure) and contributes to social, economic, and health inequalities (as those less able or likely to engage with digital technology are also most likely to need support).

The digital inverse care law [7,45] reflects the direct impact of digital exclusion on health, for example by reducing an individual's access to timely and reliable health information, services, and support (eg, from professionals or peer support groups) delivered on digital platforms (Figure 1). A focus on digitally delivered health systems, without due consideration of digital exclusion, has the potential to prevent "intervention generated inequalities" from reinforcing the underlying inequalities in health [46,47]. In a nationwide representative study of engagement with digital technology for health purposes in Wales, the use of digital technology was lower among groups of people who are likely to have greater health needs, including older people, those living in less affluent areas, those with poorer underlying health, and those reporting health-harming behaviors (ie, smoking, drinking, and physical inactivity) [48]. This has also been highlighted within the context of COVID-19, a period when having an internet presence is crucial to rapidly access not only health information but also digital health consultations and health monitoring apps. Furthermore, people who are at the highest risk of poor health outcomes are also those most likely to be digitally excluded, including older people and those living in more deprived areas [49].

Figure 1. The direct and indirect impacts of digital exclusion on health inequalities. Adapted from McAuley [45].



Digital exclusion can also have an indirect impact on health outcomes, acting through the other social determinants of health [44]. For example, digital exclusion may reduce an individual's access to social support services (eg, income support), education, and employment opportunities, all of which are recognized as underlying social and economic determinants of health (Figure 1). Within the context of the COVID-19 pandemic, during the first 2 weeks of the nationwide lockdown in the United Kingdom, approximately 1,000,000 people sought financial support through universal credit; however, this was only possible on the internet, thus posing a challenge to those who are digitally excluded. When schools, crèches, universities, offices, churches, shops, restaurants, and parks are closed owing to the COVID-19 pandemic, the usual institutions around which people structure their everyday lives and gain support were no longer physically accessible. Many people transitioned to web-based platforms, which was a supportive community inaccessible to digitally excluded people [50].

Digital exclusion is concurrent with Dahlgren and Whitehead's definition of a social determinant of health [51], as a social and economic factor with the potential to increase or decrease social inequities in health. Recognition of digital exclusion as a social determinant of health by governments and systems would increase the visibility of the issue, highlight the need for routine

measurement, report variations across population groups, and focus action on the factors contributing to digital exclusion across policy, health, and social systems. There are examples where health systems display different behaviors in light of increased awareness of digital exclusion, such as addressing data poverty by enabling mobile phone users toll-free access to the NHS and national government health websites [52]. Digitally excluded people would have been hugely impacted without access to rapidly changing and reliable public health advice during the COVID-19 pandemic.

Digital Innovation, COVID-19, and Health Inequalities: an Example

In response to the COVID-19 pandemic, contact tracing digital apps were rapidly implemented in several countries [53] including the United Kingdom [54]. Engagement with the app is dependent on the individual owning a smartphone, having the skills to understand and download the app, and keeping their phone switched on with Bluetooth enabled. However, 6,500,000 adults in the United Kingdom cannot turn on a device, and 5,900,000 adults cannot open an app [28]. A survey carried out by IPSOS Mori for the Health Foundation [55] revealed a clear digital divide by age, occupation, and educational level in public readiness to download the NHS COVID-19 app. Public trust and confidence were also likely to be highly instrumental in adoption. One's motivation to download and use the NHS COVID-19 app would reflect one's views on risk and benefit within the context of a pandemic and, potentially influenced by the views of one's family, social networks, and professionals.

Contact tracing apps are likely to be disproportionately taken up by younger, more affluent, and tech savvy populations. The benefit of the app extends to users and people around them; however, if digital exclusion prevents the most vulnerable people from participating in digital technologies, then the collective societal benefits are not equitable. This highlights the importance of digital innovation being accompanied by nondigital approaches, which are efficient and effective for people who cannot or choose not to engage with digital solutions, and to evaluate the outcome across population groups [56].

Actions to Mitigate the Risk of Digital Inequalities Exacerbating Health Inequalities

There is sufficient evidence to conclude that digital exclusion should be considered across sectors as a social determinant of health, with the potential to exacerbate health inequalities if progress does not consider the structural, economic, social, and behavioral factors contributing to digital exclusion.

The COVID-19 pandemic has clearly highlighted the levels of inequality in the ownership of digital devices and demonstrated that the access to fixed and mobile internet connections is a gateway to essential health information and care and many other key services including education, food delivery, employment, and social support, all of which indirectly impact health. As described here, the underlying inequalities in the access to digital technology and the internet are evident, and if left unresolved, have the potential to exacerbate health and social inequalities in a digital era.

In order to progress towards digital inclusion, one policy approach adopted by certain countries is to declare that access to the internet is a human right [57], whereas others consider technology and the internet as an enabler of rights rather than a right in itself [58]. Declaration of access to the internet as a right may enforce action at national and international levels to address a factor that indirectly contributes to health, placing it at par with other rights including the access to food and safe housing, all of which contribute to health equality.

This study shows that the underlying drivers of digital exclusion are complex, and as such the solution cannot be obtained through a single action or from a single organization. Addressing digital exclusion would require government, public, private, and nonprofit sector organizations to collaborate to ensure that progress toward a digital future does not inadvertently leave some people behind. This will require a comprehensive understanding of the factors contributing to digital exclusion in local populations and how these factors differ among groups, areas, and over time, along with effective engagement with digitally excluded people to coproduce solutions that adequately address the contributing factors such as poverty, access, skills, or motivation.

For example, a common viewpoint is that addressing barriers to internet connectivity will address digital exclusion; however, this is not the case. Inequalities would continue to persist after access is resolved if the complex underlying factors that contribute to the digital divide are not addressed. a crucial first step is to understand the extent and drivers of digital exclusion within and between different populations through data and qualitative insights. Followed by a multidisciplinary and multisectoral response to address structural barriers (such as the lack of infrastructure to support adequate internet access in all areas), financial barriers (eg, costs of devices and data plans), digital skills and literacy, and other barriers to engagement (eg, cultural, concerns regarding trust and data privacy) (Table 1). Embedded evaluation alongside digital innovation is essential to ascertain the impact of each of these actions against key outcomes including differences in uptake, engagement and effectiveness across population groups. For progress to occur, there is a need to bring together actions against these key components in a comprehensive and coordinated plan; for example, the New York City Internet Masterplan [59] is a comprehensive roadmap to close the digital divide through enhanced infrastructure, affordability, and inclusion. To ensure collective progress, it may also be necessary to ensure that the government takes the responsibility for digital equity.

Organizations in the United Kingdom have brought together practical calls to action to mitigate the risk of digital inequalities. The Good Things Foundation has issued a blueprint for a 100% digitally included United Kingdom for a post–COVID-19 economy, focusing on 3 key steps to fix the digital divide: the need to address the digital infrastructure, data poverty, and to develop an inclusive digital strategy [60]. The Carnegie UK Trust has reported 12 recommendations for policymakers,

practitioners, academics, and industry professionals to learn from the lockdown and eliminate digital exclusion [61]. In Wales, the Older People's Commissioner has urged investment in digital inclusion as one of the preconditions for an age-friendly recovery [62]. With an international focus, particularly on low-income countries, the Human Rights, Big Data and Technology Project has postulated 5 urgent principles for leaving nobody behind through technology in the COVID-19 response [63]: guaranteeing internet access as a human right and a public good, increasing the availability and acceptability of the digital infrastructure, increasing the accessibility and affordability of digital services, empowering people by addressing disinformation and hate speech without censorship, and ensuring that internet access is not a cause for more surveillance.

Table 1. Data and insights on the extent and drivers of digital exclusion leading to digital inequalities in population subgroups and potential approaches to address inequalities and to evaluate and understand the uptake and effectiveness of digital technologies and differences across population subgroups.

Driver of digital inequality	Types of digital inequalities	Suggested approaches to address inequalities
Structural barriers (access)	• Inequalities in access to a rapid and reliable internet connection	• Infrastructure investment • Report publicly available data on internet coverage across populations
Financial barriers	• Inequalities in the affordability of internet-enabled digital technology and data plans	• Understand economic barriers including internet access, affordability of technology, data poverty through research, and engagement with service users • Examine the impact of different policies and economic responses (eg, financial support, waived data charges for essential services, subsided costs, market regulations to reduce cost, and data caps)
Digital skills and literacy	• Inequalities in digital skills among service users	• Nationwide and local systems should use existing measures of digital inclusion to understand population needs, such as the general Digital Inclusion Scale [64] or the health-specific eHealth Literacy Scale [65] • Implement structured programs to address skills gaps such as the NHS Widening Digital Participation delivered of the Good Things Foundation [35,37] and Digital Communities Wales [36] • Targeted support addressing essential digital skills among vulnerable groups (eg, through peer-to-peer support, intergenerational mentoring, and localized digital champions [66])
	• Inequalities in digital skills, knowledge, understanding, and awareness among service providers and the workforce	• Develop a model of digital competence across different roles, train and support the workforce to enable a digital future (eg, the NHS Digital Readiness program [40])
Engagement with digital platforms	• Differences in motivation, trust, and perception of risks across populations	• Develop a nationwide standard instrument routinely administered to users to understand factors contributing to engagement • Coproduce digital solutions tailored to service user needs and levels of digital skills or engagement. • Address barriers of trust and risk with transparency and clarity on the collected data and the underlying purpose
Design and development of services	• Incompatibility across different platforms (fixed or mobile) • Inoperability owing to lower internet speeds • Requiring the use of data plans	• Include digital inequality as a factor in the equality impact assessments alongside service development and delivery • Incentivize partnerships across public, private, and nonprofit sector organizations to address digital exclusion • Ensure new service developments considering digital exclusion • Contracts with suppliers to include reports on granular data on service use in local health systems across platforms to identify excluded populations

As health and care systems seek to deliver services through digital platforms, it remains important to monitor and evaluate the levels of engagement to ensure that a focus on digital approaches does not inadvertently reinforce underlying inequalities in health. Within the United Kingdom, the NHS provides a potential opportunity to monitor and address different levels of engagement with digital technology to help understand differences by clinical needs and population groups, an understanding of which could be shared to inform innovation elsewhere.

Based on theoretical frameworks and the lessons learned from the practical examples described here, the actions needed to ensure that digital health innovation helps address, rather than exacerbate, health inequalities are as follows:

1. Understand the extent and drivers of digital exclusion in population subgroups
2. Use existing measures or develop new measures of digital exclusion and digital health literacy
3. Ensure that digital strategies and new service developments consider factors contributing to local exclusion
4. Develop solutions through coproduction for a maximally extensive and appropriate user base
5. Where widespread adoption is required, ensure the technologies used have the characteristics of innovations capable of being rapidly disseminated [67-69]
6. Use realistic and evidence-based models of behavioral science to inform digital health innovation, such as the COM-B system [70].

Conclusion

Digital technology has the potential to revolutionize health and health care with growing interest among policymakers, researchers, and practitioners in exploring how digital technology can be harnessed to improve population health. Countries worldwide have been using digital technologies to respond to the COVID-19 pandemic, for applications including communication of public information, remote delivery of health care, and population surveillance [71]. Greater adoption of technology can also provide additional sources of valuable data for potential use by population health systems to understand health needs more clearly, monitor outcomes, and counter the factors underlying ill health.

Digital innovations should be accessible to everyone, empowering citizens to become active contributors to health and well-being. However, those likely to have the greatest health needs are also least likely to have access to digital platforms and the skills to use and navigate them, and they are hence less likely to engage with such digital platforms. This raises an important question of equity in population health in a digital era, an issue brought into sharp focus by the COVID-19 pandemic. This led us to ask how we can innovate through digital technology and transform population health, while leaving no one behind.

To address this question, we need to better understand who engages with digital technologies, the enablers and barriers, and the direct and indirect impact on health outcomes. The lack of consideration of these factors poses the danger that the pursuit of digital health solutions results in unintended consequences and reinforces existing social and health inequalities. Action is needed across government, public, private, and nonprofit sector organizations to ensure capitalization on the potential for digital technology to address health and minimize the risk of exacerbation of health inequalities.

Conflicts of Interest

MH works for accuRx, a health care communications software supplier. ARD and BG have no conflicts of interest to declare.

References

1. The future of healthcare: our vision for digital, data and technology in health and care. Department of Health & Social Care: Government of the United Kingdom. 2018 Oct 17. URL: https://www.gov.uk/government/publications/the-future-of-healthcare-our-vision-for-digital-data-and-technology-in-health-and-care/the-future-of-healthcare-our-vision-for-digital-data-and-technology-in-health-and-care [accessed 2021-03-23]
2. The NHS Long Term Plan. National Health Service. 2019 Jan. URL: https://www.longtermplan.nhs.uk/wp-content/uploads/2019/08/nhs-long-term-plan-version-1.2.pdf [accessed 2021-03-23]
3. Castle-Clarke S, Hutchings R. Achieving a digital NHS: Lessons for national policy from the acute sector. Nuffield Trust. 2019 May 31. URL: https://www.nuffieldtrust.org.uk/research/achieving-a-digital-nhs-lessons-for-national-policy-from-the-acute-sector [accessed 2021-03-23]
4. A healthier Wales: long term plan for health and social care. Welsh Government. 2018 Jun 08. URL: https://gov.wales/healthier-wales-long-term-plan-health-and-social-care [accessed 2021-03-23]
5. Digital health. World Health Organization. URL: https://www.who.int/health-topics/digital-health#tab=tab_1 [accessed 2021-03-23]
6. Hart JT. The inverse care law. Lancet 1971 Feb 27;1(7696):405-412 [FREE Full text] [doi: 10.1016/s0140-6736(71)92410-x] [Medline: 4100731]
7. Eysenbach G. Consumer health informatics. BMJ 2000 Jun 24;320(7251):1713-1716 [FREE Full text] [doi: 10.1136/bmj.320.7251.1713] [Medline: 10864552]
8. Milner H. Inclusion critical for digital-first health service. Health Service Journal. URL: https://www.hsj.co.uk/technology-and-innovation/inclusion-crucial-for-a-digital-first-health-service/7027443.article [accessed 2021-03-23]
9. Coronavirus response must include digital access to connect us all. Joseph Rowntree Foundation. 2020 Apr 06. URL: https://www.jrf.org.uk/blog/coronavirus-response-must-include-digital-access-connect-us-all [accessed 2021-03-23]
10. Harrison MI, Koppel R, Bar-Lev S. Unintended consequences of information technologies in health care--an interactive sociotechnical analysis. J Am Med Inform Assoc 2007;14(5):542-549 [FREE Full text] [doi: 10.1197/jamia.M2384] [Medline: 17600093]

11. Castle-Clark S. What will new technology mean for the NHS and its patients?: Four big technological trends. The Health Foundation, the Institute for Fiscal Studies, The King's Fund, and the Nuffield Trust. 2018. URL: https://www.kingsfund.org.uk/sites/default/files/2018-06/NHS_at_70_what_will_new_technology_mean_for_the_NHS_0.pdf [accessed 2021-03-23]
12. Vogels E, Perrin A, Rainie L, Anderson M. 53% of Americans Say the Internet Has Been Essential During the COVID-19 Outbreak. Pew Research Center. 2020 Apr 30. URL: https://www.pewresearch.org/internet/2020/04/30/53-of-americans-say-the-internet-has-been-essential-during-the-covid-19-outbreak/ [accessed 2021-03-23]
13. The State of Broadband: Broadband as a Foundation for Sustainable Development. Broadband Commission for Sustainable Development. 2019 Sep. URL: https://www.itu.int/dms_pub/itu-s/opb/pol/S-POL-BROADBAND.20-2019-PDF-E.pdf [accessed 2021-03-23]
14. Prescott C. Internet access – households and individuals, Great Britain: 2019. Office for National Statistics. 2019 Aug 12. URL: https://www.ons.gov.uk/peoplepopulationandcommunity/householdcharacteristics/homeinternetandsocialmediausage/bulletins/internetaccesshouseholdsandindividuals/2019 [accessed 2021-03-23]
15. Serafino P. Exploring the UK's digital divide. Office for National Statistics. 2019 Mar 04. URL: https://www.ons.gov.uk/releases/exploringtheuksdigitaldivide [accessed 2021-03-23]
16. Anderson M, Perrin A, Jiang J, Kumar M. 10% of Americans don't use the internet. Who are they? Pew Research Center. 2019 Apr 22. URL: https://www.pewresearch.org/fact-tank/2019/04/22/some-americans-dont-use-the-internet-who-are-they/ [accessed 2019-11-19]
17. König R, Seifert A, Doh M. Internet use among older Europeans: an analysis based on SHARE data. Univ Access Inf Soc 2018 Jan 19;17(3):621-633. [doi: 10.1007/s10209-018-0609-5]
18. Digital Economy and Society Index Report 2020 - Connectivity. European Commission. URL: https://ec.europa.eu/digital-single-market/en/broadband-connectivity#:~:text=The%20Broadband%20Coverage%20in%20Europe,Mbps%20or%20more%20by%202020 [accessed 2021-03-23]
19. Connected Nations 2019: UK Report. Ofcom. 2019 Dec 20. URL: https://www.ofcom.org.uk/__data/assets/pdf_file/0023/186413/Connected-Nations-2019-UK-final.pdf [accessed 2021-03-23]
20. Worst Connected Cities 2018. National Digital Inclusion Alliance. 2018. URL: https://www.digitalinclusion.org/worst-connected-2018/ [accessed 2021-03-23]
21. Nursing in the Digital Age, Using Technology to Support Patients in the Home. The Queen's Nursing Institute. URL: https://www.qni.org.uk/resources/nursing-in-the-digital-age/ [accessed 2021-03-23]
22. Kelly A. Digital divide 'isolates and endangers' millions of UK's poorest. The Guardian. 2020 Apr 28. URL: https://www.theguardian.com/world/2020/apr/28/digital-divide-isolates-and-endangers-millions-of-uk-poorest [accessed 2020-04-29]
23. Lucas PJ, Robinson R, Treacy L. What Is Data Poverty? Nesta. 2020 Dec. URL: https://media.nesta.org.uk/documents/What_is_Data_Poverty.pdf [accessed 2021-03-23]
24. Social connection in the COVID-19 crisis: Initial report from the COVID-19 Inquiry by the All Party Parliamentary Group on Social Integration. All Party Parliamentary Group on Social Integration. URL: https://cuf.org.uk/uploads/resources/APPG-Social-Connection-In-the-COVID-19-Crisis.pdf [accessed 2021-03-23]
25. CV-19 Response: DevicesDotNow. FutureDotNow. URL: https://futuredotnow.uk/devicesdotnow/ [accessed 2021-03-23]
26. Walker D. 1000 extra digital devices will enable the rapid expansion of digital health services in Wales. Digital Communities Wales: Welsh Government. URL: https://www.digitalcommunities.gov.wales/blog/2020/04/13/key-care-providers-given-digital-devices/ [accessed 2021-03-23]
27. Broadband providers to lift data caps during Covid-19 lockdown. The Guardian. 2020 Mar 29. URL: https://www.theguardian.com/world/2020/mar/29/broadband-providers-to-lift-data-caps-during-covid-19-lockdown [accessed 2021-03-23]
28. Lloyds Bank UK Consumer Digital Index 2020: The UK's largest study of transactional, behavioural and attitudinal research including the Essential Digital Skills measure. Lloyds Bank. URL: https://www.lloydsbank.com/banking-with-us/whats-happening/consumer-digital-index.html [accessed 2021-03-23]
29. Human Capital and Digital Skills. European Commission. URL: https://ec.europa.eu/digital-single-market/en/human-capital-and-digital-skills [accessed 2020-12-18]
30. NHS Widening Digital Participation. Good Things Foundation: Digital Health Lab. URL: https://digital-health-lab.org/ [accessed 2021-03-23]
31. Gann B. Digital inclusion in health and care in Wales. Wales Co-operative Centre. 2018 Nov. URL: https://wales.coop/digital-inclusion-report-2018/ [accessed 2021-03-23]
32. Robison S, Johnston CP. Scotland's Digital Health and Care Strategy: enabling, connecting and empowering. Scottish Government. 2018 Apr 25. URL: https://www.gov.scot/publications/scotlands-digital-health-care-strategy-enabling-connecting-empowering [accessed 2021-03-23]
33. The Digital Inclusion Blueprint – Te Mahere mō te Whakaurunga Matihiko. New Zealand Government. URL: https://www.digital.govt.nz/dmsdocument/113-digital-inclusion-blueprint-te-mahere-mo-te-whakaurunga-matihiko/html [accessed 2021-03-23]

34. Thomas J, Barraket J, Wilson C, Rennie E, Ewing S, MacDonald T. Measuring Australia's digital divide: the Australian digital inclusion index 2019. Analysis and Policy Observatory. 2019 Sep 17. URL: https://apo.org.au/node/255341 [accessed 2021-03-23]
35. Coronavirus (COVID-19) and digital inclusion. Good Things Foundation. URL: https://www.goodthingsfoundation.org/coronavirus-and-digital-inclusion [accessed 2021-03-23]
36. Digital Communities Wales. URL: https://www.digitalcommunities.gov.wales/ [accessed 2021-03-23]
37. Learn My Way. Good Things Foundation. URL: https://www.goodthingsfoundation.org/learn-my-way [accessed 2021-03-23]
38. Digital skills initiatives. European Commission. 2021 Mar 09. URL: https://ec.europa.eu/digital-single-market/en/digital-skills-initiatives [accessed 2021-03-23]
39. Interreg Northern Ireland – Ireland – Scotland. URL: https://mpowerhealth.eu/ [accessed 2021-03-23]
40. The Digital Readiness Programme. NHS Health Education England. URL: https://www.hee.nhs.uk/our-work/building-digital-ready-workforce [accessed 2021-03-23]
41. Greenhalgh T, Wherton J, Shaw S, Morrison C. Video consultations for covid-19. BMJ 2020 Mar 12;368:m998. [doi: 10.1136/bmj.m998] [Medline: 32165352]
42. Johns G, Khalil S, Ogonovsky M, Wright P, Williams J, Lees M, et al. The NHS Wales Video Consulting Service, TEC Cymru. 2020 Dec. URL: https://digitalhealth.wales/tec-cymru/how-we-can-help/evidence/eval-reports [accessed 2021-03-23]
43. Valdez R, Gibbons M, Siegel E, Kukafka R, Brennan P. Designing consumer health IT to enhance usability among different racial and ethnic groups within the United States. Health Technol 2012 Jul 13;2(4):225-233 [FREE Full text] [doi: 10.1007/s12553-012-0031-6]
44. Marmot M. Health equity in England: the Marmot review 10 years on. BMJ 2020 Feb 24;368:m693. [doi: 10.1136/bmj.m693] [Medline: 32094110]
45. McAuley A. Digital health interventions: widening access or widening inequalities? Public Health 2014 Dec;128(12):1118-1120. [doi: 10.1016/j.puhe.2014.10.008] [Medline: 25458115]
46. Ramsetty A, Adams C. Impact of the digital divide in the age of COVID-19. J Am Med Inform Assoc 2020 Jul 01;27(7):1147-1148 [FREE Full text] [doi: 10.1093/jamia/ocaa078] [Medline: 32343813]
47. Arroll B, Gill E, Roskvist R, Harwood M. Covid-19 in New Zealand: A move to online consultations in primary care could widen health inequalities. BMJ Opinion. 2020 Jun 25. URL: https://blogs.bmj.com/bmj/2020/06/25/covid-19-in-new-zealand-a-move-to-online-consultations-in-primary-care-could-widen-health-inequalities/ [accessed 2021-03-23]
48. Davies A, Sharp C, Homolova L, Bellis M. Population health in a digital age: The use of digital technology to support and monitor health in Wales. Public Health Wales and Bangor University. 2019. URL: https://phw.nhs.wales/files/research/population-health-in-a-digital-age/ [accessed 2021-03-23]
49. Caul S. Deaths involving COVID-19 by local area and socioeconomic deprivation: deaths occurring between 1 March and 17 April 2020. Office for National Statistics. 2020 May 01. URL: https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/bulletins/deathsinvolvingcovid19bylocalareasanddeprivation/deathsoccurringbetween1marchand17april [accessed 2020-05-20]
50. Van den Broucke S. Why health promotion matters to the COVID-19 pandemic, and vice versa. Health Promot Int 2020 Apr 01;35(2):181-186 [FREE Full text] [doi: 10.1093/heapro/daaa042] [Medline: 32297931]
51. Dahlgren G, Whitehead M. Concepts and principles for tackling social inequities in health: Levelling up Part 1. World Health Organization: Regional Office for Europe. 2006. URL: https://isg.org.ar/wp-content/uploads/2011/09/Inequidades-Sociales-original.pdf [accessed 2021-03-25]
52. Mobile networks remove data charges for online NHS coronavirus advice. Government of the United Kingdom. 2020 Mar 18. URL: https://www.gov.uk/government/news/mobile-networks-remove-data-charges-for-online-nhs-coronavirus-advice [accessed 2021-03-23]
53. Zastrow M. Coronavirus contact-tracing apps: can they slow the spread of COVID-19? Nature 2020 May 19. [doi: 10.1038/d41586-020-01514-2] [Medline: 32433633]
54. NHS COVID-19 app. National Health Service. URL: https://www.nhsx.nhs.uk/covid-19-response/nhs-covid-19-app/ [accessed 2021-03-23]
55. The Health Foundation COVID-19 Survey. Ipsos MORI. 2020 May. URL: https://www.health.org.uk/sites/default/files/2020-06/Health-Foundation-polling-contact-tracing-app-May-2020.pdf [accessed 2021-03-23]
56. COVID-19 rapid evidence review. Ada Lovelace Institute. 2020 Apr 19. URL: https://www.adalovelaceinstitute.org/our-work/covid-19/covid-19-exit-through-the-app-store/ [accessed 2021-03-23]
57. Psaila SB. Right to access the Internet: the countries and the laws that proclaim it. Diplo. 2011 May 02. URL: https://www.diplomacy.edu/blog/right-access-internet-countries-and-laws-proclaim-it [accessed 2021-03-23]
58. Cerf V. Internet access is not a human right. The New York Times. 2012 Jan 04. URL: https://www.nytimes.com/2012/01/05/opinion/internet-access-is-not-a-human-right.html [accessed 2021-03-23]
59. Internet Master Plan: A new vision for broadband infrastructure and service in New York City. New York City Mayor's Office of the Chief Technology Officer. URL: https://www1.nyc.gov/assets/cto/#/project/internet-master-plan [accessed 2021-03-23]

60. Blueprint for a 100% Digitally Included UK: For a post-COVID-19 economy. Good Things Foundation. 2020 Sep. URL: https://www.goodthingsfoundation.org/sites/default/files/blueprint-for-a-100-digitally-included-uk-0.pdf [accessed 2021-03-23]
61. Bowyer G, Grant A, White D. Learning from Lockdown: 12 Steps to Eliminate Digital Exclusion. Carnegie UK Trust. 2020. URL: https://www.carnegieuktrust.org.uk/publications/learning-from-lockdown-12-steps-to-eliminate-digital-exclusion/ [accessed 2021-03-23]
62. Leave no-one behind: Action for an age-friendly recovery. Older People's Commissioner for Wales. URL: https://www.olderpeoplewales.com/Libraries/Uploads/Leave_no-one_behind_-_Action_for_an_age-friendly_recovery.sflb.ashx [accessed 2021-03-23]
63. The COVID-19 Pandemic: Five urgent principles for leaving no one behind through technology. The Human Rights, Big Data and Technology Project. URL: https://www.hrbdt.ac.uk/the-covid-19-pandemic-five-urgent-principles-for-leaving-no-one-behind-through-technology/ [accessed 2021-03-23]
64. Essential digital skills framework. Government of the United Kingdom. 2018 Sep 12. URL: https://www.gov.uk/government/publications/essential-digital-skills-framework [accessed 2021-03-23]
65. Norman CD, Skinner HA. eHEALS: The eHealth Literacy Scale. J Med Internet Res 2006 Nov 14;8(4):e27 [FREE Full text] [doi: 10.2196/jmir.8.4.e27] [Medline: 17213046]
66. Government Digital Inclusion Strategy. Government of the United Kingdom. 2014 Dec 04. URL: https://www.gov.uk/government/publications/government-digital-inclusion-strategy/government-digital-inclusion-strategy#measuring-digital-exclusion [accessed 2021-03-23]
67. Rogers E, Singhal A. An integrated approach to communication theory and research. 5th edition. London: Routledge; 2003:409-441.
68. Michie S, van Stralen MM, West R. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. Implement Sci 2011 Apr 23;6:42 [FREE Full text] [doi: 10.1186/1748-5908-6-42] [Medline: 21513547]
69. Greenhalgh T, Procter R, Wherton J, Sugarhood P, Hinder S, Rouncefield M. What is quality in assisted living technology? The ARCHIE framework for effective telehealth and telecare services. BMC Med 2015 Apr 23;13:91 [FREE Full text] [doi: 10.1186/s12916-015-0279-6] [Medline: 25902803]
70. Michie S, Atkins L, West R. The Behaviour Change Wheel: A Guide To Designing Interventions. London: Silverback Publishing; 2014.
71. Fahy N. Cross-country analysis: How countries are using digital health tools in responding to COVID-19? World Health Organization. 2020 Apr 28. URL: https://analysis.covid19healthsystem.org/index.php/2020/04/28/how-are-countries-using-digital-health-tools-in-responding-to-covid-19/ [accessed 2021-03-23]

Abbreviations

NHS: National Health Service

Edited by G Eysenbach, R Kukafka; submitted 23.06.20; peer-reviewed by T Greenhalgh, A Sheon, E Hargittai; comments to author 28.10.20; revised version received 20.12.20; accepted 17.03.21; published 07.04.21.

Please cite as:
Davies AR, Honeyman M, Gann B
Addressing the Digital Inverse Care Law in the Time of COVID-19: Potential for Digital Technology to Exacerbate or Mitigate Health Inequalities
J Med Internet Res 2021;23(4):e21726
URL: https://www.jmir.org/2021/4/e21726
doi:10.2196/21726
PMID:33735096

Viewpoint

The Current Situation and Future Prospects of Simulators in Dental Education

Yaning Li[1,2,3,4,5*], BDS; Hongqiang Ye[1,2,3,4,5*], DDS, PhD; Fan Ye[6], BE; Yunsong Liu[1,2,3,4,5], DDS, PhD; Longwei Lv[1,2,3,4,5], DDS, PhD; Ping Zhang[1,2,3,4,5], PhD; Xiao Zhang[1,2,3,4,5], DDS, PhD; Yongsheng Zhou[1,2,3,4,5], DDS, PhD

[1]Department of Prosthodontics, Peking University School and Hospital of Stomatology, Beijing, China

[2]National Clinical Research Center for Oral Diseases, Peking University School and Hospital of Stomatology, Beijing, China

[3]National Engineering Laboratory for Digital and Material Technology of Stomatology, Peking University School and Hospital of Stomatology, Beijing, China

[4]Beijing Key Laboratory of Digital Stomatology, Peking University School and Hospital of Stomatology, Beijing, China

[5]NHC Key Laboratory of Digital Technology of Stomatology, Peking University School and Hospital of Stomatology, Beijing, China

[6]The State Key Laboratory of Virtual Reality Technology and Systems, School of Computer Science and Engineering, Beihang University, Beijing, China

[*]these authors contributed equally

Corresponding Author:
Yongsheng Zhou, DDS, PhD
Department of Prosthodontics
Peking University School and Hospital of Stomatology
22 Zhongguancun South Avenue, Haidian District
Beijing, 100081
China
Phone: 86 010 82195070
Email: kqzhouysh@hsc.pku.edu.cn

Abstract

The application of virtual reality has become increasingly extensive as this technology has developed. In dental education, virtual reality is mainly used to assist or replace traditional methods of teaching clinical skills in preclinical training for several subjects, such as endodontics, prosthodontics, periodontics, implantology, and dental surgery. The application of dental simulators in teaching can make up for the deficiency of traditional teaching methods and reduce the teaching burden, improving convenience for both teachers and students. However, because of the technology limitations of virtual reality and force feedback, dental simulators still have many hardware and software disadvantages that have prevented them from being an alternative to traditional dental simulators as a primary skill training method. In the future, when combined with big data, cloud computing, 5G, and deep learning technology, dental simulators will be able to give students individualized learning assistance, and their functions will be more diverse and suitable for preclinical training. The purpose of this review is to provide an overview of current dental simulators on related technologies, advantages and disadvantages, methods of evaluating effectiveness, and future directions for development.

(J Med Internet Res 2021;23(4):e23635) doi:10.2196/23635

KEYWORDS

dental simulator; dental education; virtual reality

Introduction

Dental skills training is a very important part of preclinical learning in dental education and has a long history [1]. Pioneers of dental education began to use extracted teeth in dental skills education in the 1800s [2]. Later, Oswald Fergus invented the first phantom head simulator in 1894, which was used to teach oral anatomy and physiology to dental students [2]. Since then, the phantom head simulator has developed rapidly. Modern phantom head simulators include water spray, dental handpieces, and other necessary items [3], providing students with a more realistic environment for diagnosis and treatment. The dental simulator appeared in the 1990s [4,5] as a result of further research into methods of dental preclinical education, concern for patient safety, improvements in computer technology, and the inappropriateness of a clinical environment for the novice

[1]. The arrival of the dental simulator marked a new era of dental preclinical education.

The dental simulator replicates both soft and hard oral tissues as well as providing a clinical diagnosis and treatment environment through virtual reality (VR). It also simulates the interaction force between the bur and the tooth, the mouth mirror, and soft and hard tissues through force feedback technology to reproduce the whole training process for dental clinical skills as closely as possible [6,7]. In recent years, the dental simulator has mainly been used for adult vocational training and university education [8].

Traditional preclinical dental skills training, which was based on a phantom head, extracted teeth, or plastic teeth [9,10], is generally used for practicing tooth preparation, for which the processes are irreversible. The acquisition of extracted teeth becomes more and more difficult, and the sensory feedback of preparing plastic teeth is different from that of real teeth [3,11]. The dental simulator, simulating realistic clinical conditions via VR and force feedback, makes training reversible, repeatable, and environmentally friendly [12,13]. Training via a dental simulator is varied [14] since different training content and tooth positions are available. These can be displayed in 3D on a computer screen for real-time evaluation by and feedback from teachers.

Technologies Included in Dental Simulators

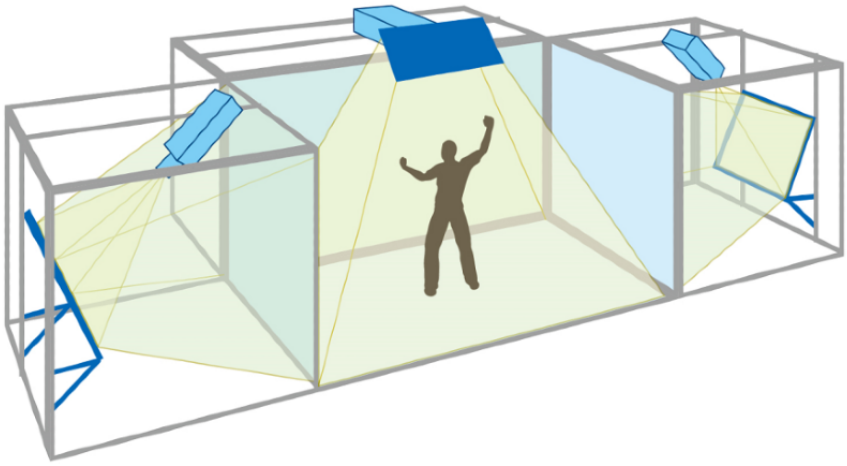
The dental simulator is a deep integration of computer and dental technology, mainly consisting of VR and force feedback technologies [15].

Virtual Reality

VR uses computer technology [16] to generate a digital environment similar to the real environment in the visual, auditory, tactile, and other senses [17], which is used in many different fields. The operator interacts with and feels feedback from the virtual objects using specialized equipment [18]. A complete VR system consists of a stereo display device, a motion tracking device, an input device, and a computing platform. The stereo display device is usually a head-mounted display (HMD; Figure 1) [19]; another type of stereo display is the Cave Automatic Virtual Environment (CAVE; Figure 2) [20]. HMDs are more widely used than CAVEs because HMD application is more flexible and requires less space.

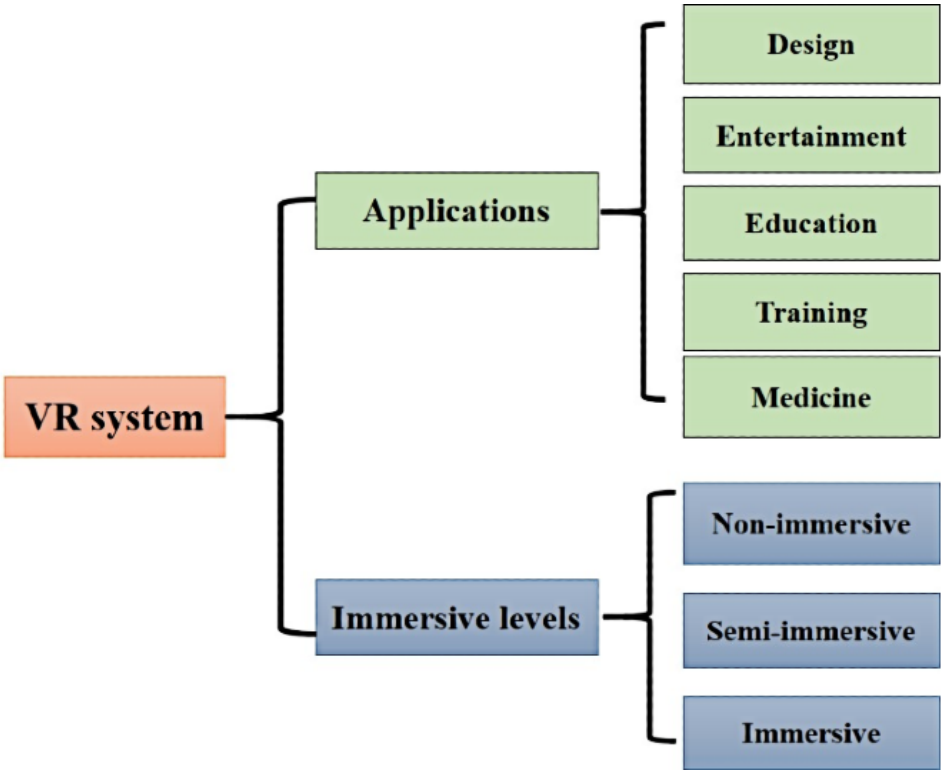
Figure 1. Head-mounted display and body motion sensors.

Figure 2. Cave Automatic Virtual Environment (CAVE).



VR systems are divided into three immersive levels based on the degree of stimulated senses and interactions: nonimmersive system, semi-immersion system, and immersion system (Figure 3) [21]. A nonimmersive system only reproduces images on desktops; an immersion system places the user in a complete virtual environment with the support of several sensory output devices including visual devices such as HMDs, audio devices, and haptic devices [22]; and a semi-immersion system provides the user with a simulated environment between the two above.

Figure 3. Applications and immersive levels of virtual reality (VR) systems.



In the implementation of VR technology, the key points are modeling and interaction. In the medical field, images from computed tomography [7,23], magnetic resonance imaging (MRI) [24], and dental scanners [25] can be used to rehabilitate virtual models. Interactions mainly contain visual interaction and tactile sensation interaction, which are actualized by display device and force feedback device, respectively.

In the medical field, VR and related technology are gradually being applied in surgical training [26] and surgical navigation systems [27]. Combining VR with medical training is a new research field that has emerged with the development of computer science, sensor technology, and automation technology. A virtual operating environment with high fidelity and real-time performance can be created through virtual simulation technology [14]. The system is created and processed by a computer, with dedicated devices such as a helmet display and a force feedback handle that allow the user to observe and interact with the scene while experiencing multisensory feedback that give a near-realistic training effect [28].

Force Feedback

Haptic devices can receive and transmit motion signals to improve the operator's sense of reality [29], which is important for dental training assisted by VR.

To simulate the tactile sense of real dental training as much as possible, the reaction force of the virtual object is calculated using the appropriate force generation algorithm in the VR system [30]. Due to the characteristics of human touch receptors, real-time haptic rendering requires a refresh frequency of at least 1 kHz [31]. For force feedback interaction in dental skills training, several kinds of interaction algorithms have been proposed for various scenarios, including teeth preparation, scaling, and bone drilling [32,33].

Introduction to Existing Dental Simulators

Overview

There is a wide variety of dental education simulators available, each with advantages and disadvantages in terms of training content, training process, hardware device, and software design. These are briefly described in Table 1.

Table 1. Comparison of dental simulators.

Characteristic	DentSim	Virtual Education System for Dentistry	IDEA[a]	Periosim	iDental	Simodont Dental Trainer	VirTeaSy	IDEAL[b]	Voxel-Man
Hardware facility									
Display type	2D display+ phantom head	2D display+ phantom head	2D display	3D glasses+ 2D display	2D display	3D glasses+ 2D display	3D glasses+ 2D display	2D display+ phantom head	3D glasses+ 2D display
Operation with two hands	Available	Available	Available	Available	Available	Available	No	Available	Available
Fixed finger rest	Yes	Yes	No	No	Yes	Yes	Yes	No	No
Ergonomic postures	Yes	Yes	No	No	Yes	Yes	Yes	Yes	Yes
Software design									
Application field	Tooth preparation	Crown preparation	Manual dexterity training	Periodontal training	Periodontal training	Caries removal and crown preparation	Dental implant surgery	Dental radiography	Dental surgery
Fundamental skills	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Clinical cases	No	No	No	No	Yes	Yes	Yes	Yes	No
Exam simulation	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes
Repetitive practice	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Practice at different levels	No	No	Yes	No	Yes	Yes	Yes	No	Yes
Individualized learning	No	No	No	No	No	No	Yes	No	No
Result evaluation	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes
Force feedback	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
VR[c] immersive level	Non[d]	Non	Non	Non	Non	Non	Non	Non	Non

[a]IDEA: Individual Dental Education Assistant.

[b]IDEAL: Internet of Things–based dental education and learning.

[c]VR: virtual reality.

[d]Non: nonimmersive.

DentSim

The DentSim system unit, consisting of a phantom head and dentoform, dental instruments, infrared sensors, infrared cameras, and two computers, was born in 1997 [34]. The infrared cameras can capture the orientation and movement of resin teeth and a handpiece so as to show the students' work virtually on the computer screen in real time. The unit allows students to see the evaluation of their tooth preparation compared with the ideal preparation on the screen, while also providing them the ability to continue working on the resin teeth [35]. The training using DentSim is more efficient and standardized than that using traditional preclinical teaching methods [36]. The disadvantage of this unit is that it relies on physical resin teeth, which are disposable consumables.

Virtual Education System for Dentistry

Virtual Education System for Dentistry is a dental simulator for prosthodontics developed by the Affiliated Stomatological Hospital of Nanjing Medical University and Suzhou Digital-health Care Company. The system contains the Virtual Learning Network Platform (VLNP) and Real-time Dental Training and Evaluation System (RDTES) [37]. Prior to practical work, students are requested to learn courses on the VLNP, including reading the operational instructions and predefined criteria of crown preparation and watching the standard operational videos. Afterwards, students can perform crown preparations on the phantom head under the guidance of the RDTES; the processes and results can be recorded by the RDTES. When the students finish their preparations, the RDTES can automatically assess the procedures and results of preparation based on the predefined tooth preparation criteria. As well, the students can visually compare their own procedures and results with the predefined assessment criteria on the computer screen [37,38].

Individual Dental Education Assistant

Individual Dental Education Assistant (IDEA) is a VR hand flexibility training simulator consisting of a handheld stylus that simulates a dental handpiece and provides force feedback and a computer installed with simulation software. Unlike other dental simulators, IDEA is designed to enable students to be flexible and proficient in the use of dental handpieces by practicing removing predesigned virtual materials with different shapes (eg, straight line or circle). Therefore, IDEA aims to train dental students in hand flexibility, not to train students on a particular teaching component such as crown preparation or scaling. The main advantage of the system is its evaluation system. During the training process, two parameters determine the score obtained: drilling speed and drilling accuracy. Deviation from the trajectory or to an inappropriate depth can lead to a decrease in accuracy, and this is displayed as an accuracy bar on the screen. Complete depletion of the bar means that the student fails the test [39]. Some researchers have reported that IDEA could improve students' performance in the dental skill test; in addition, it can be used to identify students with troubles with hand flexibility at an early stage that may allow for early intervention to prevent failure [40,41].

PerioSim

PerioSim, consisting of a stereoscopic display, a computer, and a haptic device, allows students to use virtual dental instruments to visualize and detect caries and periodontal diseases in a haptic environment [12]. The system is available online for students and allows teachers to upload different training programs, which can be downloaded and replayed by students at any time, making this system convenient and efficient [35]. Steinberg et al reported that the image display and force feedback were very realistic for teeth and dental instruments but not for gingival tissue [42].

iDental

*i*Dental is a periodontal skill education simulator developed by Peking University School and Hospital of Stomatology and Beihang University, which can simulate periodontal examinations and treatment procedures including periodontal probing and calculus detection and removal. Unlike PerioSim, the device mainly uses a 2D monitor. However, it is equipped with an odontoscope handle, so it can be used to practice two-handed cooperative operation, making it realistic. *i*Dental also has a basic periodontal knowledge teaching module, which enables students to review basic knowledge before operation training. A combination of the two training parts improves the teaching effectiveness [43].

Simodont Dental Trainer

Simodont Dental Trainer is a widely used teaching simulator for dental skills training that is currently available in many dental schools. It mainly includes modules for hand flexibility, cariology, crown and bridge preparations, clinical cases, and a full mouth simulation experience [13,44]. One of the highlights of the system is that an X-ray of the working tooth is attached to each individual case, which can allow students to make a diagnosis assisted by both the appearance of teeth and the X-ray films [13].

The system does have some disadvantages. It requires 3D glasses for 3D display. Excessive training on a single tooth fails to create a realistic sense of manipulation. The training process for crown preparation is single-jawed and does not fully mimic the narrow operating space of the mouth. Besides, it cannot be used to train students about positioning requirements because operation postures during training are fixed and visual angle conversion requires manual rotation of the rotary button.

VirTeaSy Project

The VirTeaSy project is a dental implant training simulator and is composed of VirTeaSy Scan Implant and VirTeaSy Implant Pro. VirTeaSy Scan Implant is used for implanting scheme designs by students. Radiographs are used by students to perform the implant treatment plan, including the implant's characteristics (shape, diameter, length) and its location (location, angle, insertion depth) in the jaw. When the treatment plan is complete, it can be compared with the design planned by an expert that is stored in the database to identify where improvements could be made in their own plans. In addition, the device's database contains cases of varying difficulty, so students with different skill levels can use it to practice accordingly [45].

VirTeaSy Implant Pro is a virtual implant surgery training system, which allows students to perform surgery cases planned in VirTeaSy Scan Implant. VirTeaSy Implant Pro is supplemented by an auxiliary system that can alert students if the drilling's location, angle, and depth are incorrect as well as if there is overheating of the bone. In addition, the VirTeaSy Implant Pro has a display through which the teacher can interact with the student in real time and assist the student as necessary. VirTeaSy Implant Pro can improve students' skills in bone mineral density perception through force feedback, thus allowing them to perceive whether their bone density measurements in VirTeaSy Scan Implant match the reality [46]. VirTeaSy Scan Implant and Implant Pro complement each other, forming an efficient learning tool [45].

Internet of Things–Based Education and Learning System

The Internet of Things–based education and learning (IDEAL) system is an oral radiology education simulator. The simulator is mainly used for teaching students to take intraoral X-ray images without using X-rays. The IDEAL system consists of a simulator cone, simulator main body, sensor, detector, and stool. Training contents comprise basic education before X-ray imaging, information on X-ray imaging techniques (such as periapical radiography, bisecting angle technique, and paralleling technique), a test bank, and an evaluation and feedback system. During the training process, students can practice taking X-ray images for different tooth positions by adjusting the angle of the X-ray tube to improve their skills in taking X-ray images. At the end of training, students submit their own imaging results to the system and receive automatic feedback and evaluations from the system. The system allows students to avoid using radiographic devices while learning to take dental films, reducing the risk of radiation exposure to both students and patients. The system is safe and affordable [47].

Voxel-Man Simulator

The Voxel-Man simulator is a VR surgical simulator that primarily consists of a 2D monitor, simulated surgical operating handle, foot pedal, and 3D glasses. At present, the device is mainly used to simulate apical surgery, such as apical resection and apical cystectomy [48].

The system provides different training modes depending on the trainees' level, with different display interfaces and different operating instructions. The three modes to choose from are primary mode, advanced mode, and exam mode. In primary mode, the lesions and surrounding important anatomical structures (such as the alveolar nerve) are marked with bright colors. This not only helps operators to understand the anatomical characteristics of the corresponding surgical area, but also reminds them of the scope and amplitude of the operation at all times to help them avoid damaging these important anatomical structures. When the surgical instrument is close to an important anatomical structure, the system will emit a danger alarm. In advanced mode and exam mode, some of the functions and hints are turned off [48,49].

In addition, the system can record the operation process for later replay so that the operator can identify errors and mistakes to improve on [48]. The operator can stop the operation at any time and can undo a wrong step or restart a new one, saving a lot of time. Cases vary in difficulty, allowing operators with different experience levels to practice.

Advantages of Dental Simulators

Compared to phantom-based traditional training methods, dental simulators have many strengths that will offer students a better learning environment. Besides dental operation skills, students are also able to acquire relevant theoretical knowledge through dental simulators [37]. Since the dental simulators allow repeatable and reversible preclinical training of clinical skills [50,51], they give students a more flexible training experience [14]. They also allow digital objective evaluation and tutorial feedback [34,52] by recording the training processes [53]. In addition, training in dental simulators is more clinically relevant [54] because they recreate situations that are similar to those encountered in a real clinical environment [14]. It is certain that dental simulators can eliminate the risk of treatment and enhance the safety of patients [55,56]. Previous studies showed that training using dental simulators can save the time of faculty [50,57] and allow students to practice repeatedly whenever they want until they achieve mastery. Some studies reported that training in dental simulators can reduce training time compared with traditional training methods [57]. Therefore, the application of dental simulators in teaching can make up for the deficiency of traditional teaching methods and reduce the teaching burden, improving convenience for both teachers and students.

Current Disadvantages of Dental Simulators

The dental education simulators that have been described all have certain disadvantages in terms of both hardware and software.

Disadvantages of Hardware

The Stereo Vision of the Display Is Not Visible Enough and the Resolution Is Not High Enough

Currently, 3D displays in dental simulators are relatively behindhand. To achieve 3D display, most simulators use 3D glasses; these can slightly change the color of the oral tissues and can produce unpleasant effects such as vertigo and nausea [45,58]. In addition, 3D glasses change the image quality to a certain extent, reducing the resolution of the image. The display of oral tissue needs to be precise since the dental operation needs to identify subtle differences between different kinds of tissues, so a lower resolution may result in inaccuracy in the operation [59]. Image display can also affect the VR immersiveness, so it is necessary for researchers to find a higher resolution of 3D display to replace 3D glasses.

No Fixed Physical Finger Rest

A stable finger rest is of great significance in dental operations because of the small intraoral space and the requirement for precise dental operations [15,43]; without it, accidental injury to the surrounding soft and hard tissues may occur. Training on the use of finger rests is crucial in dental skills training, so finger rests should be provided for optimal simulation during dental skill training.

Lack of Bimanual Cooperative Operation

Bimanual operation is extremely important in dental operations. In general, the left hand of the operator manipulates an odontoscope to stretch and protect the soft tissues and reflect light, to ensure accurate and safe intraoral manipulation of the instrument held in the right hand. For optimal simulation of an operation on a real patient, training on bimanual operation in a dental simulator is essential [53,60].

Disadvantages of Software Design

Simulation of Force Feedback Is Not Realistic Enough

In clinical operations, dentists continually judge the process to determine whether to continue the operation by perceiving the different force feedback of the different oral tissues. Therefore, the fidelity of force feedback in the simulator is vital to dental training [27]. At present, the force feedback of oral simulators is based on a force-generating algorithm. This is different from the force feedback of a real instrument in contact with the oral tissue, especially soft tissue. Consequently, researchers need to work on improving the fidelity of force feedback in the future [14,61].

Simulation of Soft Tissue Deformation Is Not Realistic Enough

An oral environment simulation includes simulation of tongue and facial tissue deformation [62]. When a simulated surgical instrument collides with the deformable object, the object needs to deform accordingly. Deformation simulation is based on the physical properties of the soft body, such as density and elasticity. To express these properties, a physical model must be established; these commonly include the mass-spring method [63,64] and the finite element method [65]. The mass-spring method is simple and commonly used, but it is hard to maintain tissue volume information and set the stiffness of the spring. The drawback of the finite element method is that it is hard to use in real-time simulation [65]. Therefore, a physical model with higher calculation efficiency and more accurate simulation that can better represent the physical characteristics of oral tissues needs to be established.

The Training Content Is Insufficient

At present, many dental simulators do not have different degrees of difficulty in training, so students cannot learn step by step in the process of training. Moreover, the simulators have only basic skills training and lack the application of skills in clinical cases, so they cannot assess students' progressive mastery of skills. More comprehensive and systematic training content should be developed in the future [15].

The Evaluation of Training Results Cannot Be Accurately Quantified

It has been reported that professional instruction and performance feedback are beneficial for students' skills acquisition [66,67]. Some simulators cannot give an accurate quantitative evaluation of the student's operational results after training, so students do not know whether their operation results meet the requirement or not.

In summary, a good dental skills training simulator should be able to overcome these shortcomings, and its comprehensive features are summarized in Figures 4 and 5.

Figure 4. Hardware facilities required for a dental simulator. CAVE: Cave Automatic Virtual Environment; HMD: head-mounted display.

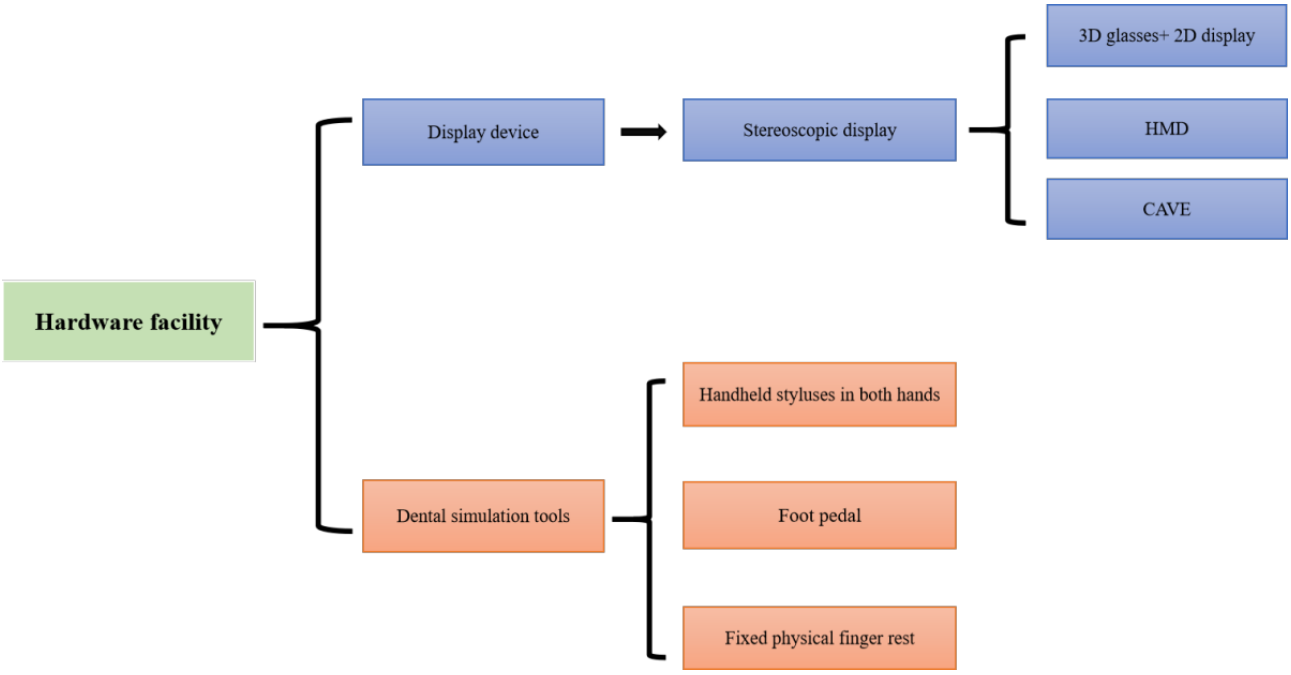


Figure 5. Software design required for an ideal dental simulator.

Effectiveness Evaluation of Dental Simulators

Methods for Effectiveness Evaluation

The effectiveness of VR systems can be evaluated using two analytical methods, qualitative and quantitative [15,43], which can be analyzed by questionnaires and comparative studies, respectively.

Questionnaires

Questionnaires are mainly used to investigate the subjective experience of people using the equipment, providing possible ways to improve the next iteration of the simulator. The content of the questionnaire is mainly based on the relationship between the user and the simulator. The simulator has two categories of users: trainers and trainees. For trainers, the device is a teaching tool and therefore needs to be evaluated for its auxiliary role in teaching, such as whether it reduces teaching cost or improves teaching efficiency. For trainees, the device is a learning tool and needs to be evaluated for learning effectiveness and construction of a learning environment [68]. Therefore, the content of the questionnaire should be designed according to the characteristics of the respondents.

The content of the questionnaire is mainly classified into two categories. The first category is the evaluation of the software and hardware of the device, involving force feedback, 3D modeling, the ease and convenience of using the device, the simulation degree, and the design of the device. Wang et al proposed a qualitative evaluation architecture based on the analysis of function components, which included the performance and usability of the simulator [15]. The second

category is the evaluation of the teaching effect of the device, which involves the subjective evaluation of the equipment effect by users [69]. The main questions in such questionnaires are whether the students think that the simulator assisted their study, improving their understanding of the related curriculum, and whether they are willing to use the simulator in their future study [70]. In addition, Venkatesh et al made the Unified Theory of Acceptance and Use of Technology questionnaire proposal to measure the acceptance of a new technology [71], and Bravo et al applied it to the use of dental simulators and concluded that it can assess the acceptance of dental simulators in dental education [72].

Comparative Studies

Comparative studies are mainly suitable for quantitative evaluation of objective function of a simulator, such as evaluating the role of the simulator in education compared with traditional teaching methods [73,74], the accuracy of the simulator at evaluating different kinds of trainees [46,75], and its accuracy in predicting the skill levels of students [76]. Barsom et al proposed a matrix of validity type to train medical professionals and classified the existing research methods including surface validity, content validity, construct validity, concurrent validity, and predictive validity [77,78].

Construct validity is the most widely used criterion in the functional evaluation of simulators [79]. The objects of the study of construct validity are populations with different clinical experience and skills in a corresponding specialty, such as experts and students [80]. By comparing the relevant parameters of using the simulator in different groups, construct validity of the simulator aims to distinguish different levels of dental skills. Dental trainees with diverse training time in a study by Mirghani et al [75] and dental students in their first year and experienced dentists in a study by Eve et al [45] were chosen to be trained using the simulator to evaluate its construct validity.

Influence Factors

Dental skills training takes the form of gradual acquisition, so the proficiency of training using the simulator can improve along with increasing familiarity with the training content. Urbankova et al concluded that repeating the training led to learning the test content, which could affect the results of the study [40]. Since VR is a new technology, teachers' sensitivity to the simulator may depend on their age, which could affect the results of the experiment. Therefore, questionnaires are often designed to eliminate the influence of these two factors [81].

Training and demonstration before the experiment, which allow objects to know the experimental flow and operational specification and eliminate the interference factors, are necessary [47]. Seymour et al tested trainees' visual ability, perception ability, and mental condition to exclude interference factors that affected experimental results [82].

Effectiveness of Dental Simulators

To date, many pilot tests have been conducted to evaluate the validity of dental simulators in endodontics [3,83], periodontics [12,42,43,84,85], oral and maxillofacial surgery [65,86-88], dental radiography [41], prosthodontics [37,59,85], implantology [89,90], and orthodontics [90]; the specific applications of dental simulators in dental education are summarized in Table 2.

Table 2. Applications of dental simulators in dental fields.

Dental fields	Training contents
Endodontics	• Dental caries detection • Caries removal • Light-curing skills • Endodontic cavity preparation
Periodontics	• Periodontal probing • Calculus detection and removal • Ultrasonic scaling
Oral and maxillofacial surgery	• Dental anesthesia training • Maxillofacial palpation • Dental extraction skills • Orthognathic surgery • Dental surgery
Dental radiography	Intraoral X-ray imaging
Prosthodontics	Tooth preparation
Implantology	• Implanting scheme design • Implant drilling skills
Orthodontics	Training and treatment planning in orthodontics

Many researchers are positive about the roles that dental simulators play in skills training. Al-Saud et al reported that students could gain basic manual dexterity at a quicker pace when they practiced skills with the guidance of experienced faculty, and training with dental simulators helped students to retain the skills they had learned [91]. A study performed by Plessas showed that students could develop a better understanding of the material by using dental simulators since they create a more varied learning environment compared with traditional training methods [74]. Suebnukarn et al evaluated

the effectiveness of dental simulators in cavity preparation and concluded that haptic VR simulators are equivalent to conventional dental phantom heads in reducing operation errors [92]. Other research provided evidence that students can improve their tooth extraction skills [93], dental radiology skills [47], and implant skills [32,90] through dental simulators. Dental simulators were also reported to be able to enhance students' confidence [94] and improve their attitudes toward patients [53] and abilities to discern and solve medical emergencies [70,95,96]. Some studies focused on students' acceptance of dental simulators and found that most students are willing to learn with dental simulators, which would boost their enthusiasm to learn [97]. Therefore, many research studies indicated that dental simulators had the potential to be an alternative to conventional training methods [36,52,98].

However, as mentioned above, there are still some disadvantages of dental simulators that can't be ignored, and these disadvantages may directly influence the effectiveness of dental simulators. In addition, the effectiveness of some dental simulators has not been validated [88,98]. Therefore, it is suggested that dental simulators cannot completely replace traditional skill training methods. The automated evaluation and tutorial feedback offered by dental simulators are considered to be complementary components to traditional methods [84]. Some studies concluded that a combination of traditional and virtual methods would be an optimal approach to choose in skills training [52,99,100].

Prospects of Dental Education Simulators

Current simulators are deficient in stereovision, video resolution, force feedback, instructional content, and outcome assessment; in response to these issues, we offer an outlook for their future development.

Visualized Analyses of Education Data Based on Big Data Technology

By applying big data technology to medical education, we can establish statistical models by the visual analysis method [101] via educational data mining and learning analysis [102], which allow trainers to analyze and understand the learning status of trainees intuitively. Currently, big data technology has been applied in educational analysis of massive open online courses [103] and learning effectiveness prediction for medical courses [101].

A similar big data analysis platform can be set up in dental skills training. We can collect dental operation data and establish a database of correctly performed operations. Then, a scientific evaluation system to score the training results by comparing them with the correctly performed operations can be created. Trainers can identify information on general problems in operations, as well as problems experienced by the individual trainee, to analyze and predict trainees' ability to master the skills, error-prone points, and the pass rate of examinations. Based on the analysis, trainers can create or adjust different plans for all trainees to enhance their learning experience and performance. Therefore, the combination of big data and education will make it possible to deeply understand and study the training process in order to provide more trainees with quality training.

Video Transmission With Quicker Speeds and Less Delay Supported by 5G

The development of 5G technology will allow transmission links with high bandwidth and low delay [104-106]. In the virtual oral training system, calculations are often performed on the server side to improve accuracy. The use of 5G technology will greatly increase the transmission rate of high-definition video data from the server host to the display device, reducing the response delay [107-109]. This will provide users with more reliable visual and tactile feedback and improve comfort [110].

Currently, the 5G quality of experience framework is proposed to solve the problem of transmitting ultra-high-definition video on the 5G network [111]. This could provide a basis for the establishment of a VR dental educational video network. Large-scale synchronous training and one-to-one demonstrations could be performed via 5G transmission to improve training efficiency.

Improvement of the Simulation of Force Feedback by Deep Learning

Deep learning techniques are widely used to find rules from a large amount of data, abstract problems through neural networks, and establish input and output mapping.

In the field of haptics, deep learning has been used to obtain haptic properties of objects from their images [112], whereby force feedback devices can directly generate a sense of force feedback that mimics the real physical properties without having to manually adjust physical parameters.

We can use deep learning to obtain the force feedback properties of teeth and build different deep learning models for different parts of the teeth. Therefore, the instrument would have different force feedback in contact with normal teeth, dental caries, tooth enamel, dentin, and other areas.

Deep learning can also play a role in oral modeling. Now, deep learning has been widely used in medical image analysis and modeling [113]. Through this type of method, we can easily extract relevant regions from computed tomography or MRI data and reconstruct mesh data [114]. These technologies can be transferred to oral cavity modeling, which can greatly facilitate model construction in a virtual surgery system and quickly establish a personalized oral cavity model.

Improvement of the Immersiveness of Dental Simulators by Augmented Reality

Augmented reality technology is an extension of VR technology, combining virtual images with real environments to support interaction with virtual objects in real time.

Nowadays, augmented reality technology has begun to be used in the medical field. Based on an augmented reality helmet (HoloLens by Microsoft), various applications have been developed such as surgical navigation applications [115] and intelligent medical management systems [116]. HoloLens also

supports a display of high-definition video derived from high-performance computing.

The virtual tooth and the phantom head can be combined with augmented reality technology in a dental training system. Students can feel the phantom head directly with their hands and interact with a virtual tooth through a force feedback device. At the same time, high-definition images allow users to observe finer tooth details, which can greatly improve the reality and the immersiveness of the experience. The users can also observe patients from multiple views, which will greatly enhance the trainees' understanding and learning of training objectives [117,118]. Therefore, augmented reality will have a significant effect on dental education in the future.

Strengths and Limitations

The strengths of this paper are as follows: First, we have come up with a summary of comprehensive features that a good dental simulator should possess based on the analysis of the advantages and disadvantages of current dental simulators. Second, this paper proposes the future prospects of dental simulators and provides researchers with new ideas for further studies.

Our paper also has some limitations. First, our paper lists the nine dental simulators that are commonly mentioned and studied in most articles instead of all kinds of dental simulators, which may have a little effect on the results of our study. Second, our study is focused on VR- and haptic-based dental simulators; as a consequence, some relevant papers may not be included due to inappropriate use of keywords.

Conclusion

Despite the fact that dental simulators are not currently able to rival traditional training modalities for skills training in some disciplines, they still have some advantages over traditional methods, and their effectiveness has been validated in some cases. More studies should be conducted to improve the force feedback, video transmission, and immersiveness of dental simulators. With scientific and technological development, dental simulators that gradually combine with big data, cloud computing, 5G, and deep learning technology will offer students individualized learning assistance, and their functions will be more diverse and suitable for preclinical training.

Acknowledgments

This work is supported by the Teaching Reformation Fund of Peking University School and Hospital of Stomatology (2018-ZD-04, HY; 2019-ZD-07, Y Liu), and all authors thank Beijing Unidraw VR Technology Research Institute Co Ltd for providing Figure 1.

Authors' Contributions

Y Li and FY designed the structure of this manuscript and wrote the paper; Y Liu, LL, PZ and XZ collected the data and revised the manuscript; and HY and YZ conceived and revised the manuscript. All authors approved the manuscript.

Conflicts of Interest

None declared.

References

1. Arigbede A, Denloye O, Dosumu O. Use of simulators in operative dental education: experience in southern Nigeria. Afr Health Sci 2015 Mar;15(1):269-277 [FREE Full text] [doi: 10.4314/ahs.v15i1.35] [Medline: 25834558]
2. Perry S, Bridges SM, Burrow MF. A review of the use of simulation in dental education. Simul Healthc 2015 Feb;10(1):31-37. [doi: 10.1097/SIH.0000000000000059] [Medline: 25574865]
3. Coro Montanet G, Gómez Sánchez M, Suárez García A. Haptic simulators withvirtual reality environments in dental education: A preliminaryteaching diagnosis. @tic 2017 Jun 23(18):14-21. [doi: 10.7203/attic.18.9077]
4. Serrano CM, Wesselink PR, Vervoorn JM. First experiences with patient-centered training in virtual reality. J Dent Educ 2020 May;84(5):607-614. [doi: 10.1002/jdd.12037] [Medline: 31971611]
5. Xia P, Lopes AM, Restivo MT. Virtual reality and haptics for dental surgery: a personal review. Vis Comput 2012 Sep 8;29(5):433-447. [doi: 10.1007/s00371-012-0748-2]
6. Vannaprathip N, Haddawy P, Schultheis H, Suebnukarn S. Generating tutorial interventions for teaching situation awareness in dental surgery – preliminary report. 2017 Presented at: International Workshop on Multi-disciplinary Trends in Artificial Intelligence; 20-22 November; Gadong, Brunei Darussalam.
7. Pavaloiu I, Ioanitescu R, Dragoi G, Grigorescu S, Sandu S. Virtual reality for educationtraining in dentistry. In: eLearning and Software for Education. Bucharest: Carol I Natl Defence Univ Publishing House; 2016.
8. Freina L, Ott M. A literature review on immersive virtual reality in education: state of the art and perspectives. 2015 Presented at: The International Scientific Conference eLearning and Software for Education; April 2015; Bucharest (Romania) p. 23-24.
9. Ferguson MB, Sobel M, Niederman R. Preclinical restorative training. J Dent Educ 2002 Oct;66(10):1159-1162. [Medline: 12449210]

10. Rhienmora P, Haddawy P, Suebnukarn S, Dailey M. Providing objective feedback on skill assessment in a dental surgical training simulator. 2009 Presented at: Conference on Artificial Intelligence in Medicine in Europe; 18-22 July; Verona, Italy.
11. de Boer IR, Wesselink P, Vervoorn J. The creation of virtual teeth with and without tooth pathology for a virtual learning environment in dental education. Eur J Dent Educ 2013 Nov;17(4):191-197. [doi: 10.1111/eje.12027] [Medline: 24127759]
12. Luciano C, Banerjee P, DeFanti T. Haptics-based virtual reality periodontal training simulator. Virtual Reality 2009 Feb 13;13(2):69-85. [doi: 10.1007/s10055-009-0112-7]
13. de Boer IR, Bakker DR, Wesselink PR, Vervoorn JM. [The Simodont in dental education]. Ned Tijdschr Tandheelkd 2012 Jun;119(6):294-300. [doi: 10.5177/ntvt.2012.06.12105] [Medline: 22812267]
14. Escobar-Castillejos D, Noguez J, Neri L, Magana A, Benes B. A Review of Simulators with Haptic Devices for Medical Training. J Med Syst 2016 Apr;40(4):104. [doi: 10.1007/s10916-016-0459-8] [Medline: 26888655]
15. Wang D, Li T, Zhang Y, Hou J. Survey on multisensory feedback virtual reality dental training systems. Eur J Dent Educ 2016 Nov;20(4):248-260. [doi: 10.1111/eje.12173] [Medline: 26547278]
16. Jayaram S, Connacher HI, Lyons KW. Virtual assembly using virtual reality techniques. Computer-Aided Design 1997 Aug;29(8):575-584. [doi: 10.1016/S0010-4485(96)00094-2]
17. Steuer J. Defining Virtual Reality: Dimensions Determining Telepresence. J Commun 1992;42(4):73-93. [doi: 10.1111/j.1460-2466.1992.tb00812.x]
18. Ayoub A, Pulijala Y. BMC Oral Health 2019 Nov 08;19(1):238 [FREE Full text] [doi: 10.1186/s12903-019-0937-8] [Medline: 31703708]
19. Rolland J, Hua H. Head-Mounted Display Systems. 2005. URL: https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.465.4293&rep=rep1&type=pdf [accessed 2021-03-18]
20. Cruz-Neira C, Sandin D, Defanti T. Surround-screen projection-based virtual reality: the design and implementation of the CAVE. : Association for Computing Machinery; 1993 Presented at: 20th Annual Conference and Exhibition on Computer Graphics and Interactive Techniques; August 1993; Anaheim CA. [doi: 10.1145/166117.166134]
21. Cipresso P, Giglioli IAC, Raya MA, Riva G. The Past, Present, and Future of Virtual and Augmented Reality Research: A Network and Cluster Analysis of the Literature. Front Psychol 2018;9:2086 [FREE Full text] [doi: 10.3389/fpsyg.2018.02086] [Medline: 30459681]
22. Shahrbanian S, Ma X, Aghaei N, Korner-Bitensky N, Simmonds M. Use of virtual reality (immersive vs. non immersive) for pain management in children and adults: A systematic review of evidence from randomized controlled trials. Eur J Exp Biol 2012;2(5):1408-1422.
23. Vasco MAA, Souza JTAD, Las Casas EBD, de Castro e Silva ALR, Hecke M. A method for constructing teeth and maxillary bone parametric model from clinical CT scans. Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization 2014 Mar 13;3(3):117-122. [doi: 10.1080/21681163.2014.889579]
24. Quan T, Han S, Cho H, Jeong W. Multi-GPU reconstruction of dynamic compressed sensing MRI. : Springer; 2015 Presented at: International Conference on Medical Image Computing and Computer-Assisted Intervention; Springer; 5-9 October 2015; Munich, Germany.
25. Lian C, Wang L, Wu T, Wang F, Yap P, Ko C, et al. Deep Multi-Scale Mesh Feature Learning for Automated Labeling of Raw Dental Surfaces From 3D Intraoral Scanners. IEEE Trans Med Imaging 2020 Jul;39(7):2440-2450. [doi: 10.1109/TMI.2020.2971730] [Medline: 32031933]
26. Gurusamy K, Aggarwal R, Palanivelu L, Davidson B. Virtual reality training for surgical trainees in laparoscopic surgery. Cochrane Database Syst Rev 2009 Jan 21(1):CD006575. [doi: 10.1002/14651858.CD006575.pub2] [Medline: 19160288]
27. Haluck RS, Webster RW, Snyder AJ, Melkonian MG, Mohler BJ, Dise ML, et al. A virtual reality surgical trainer for navigation in laparoscopic surgery. Stud Health Technol Inform 2001;81:171-176. [Medline: 11317733]
28. Riva G. Applications of virtual environments in medicine. Methods Inf Med 2003;42(5):524-534. [doi: 10.1055/s-0038-1634379] [Medline: 14654887]
29. de Boer IR, Lagerweij MD, de Vries MW, Wesselink PR, Vervoorn JM. The Effect of Force Feedback in a Virtual Learning Environment on the Performance and Satisfaction of Dental Students. Simul Healthc 2017 Apr;12(2):83-90. [doi: 10.1097/SIH.0000000000000208] [Medline: 28704286]
30. Wang D, Zhao X, Shi Y, Zhang Y, Hou J, Xiao J. Six Degree-of-Freedom Haptic Simulation of Probing Dental Caries Within a Narrow Oral Cavity. IEEE Trans Haptics 2016;9(2):279-291. [doi: 10.1109/TOH.2016.2531660] [Medline: 26915130]
31. Wang D, Shi Y, Liu S, Zhang Y, Xiao J. Haptic simulation of organ deformation and hybrid contacts in dental operations. IEEE Trans Haptics 2014 Mar;7(1):48-60. [doi: 10.1109/TOH.2014.2304734] [Medline: 24845745]
32. Kinoshita H, Nagahata M, Takano N, Takemoto S, Matsunaga S, Abe S, et al. Development of a Drilling Simulator for Dental Implant Surgery. J Dent Educ 2016 Jan;80(1):83-90. [Medline: 26729688]
33. Zhao X, Zhu Z, Cong Y, Zhao Y, Zhang Y, Wang D. Haptic Rendering of Diverse Tool-Tissue Contact Constraints During Dental Implantation Procedures. Front Robot AI 2020 Mar 20;7:35 [FREE Full text] [doi: 10.3389/frobt.2020.00035] [Medline: 33501203]

34. Buchanan JA. Use of simulation technology in dental education. J Dent Educ 2001 Nov;65(11):1225-1231. [Medline: 11765868]
35. Roy E, Bakr MM, George R. The need for virtual reality simulators in dental education: A review. Saudi Dent J 2017 Apr;29(2):41-47 [FREE Full text] [doi: 10.1016/j.sdentj.2017.02.001] [Medline: 28490842]
36. Urbankova A. Impact of computerized dental simulation training on preclinical operative dentistry examination scores. J Dent Educ 2010 Apr;74(4):402-409. [Medline: 20388813]
37. Liu L, Zhou R, Yuan S, Sun Z, Lu X, Li J, et al. Simulation training for ceramic crown preparation in the dental setting using a virtual educational system. Eur J Dent Educ 2020 May;24(2):199-206. [doi: 10.1111/eje.12485] [Medline: 31766075]
38. Liu L, Li J, Yuan S, Wang T, Chu F, Lu X, et al. Evaluating the effectiveness of a preclinical practice of tooth preparation using digital training system: A randomised controlled trial. Eur J Dent Educ 2018 Nov;22(4):e679-e686. [doi: 10.1111/eje.12378] [Medline: 29952122]
39. Urbankova A, Engebretson SP. The use of haptics to predict preclinic operative dentistry performance and perceptual ability. J Dent Educ 2011 Dec;75(12):1548-1557. [Medline: 22184593]
40. Urbankova A, Eber M, Engebretson SP. A complex haptic exercise to predict preclinical operative dentistry performance: a retrospective study. J Dent Educ 2013 Nov;77(11):1443-1450. [Medline: 24192409]
41. Ben-Gal G, Weiss E, Gafni N, Ziv A. Testing manual dexterity using a virtual reality simulator: reliability and validity. Eur J Dent Educ 2013 Aug;17(3):138-142. [doi: 10.1111/eje.12023] [Medline: 23815690]
42. Steinberg AD, Bashook PG, Drummond J, Ashrafi S, Zefran M. Assessment of faculty perception of content validity of PerioSim, a haptic-3D virtual reality dental training simulator. J Dent Educ 2007 Dec;71(12):1574-1582. [Medline: 18096883]
43. Dangxiao Wang, Yuru Zhang, Jianxia Hou, Yong Wang, Peijun Lv, Yonggang Chen, Hui Zhao. iDental: A Haptic-Based Dental Simulator and Its Preliminary User Evaluation. IEEE Trans Haptics 2012;5(4):332-343. [doi: 10.1109/TOH.2011.59] [Medline: 26964131]
44. Wang F, Liu Y, Tian M, Zhang Y, Chen J. Application of a 3D Haptic Virtual Reality Simulation System for Dental Crown Preparation Training. 2017 Presented at: 8th International Conference on Information Technology in Medicine and Education (ITME); 23-25 Dec 2016; Fuzhou, China.
45. Eve EJ, Koo S, Alshihri AA, Cormier J, Kozhenikov M, Donoff RB, et al. Performance of dental students versus prosthodontics residents on a 3D immersive haptic simulator. J Dent Educ 2014 Apr;78(4):630-637. [Medline: 24706694]
46. Cormier J, Pasco D, Syllebranque C, Querrec R. VirTeaSy a haptic simulator for dental education. Bucharest: Bucharest University Press; 2011 Presented at: Proceedings of the 6th International Conference on Virtual Learning; Oct 28-29, 2011; Babes Bolyai Univ Cluj Napoca, Cluj Napoca, Romania.
47. Kim S, Min J, Kim H. The Development of an IoT-Based Educational Simulator for Dental Radiography. IEEE Access 2019;7:12476-12483. [doi: 10.1109/access.2019.2891569]
48. Pohlenz P, Gröbe A, Petersik A, von Sternberg N, Pflesser B, Pommert A, et al. Virtual dental surgery as a new educational tool in dental school. J Craniomaxillofac Surg 2010 Dec;38(8):560-564. [doi: 10.1016/j.jcms.2010.02.011] [Medline: 20303282]
49. Heiland M, Petersik A, Pflesser B, Tiede U, Schmelzle R, Höhne K. Realistic haptic interaction for computer simulation of dental surgery. Amsterdam: Elsevier; 2004 Presented at: The 18th International Congress and Exhibition on Computer Assisted Radiology and Surgery; June 23-26, 2004; Chicago, IL.
50. Nassar HM, Tekian A. Computer simulation and virtual reality in undergraduate operative and restorative dental education: A critical review. J Dent Educ 2020 Jul;84(7):812-829. [doi: 10.1002/jdd.12138] [Medline: 32147841]
51. Albuha Al-Mussawi RM, Farid F. Computer-Based Technologies in Dentistry: Types and Applications. J Dent (Tehran) 2016 Jun;13(3):215-222 [FREE Full text] [Medline: 28392819]
52. Quinn F, Keogh P, McDonald A, Hussey D. A study comparing the effectiveness of conventional training and virtual reality simulation in the skills acquisition of junior dental students. Eur J Dent Educ 2003 Nov;7(4):164-169. [doi: 10.1034/j.1600-0579.2003.00309.x] [Medline: 14753762]
53. Welk A, Splieth C, Rosin M, Kordass B, Meyer G. DentSim - a future teaching option for dentists. Int J Comput Dent 2004 Apr;7(2):123-130. [Medline: 15516090]
54. Dixon J, Towers A, Martin N, Field J. Re-defining the virtual reality dental simulator: Demonstrating concurrent validity of clinically relevant assessment and feedback. Eur J Dent Educ 2021 Feb;25(1):108-116. [doi: 10.1111/eje.12581] [Medline: 32780500]
55. Cheng A, Kessler D, Mackinnon R, Chang TP, Nadkarni VM, Hunt EA, International Network for Simulation-based Pediatric Innovation, Research, and Education (INSPIRE) Reporting Guidelines Investigators. Reporting Guidelines for Health Care Simulation Research: Extensions to the CONSORT and STROBE Statements. Simul Healthc 2016 Aug;11(4):238-248. [doi: 10.1097/SIH.0000000000000150] [Medline: 27465839]
56. Abe S, Noguchi N, Matsuka Y, Shinohara C, Kimura T, Oka K, et al. Educational effects using a robot patient simulation system for development of clinical attitude. Eur J Dent Educ 2018 Aug;22(3):e327-e336. [doi: 10.1111/eje.12298] [Medline: 29091328]
57. Jasinevicius TR, Landers M, Nelson S, Urbankova A. An evaluation of two dental simulation systems: virtual reality versus contemporary non-computer-assisted. J Dent Educ 2004 Nov;68(11):1151-1162. [Medline: 15520234]

58. Polcar J, Horejsi P. Knowledge acquisition and cyber sickness: A comparison of VR devices in virtual tours. MM SJ 2015 Jun 03;2015(02):613-616. [doi: 10.17973/mmsj.2015_06_201516]
59. Jung H, Kim H, Moon S. Virtual reality training simulator for tooth preparation techniques. Oral Biol Res 2018 Dec 31;42(4):235-240. [doi: 10.21851/obr.42.04.201812.235]
60. Yamaguchi S, Kazumichi W, Yoshida Y, Nagashima T, Takeshige F, Kawamoto Y, et al. Dental haptic simulator to train hand skill of student - calibration method to realize visual haptic environment. Stud Health Technol Inform 2009;142:429-431. [Medline: 19377200]
61. Ruthenbeck GS, Reynolds KJ. Virtual reality for medical training: the state-of-the-art. Journal of Simulation 2017 Dec 19;9(1):16-26. [doi: 10.1057/jos.2014.14]
62. Keeve E, Girod S, Kikinis R, Girod B. Deformable modeling of facial tissue for craniofacial surgery simulation. Comput Aided Surg 1998;3(5):228-238. [doi: 10.1002/(SICI)1097-0150(1998)3:5<228::AID-IGS2>3.0.CO;2-I] [Medline: 10207647]
63. Mesit J, Guha R, Chaudhry S. 3D Soft Body Simulation Using Mass-spring System with Internal Pressure Force and Simplified Implicit Integration. JCP 2007 Oct 01;2(8):34-43. [doi: 10.4304/jcp.2.8.34-43]
64. Teschner M, Girod S, Girod B. Direct Computation of Nonlinear Soft-Tissue Deformation. URL: https://static.aminer.org/pdf/PDF/000/647/215/direct_computation_of_nonlinear_soft_tissue_deformation.pdf [accessed 2021-03-18]
65. Nisansala A, Weerasinghe M, Dias G, Sandaruwan D, Kodikara N, Technology I. Soft tissue modeling techniques in surgery simulation. Int I Comput Inf Sci 2015;4(5):826-831.
66. Lazarski M, Susarla S, Bennett N, Seldin E. How do feedback and instructions affect the performance of a simulated surgical task? J Oral Maxillofac Surg 2007 Jun;65(6):1155-1161. [doi: 10.1016/j.joms.2006.11.021] [Medline: 17517300]
67. Wierinck E, Puttemans V, van Steenberghe D. Effect of tutorial input in addition to augmented feedback on manual dexterity training and its retention. Eur J Dent Educ 2006 Feb;10(1):24-31. [doi: 10.1111/j.1600-0579.2006.00392.x] [Medline: 16436081]
68. Espejo-Trung LC, Elian SN, Luz MAADC. Development and Application of a New Learning Object for Teaching Operative Dentistry Using Augmented Reality. J Dent Educ 2015 Nov;79(11):1356-1362. [Medline: 26522642]
69. Sabalic M, Schoener JD. Virtual Reality-Based Technologies in Dental Medicine: Knowledge, Attitudes and Practice Among Students and Practitioners. Tech Know Learn 2017 Feb 8;22(2):199-207. [doi: 10.1007/s10758-017-9305-4]
70. Kishimoto N, Mukai N, Honda Y, Hirata Y, Tanaka M, Momota Y. Simulation training for medical emergencies in the dental setting using an inexpensive software application. Eur J Dent Educ 2018 Aug;22(3):e350-e357. [doi: 10.1111/eje.12301] [Medline: 29120509]
71. Venkatesh V, Thong JYL, Xu X. Consumer Acceptance and Use of Information Technology: Extending the Unified Theory of Acceptance and Use of Technology. MIS Quarterly 2012 Mar;36(1):157-117. [doi: 10.2307/41410412] [Medline: 22217802]
72. González Bravo L, Fernández Sagredo M, Torres Martínez P, Barrios Penna C, Fonseca Molina J, Stanciu ID, et al. Psychometric analysis of a measure of acceptance of new technologies (UTAUT), applied to the use of haptic virtual simulators in dental students. Eur J Dent Educ 2020 Nov;24(4):706-714. [doi: 10.1111/eje.12559] [Medline: 32567797]
73. Quinn F, Keogh P, McDonald A, Hussey D. A study comparing the effectiveness of conventional training and virtual reality simulation in the skills acquisition of junior dental students. Eur J Dent Educ 2003 Nov;7(4):164-169. [doi: 10.1034/j.1600-0579.2003.00309.x] [Medline: 14753762]
74. Plessas A. Computerized Virtual Reality Simulation in Preclinical Dentistry: Can a Computerized Simulator Replace the Conventional Phantom Heads and Human Instruction? Simul Healthc 2017 Oct;12(5):332-338. [doi: 10.1097/SIH.0000000000000250] [Medline: 28697057]
75. Mirghani I, Mushtaq F, Allsop MJ, Al-Saud LM, Tickhill N, Potter C, et al. Capturing differences in dental training using a virtual reality simulator. Eur J Dent Educ 2018 Feb;22(1):67-71. [doi: 10.1111/eje.12245] [Medline: 27864856]
76. Imber S, Shapira G, Gordon M, Judes H, Metzger Z. A virtual reality dental simulator predicts performance in an operative dentistry manikin course. Eur J Dent Educ 2003 Nov;7(4):160-163. [doi: 10.1034/j.1600-0579.2003.00299.x] [Medline: 14753761]
77. Barsom EZ, Graafland M, Schijven MP. Systematic review on the effectiveness of augmented reality applications in medical training. Surg Endosc 2016 Oct;30(10):4174-4183 [FREE Full text] [doi: 10.1007/s00464-016-4800-6] [Medline: 26905573]
78. Schijven MP, Jakimowicz JJ. Validation of virtual reality simulators: Key to the successful integration of a novel teaching technology into minimal access surgery. Minim Invasive Ther Allied Technol 2005;14(4):244-246. [doi: 10.1080/13645700500221881] [Medline: 16754170]
79. van Ginkel MPH, Schijven MP, van Grevenstein WMU, Schreuder HWR. Bimanual Fundamentals: Validation of a New Curriculum for Virtual Reality Training of Laparoscopic Skills. Surg Innov 2020 Oct;27(5):523-533 [FREE Full text] [doi: 10.1177/1553350620953030] [Medline: 32865136]
80. Suebnukarn S, Phatthanasathiankul N, Sombatweroje S, Rhienmora P, Haddawy P. Process and outcome measures of expert/novice performance on a haptic virtual reality system. J Dent 2009 Sep;37(9):658-665. [doi: 10.1016/j.jdent.2009.04.008] [Medline: 19481326]
81. Gal GB, Weiss EI, Gafni N, Ziv A. Preliminary assessment of faculty and student perception of a haptic virtual reality simulator for training dental manual dexterity. J Dent Educ 2011 Apr;75(4):496-504. [Medline: 21460270]

82. Seymour NE, Gallagher AG, Roman SA, O'Brien MK, Bansal VK, Andersen DK, et al. Virtual reality training improves operating room performance: results of a randomized, double-blinded study. Ann Surg 2002 Oct;236(4):458-63; discussion 463. [doi: 10.1097/00000658-200210000-00008] [Medline: 12368674]
83. Yin MS, Haddawy P, Suebnukarn S, Rhienmora P. Automated outcome scoring in a virtual reality simulator for endodontic surgery. Comput Methods Programs Biomed 2018 Jan;153:53-59. [doi: 10.1016/j.cmpb.2017.10.001] [Medline: 29157461]
84. Yamaguchi S, Yoshida Y, Noborio H, Murakami S, Imazato S. The usefulness of a haptic virtual reality simulator with repetitive training to teach caries removal and periodontal pocket probing skills. Dent Mater J 2013;32(5):847-852 [FREE Full text] [doi: 10.4012/dmj.2013-174] [Medline: 24088844]
85. Kim K, Cho J, Kim J, Park J. A Dental Simulator for Training of Prevalent Interventions: Tooth Restoration and Ultrasonic Scaling. Berlin, Heidelberg: Springer; 2012 Presented at: Haptics: Perception, Devices, Mobility, and Communication; 13-15 June 2012; Tampere, Finland.
86. Mladenovic R, Dakovic D, Pereira L, Matvijenko V, Mladenovic K. Effect of augmented reality simulation on administration of local anaesthesia in paediatric patients. Eur J Dent Educ 2020 Aug;24(3):507-512. [doi: 10.1111/eje.12529] [Medline: 32243051]
87. Maliha SG, Diaz-Siso JR, Plana NM, Torroni A, Flores RL. Haptic, Physical, and Web-Based Simulators: Are They Underused in Maxillofacial Surgery Training? Journal of Oral and Maxillofacial Surgery 2018 Nov;76(11):2424.e1-2424.e11. [doi: 10.1016/j.joms.2018.06.177] [Medline: 30081008]
88. Chen X, Sun P, Liao D. A patient-specific haptic drilling simulator based on virtual reality for dental implant surgery. Int J Comput Assist Radiol Surg 2018 Nov;13(11):1861-1870. [doi: 10.1007/s11548-018-1845-0] [Medline: 30097957]
89. Joseph D, Jehl J, Maureira P, Perrenot C, Miller N, Bravetti P, et al. Relative contribution of haptic technology to assessment and training in implantology. Biomed Res Int 2014;2014:413951 [FREE Full text] [doi: 10.1155/2014/413951] [Medline: 24701577]
90. Rodrigues MAF, Silva WB, Barbosa Neto ME, Gillies DF, Ribeiro IMMP. An interactive simulation system for training and treatment planning in orthodontics. Computers & Graphics 2007 Oct;31(5):688-697. [doi: 10.1016/j.cag.2007.04.010]
91. Al-Saud L, Mushtaq F, Allsop M, Culmer P, Mirghani I, Yates E, et al. Feedback and motor skill acquisition using a haptic dental simulator. Eur J Dent Educ 2017 Nov;21(4):240-247. [doi: 10.1111/eje.12214] [Medline: 27324833]
92. Suebnukarn S, Hataidechadusadee R, Suwannasri N, Suprasert N, Rhienmora P, Haddawy P. Access cavity preparation training using haptic virtual reality and microcomputed tomography tooth models. Int Endod J 2011 Nov;44(11):983-989. [doi: 10.1111/j.1365-2591.2011.01899.x] [Medline: 21623838]
93. Lee SH. Research and development of haptic simulator for Dental education using Virtual reality and User motion. The International Journal of Advanced Smart Convergence 2018;6(4):114-120. [doi: 10.7236/IJASC.2018.7.4.114]
94. Serrano C, Wesselink PR, Vervoorn JM. [Real patients in virtual reality: the link between phantom heads and clinical dentistry]. Ned Tijdschr Tandheelkd 2018 May;125(5):263-267. [doi: 10.5177/ntvt.2018.05.17192] [Medline: 29754155]
95. Roy E, Quinsat VE, Bazin O, Lesclous P, Lejus-Bourdeau C. High-fidelity simulation in training dental students for medical life-threatening emergency. Eur J Dent Educ 2018 May;22(2):e261-e268. [doi: 10.1111/eje.12284] [Medline: 28833993]
96. Sararit N, Haddawy P, Suebnukarn S. Effectiveness of a low-cost VR simulator for emergency management training in dental surgery. : IEEE; 2018 Presented at: 2018 15th International Joint Conference on Computer Science and Software Engineering (JCSSE); 11-13 July 2018; Nakhonpathom, Thailand. [doi: 10.1109/JCSSE.2018.8457353]
97. Curnier F. Teaching dentistry by means of virtual reality--the Geneva project. Int J Comput Dent 2010;13(3):251-263. [Medline: 20879463]
98. LeBlanc VR, Urbankova A, Hadavi F, Lichtenthal RM. A preliminary study in using virtual reality to train dental students. J Dent Educ 2004 Mar;68(3):378-383. [Medline: 15038639]
99. Vincent M, Joseph D, Amory C, Paoli N, Ambrosini P, Mortier É, et al. Contribution of Haptic Simulation to Analogic Training Environment in Restorative Dentistry. J Dent Educ 2020 Mar;84(3):367-376. [doi: 10.21815/JDE.019.187] [Medline: 32176342]
100. Murbay S, Neelakantan P, Chang JWW, Yeung S. 'Evaluation of the introduction of a dental virtual simulator on the performance of undergraduate dental students in the pre-clinical operative dentistry course'. Eur J Dent Educ 2020 Feb;24(1):5-16. [doi: 10.1111/eje.12453] [Medline: 31278815]
101. Soto S, Luna J, Cano A. Big Data on Real-World Applications. London, UK: IntechOpen; 2016.
102. Seufert S, Meier C, Soellner M, Rietsche R. A Pedagogical Perspective on Big Data and Learning Analytics: A Conceptual Model for Digital Learning Support. Tech Know Learn 2019 Mar 1;24(4):599-619. [doi: 10.1007/s10758-019-09399-5]
103. Tang JKT, Xie H, Wong TL. A Big Data Framework for Early Identification of Dropout Students in MOOC. Berlin: Springer; 2015 Presented at: International Conference on Technology in Education; 2-4 July 2015; Hong Kong, China.
104. Shafi M, Molisch AF, Smith PJ, Haustein T, Zhu P, De Silva P, et al. 5G: A Tutorial Overview of Standards, Trials, Challenges, Deployment, and Practice. IEEE J. Select. Areas Commun 2017 Jun;35(6):1201-1221. [doi: 10.1109/jsac.2017.2692307]
105. Gupta A, Jha RK. A Survey of 5G Network: Architecture and Emerging Technologies. IEEE Access 2015;3:1206-1232. [doi: 10.1109/access.2015.2461602]

106. Agiwal M, Roy A, Saxena N. Next Generation 5G Wireless Networks: A Comprehensive Survey. IEEE Commun Surv Tutorials 2016;18(3):1617-1655. [doi: 10.1109/comst.2016.2532458]
107. Vo N, Duong TQ, Tuan HD, Kortun A. Optimal Video Streaming in Dense 5G Networks With D2D Communications. IEEE Access 2018;6:209-223. [doi: 10.1109/access.2017.2761978]
108. Awobuluyi O, Nightingale J, Wang Q, Alcaraz-Calero J. Video Quality in 5G Networks: Context-Aware QoE Management in the SDN Control Plane. Piscataway, NJ: IEEE; 2015 Presented at: International Conference on Computer and Information Technology (CIT); 26-28 Oct. 2015; Liverpool, UK. [doi: 10.1109/CIT/IUCC/DASC/PICOM.2015.250]
109. Swetha S, Raj D. Optimized video content delivery over 5G networks. 2017 Presented at: 2017 2nd International Conference on Communication and Electronics Systems (ICCES); 19-20 Oct 2017; Coimbatore, India p. 1000-1002. [doi: 10.1109/CESYS.2017.8321232]
110. Raaen K, Kjellmo I. Measuring Latency in Virtual Reality Systems. Cham: Springer; 2015 Presented at: International Conference on Entertainment Computing; 29 September - 2 October; Trondheim, Norway.
111. Nightingale J, Salva-Garcia P, Calero JMA, Wang Q. 5G-QoE: QoE Modelling for Ultra-HD Video Streaming in 5G Networks. IEEE Trans on Broadcast 2018 Jun;64(2):621-634. [doi: 10.1109/tbc.2018.2816786]
112. Zheng H, Fang L, Ji M, Strese M, Ozer Y, Steinbach E. Deep Learning for Surface Material Classification Using Haptic and Visual Information. IEEE Trans Multimedia 2016 Dec;18(12):2407-2416. [doi: 10.1109/tmm.2016.2598140]
113. Wang S, Su Z, Ying L, Peng X, Zhu S, Liang F. Accelerating magnetic resonance imaging via deep learning. New York: IEEE; 2016 Presented at: 2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI); 13-16 April 2016; Prague, Czech Republic p. 13.
114. Georgescu M, Ionescu RT, Verga N. Convolutional Neural Networks With Intermediate Loss for 3D Super-Resolution of CT and MRI Scans. IEEE Access 2020;8:49112-49124. [doi: 10.1109/access.2020.2980266]
115. Zuo Y, Jiang T, Dou J, Yu D, Ndaro ZN, Du Y, et al. A Novel Evaluation Model for a Mixed-Reality Surgical Navigation System: Where Microsoft HoloLens Meets the Operating Room. Surg Innov 2020 Apr;27(2):193-202. [doi: 10.1177/1553350619893236] [Medline: 31920155]
116. Ingeson M, Blusi M, Nieves J. Microsoft Hololens - A mHealth Solution for Medication Adherence. Cham: Springer; 2018 Presented at: International Workshop on Artificial Intelligence in Health; 13-14 July; Stockholm, Sweden.
117. Kwon H, Park Y, Han J. Augmented reality in dentistry: a current perspective. Acta Odontol Scand 2018 Oct;76(7):497-503. [doi: 10.1080/00016357.2018.1441437] [Medline: 29465283]
118. Azuma RT. A Survey of Augmented Reality. Presence: Teleoperators & Virtual Environments 1997 Aug;6(4):355-385. [doi: 10.1162/pres.1997.6.4.355]

Abbreviations

CAVE: Cave Automatic Virtual Environment
HMD: head-mounted display
IDEA: Individual Dental Education Assistant
IDEAL: Internet of Things–based education and learning
MRI: magnetic resonance imaging
RDTES: Real-time Dental Training and Evaluation System
VLNP: Virtual Learning Network Platform
VR: virtual reality

Edited by G Eysenbach; submitted 22.08.20; peer-reviewed by J Song, A Rung, P Kanzow; comments to author 06.09.20; revised version received 26.11.20; accepted 29.01.21; published 08.04.21.

Please cite as:
Li Y, Ye H, Ye F, Liu Y, Lv L, Zhang P, Zhang X, Zhou Y
The Current Situation and Future Prospects of Simulators in Dental Education
J Med Internet Res 2021;23(4):e23635
URL: https://www.jmir.org/2021/4/e23635
doi:10.2196/23635
PMID:33830059

Viewpoint

Beyond Notes: Why It Is Time to Abandon an Outdated Documentation Paradigm

Jackson Steinkamp[1*], MD; Jacob Kantrowitz[2*], MD, PhD; Abhinav Sharma[3*], MD; Wasif Bala[4*], MD

[1]Department of Medicine, University of Pennsylvania, Philadelphia, PA, United States

[2]Department of Medicine, Carney Hospital, Boston, MA, United States

[3]Department of Family and Community Medicine, University of Toronto, Toronto, ON, Canada

[4]Department of Radiology, Emory University, Atlanta, GA, United States

[*]all authors contributed equally

Corresponding Author:
Abhinav Sharma, MD
Department of Family and Community Medicine
University of Toronto
500 University Avenue, 5th Floor
Toronto, ON, M5G 1V7
Canada
Phone: 1 7789387714
Email: abhinavarun@gmail.com

Abstract

Clinicians spend a substantial part of their workday reviewing and writing electronic medical notes. Here we describe how the current, widely accepted paradigm for electronic medical notes represents a poor organizational framework for both the individual clinician and the broader medical team. As described in this viewpoint, the medical chart—including notes, labs, and imaging results—can be reconceptualized as a dynamic, fully collaborative workspace organized by topic rather than time, writer, or data type. This revised framework enables a more accurate and complete assessment of the current state of the patient and easy historical review, saving clinicians substantial time on both data input and retrieval. Collectively, this approach has the potential to improve health care delivery effectiveness and efficiency.

(J Med Internet Res 2021;23(4):e24179) doi:10.2196/24179

KEYWORDS

electronic medical records; health informatics; information chaos; medical documentation; clinicians; medical notes; electronic medical notes; medical team

Electronic medical record (EMR) documentation is used for clinical, medicolegal, and billing purposes, as well as research, quality improvement, and population health initiatives. Each of these downstream use cases drives documentation requirements and influences day-to-day efficiency, work satisfaction, and quality of patient care. Poorly designed systems lead to the proliferation of out-of-date or incorrect information [1,2], increased time spent searching the medical chart [3], medical errors [2], and clinician burnout [3-5] while limiting the effective use of EMR data for individual and population-level applications.

Many clinicians are intimately familiar with the pathologies of "information chaos" in electronic patient charts [1]: large volumes of duplicate or *copy-pasted* information, scattered information requiring multiple navigation steps to locate, conflicting information, and outright erroneous information. Well-designed systems and workflows should incentivize data entry, storage, and retrieval behaviors that minimize these forms of information chaos. Conversely, when we see rampant information chaos, we should question and re-evaluate our default assumptions about documentation.

Currently, most medical documentation is organized in bundles or containers called "notes": documents containing many distinct but loosely related clinical observations, interpretations, and plans. Although notes lump together varied data from different sources regarding different topics, they are generally stored as indivisible units, written largely by a single author, and are rarely edited after initial creation. In their current form, notes organize information by time, by clinical thread (subject matter or team ownership; eg, primary team vs consultant), and by responsibility (the writer of a note simultaneously attests to the truth of, and takes responsibility for, all assertions within it).

These are the specific organizational principles and design choices that form the *note* paradigm. These principles—and their consequences for documentation—are easy to overlook when the note paradigm is the only paradigm most clinicians know. Notes have been the predominant organizational principle since the era of paper records, but what if *notes* are the wrong organizational paradigm and are largely responsible for the information chaos that plagues modern electronic charts?

Consider a set of notes written by a single clinician over time. Since each note is a disjointed bundle of information, recorded and accessed *separately*, there are two ways for a clinician to treat a set of sequential notes. The first is to treat each note as a complete *state of the patient* at a particular time. This approach has at least one major problem: the vast majority of a patient's medical information remains the same from time t to time t + 1. With this approach, clinicians not only habitually document the new updates but also copy and paste large portions of unchanged information from one note to the next to keep all of the information about the patient's current state in one place. This habit contributes to charts overloaded with duplicate information, makes it difficult for a later reader to easily identify what has changed from time t to t + 1, and makes it onerous to truly expunge errors from the chart because errors now require correction in multiple locations. On the other hand, some clinicians use each note to record only the *new updates* in the patient's state from t to t + 1. This practice scatters data that should be stored in a single place across multiple notes (eg, the time course of a patient's chronic medical problems), making it difficult for later readers to find relevant information and to piece together complicated medical histories. So, in the note's organizational paradigm, clinicians are in a lose-lose situation—forced to choose between large-scale information duplication (if each note is treated as a complete *state of the patient*) or information scatter (if each note is treated as a *bundle of updates*). Most clinicians choose not to adhere strictly to either of these strategies, leading to charts bloated with duplicate and yet seemingly incomplete data, medical histories that are onerous to synthesize, and hard-to-correct errors.

Notes are a poor organizational framework for the individual clinician, but they may be even worse for a collaborative medical team. Although little of a patient's medical information changes depending on the team or physician viewing it, different teams habitually redocument the same information (eg, the history of present illness) in separate notes, representing another large source of duplicated information. When information does differ between teams (eg, differing physical exam findings or a more in-depth cardiology history), it can only be identified by navigating between separate notes. Such practices not only contribute to information scatter and overload but also waste time on duplicate effort and limit optimal collaborative potential within the EMR.

The current note paradigm thus is a key source of information chaos and associated frustrations. We suggest the value of moving toward a nonnote paradigm in a wholesale effort to reimagine what the electronic medical chart should be. First, we suggest organizing information in the chart primarily by topic rather than by time, team, writer, or data type. This means rethinking how information is stored in the chart—not only the narrative text but also the structured data, including orders, medications, laboratory values, imaging, and other diagnostic results. These data should be tightly coupled to the relevant medical problems and associated free-text data, and update in place dynamically.

Next, the narrative medical chart should be reconceived as a dynamic living workspace—a set of shared and editable topics that contain only the most up-to-date information, rather than a set of separate, fixed time slices. Instead of requiring clinicians to painstakingly identify the information in the medical chart that is true at a given time, this approach ensures that the information present at a glance accurately depicts the most current state of the patient by default. However, the past patient states would still be easily accessible, by navigating forward and back through time, with a version history feature or *track changes* functionality (similar to modern word processing software). This feature would enable readers to quickly identify updates to the chart from t to t + 1 or to follow individual medical problems over time at a more granular level than allowed by notes. When documenting, clinicians would update only the changed information and attest to, but not redocument, the unchanged information. This would facilitate granular clinical updates as they occur (eg, a single medication change or a new symptom) without requiring an entirely new note or addendum. This dynamic workspace paradigm would save substantial time on both the data input and retrieval sides while mitigating incentives for information scatter and copy-paste. Problem-based charting is an appropriate but insufficient step toward this ideal.

A nonnote paradigm would also reduce information chaos at the team level. The chart could be reconceived as a fully collaborative workspace where information is, by default, shared among clinicians. Attestation would replace redocumentation as the primary mode of agreement, even across different clinical services. This is something already done in parts of the chart, for instance to confirm the allergies or family history. Disagreements could easily be documented under this system, but in the same topic-defined place, so that future readers can easily see where opinions differ. Customizable views of certain *slices* of information, depending on a user's needs, could facilitate service- or individual–specific dashboards for particular workflows. The version history system would enable granular tracking of which clinicians made which changes to a given chart, allowing for appropriate assignment of responsibility (for clinical, medicolegal, or billing purposes). Such a system could be applied to reduce duplication and scatter within a single hospital stay (consultants collaborating with primary teams) and across clinical contexts (an inpatient clinician collaborating with an outpatient clinician). Under such a system, to copy and paste text from another person's *note* would seem absurd—after all, the two clinicians share and edit the same dynamic workspace.

Deeper than specific software interfaces or billing requirements, our reliance on the note as the primary mode of information organization is a major contributor to clinicians' EMR frustrations. The note is largely a historical relic from the era of physical paper records, which has outlived its usefulness in the digital era. By questioning historical assumptions and

developing alternative documentation paradigms, we can reduce information chaos, promoting improved clinical care, efficiency, and health care worker satisfaction.

Conflicts of Interest

JS, JK, AS, and WB are cofounders of River Records, a health care technology company focused on improving clinician workflows.

References

1. Beasley JW, Wetterneck TB, Temte J, Lapin JA, Smith P, Rivera-Rodriguez AJ, et al. Information chaos in primary care: implications for physician performance and patient safety. J Am Board Fam Med 2011;24(6):745-751 [FREE Full text] [doi: 10.3122/jabfm.2011.06.100255] [Medline: 22086819]
2. Bowman S. Impact of electronic health record systems on information integrity: quality and safety implications. Perspect Health Inf Manag 2013;10:1c [FREE Full text] [Medline: 24159271]
3. Arndt BG, Beasley JW, Watkinson MD, Temte JL, Tuan W, Sinsky CA, et al. Tethered to the EHR: primary care physician workload assessment using EHR event log data and time-motion observations. Ann Fam Med 2017 Sep;15(5):419-426 [FREE Full text] [doi: 10.1370/afm.2121] [Medline: 28893811]
4. Gardner RL, Cooper E, Haskell J, Harris DA, Poplau S, Kroth PJ, et al. Physician stress and burnout: the impact of health information technology. J Am Med Inform Assoc 2019 Feb 01;26(2):106-114 [FREE Full text] [doi: 10.1093/jamia/ocy145] [Medline: 30517663]
5. Babbott S, Manwell LB, Brown R, Montague E, Williams E, Schwartz M, et al. Electronic medical records and physician stress in primary care: results from the MEMO Study. J Am Med Inform Assoc 2014 Feb;21(e1):e100-e106 [FREE Full text] [doi: 10.1136/amiajnl-2013-001875] [Medline: 24005796]

Abbreviations

EMR: electronic medical record

Edited by R Kukafka; submitted 07.09.20; peer-reviewed by J Glaser, D Hu; comments to author 16.09.20; revised version received 30.09.20; accepted 18.03.21; published 20.04.21.

Please cite as:
Steinkamp J, Kantrowitz J, Sharma A, Bala W
Beyond Notes: Why It Is Time to Abandon an Outdated Documentation Paradigm
J Med Internet Res 2021;23(4):e24179
URL: https://www.jmir.org/2021/4/e24179
doi:10.2196/24179
PMID:33877053

Viewpoint

Leveraging Virtual Reality and Augmented Reality to Combat Chronic Pain in Youth: Position Paper From the Interdisciplinary Network on Virtual and Augmented Technologies for Pain Management

Deirdre E Logan[1,2], PhD; Laura E Simons[3,4], PhD; Thomas J Caruso[3,4*], MD; Jeffrey I Gold[5,6,7,8*], PhD; Walter Greenleaf[9*], PhD; Anya Griffin[3,4*], PhD; Christopher D King[10,11,12*], PhD; Maria Menendez[4*], MD; Vanessa A Olbrecht[10,12,13*], MD, MBA; Samuel Rodriguez[3,4*], MD; Megan Silvia[14,15*], OTR, MSc; Jennifer N Stinson[16*], PhD; Ellen Wang[3,4*], MD; Sara E Williams[10,11,12*], PhD; Luke Wilson[17*], BA

[1]Department of Anesthesiology, Critical Care and Pain Medicine, Boston Children's Hospital, Boston, MA, United States
[2]Department of Psychiatry, Harvard Medical School, Boston, MA, United States
[3]Department of Anesthesiology, Perioperative, and Pain Medicine, Stanford University School of Medicine, Palo Alto, CA, United States
[4]Lucile Packard Children's Hospital Stanford, Stanford University, Stanford, CA, United States
[5]Department of Anesthesiology, Keck School of Medicine, University of Southern California, Los Angeles, CA, United States
[6]The Saban Research Institute, Children's Hospital Los Angeles, Los Angeles, CA, United States
[7]Department of Pediatrics, Keck School of Medicine, University of Southern California, Los Angeles, CA, United States
[8]Department of Psychiatry and Behavioral Sciences, Keck School of Medicine, University of Southern California, Los Angeles, CA, United States
[9]Stanford University, Stanford, CA, United States
[10]Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, OH, United States
[11]Division of Behavioral Medicine and Clinical Psychology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, United States
[12]Center for Understanding Pediatric Pain (CUPP), Cincinnati Children's Hospital Medical Center, Cincinnati, OH, United States
[13]Department of Anesthesiology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, United States
[14]Department of Physical Therapy and Occupational Therapy Services, Boston Children's Hospital, Boston, MA, United States
[15]Pediatric Pain Rehabilitation Center, Boston Children's Hospital, Boston, MA, United States
[16]Hospital for Sick Children, Toronto, ON, Canada
[17]Mighty Immersion, Inc., New York, NY, United States
[*]these authors contributed equally

Corresponding Author:
Laura E Simons, PhD
Department of Anesthesiology, Perioperative, and Pain Medicine
Stanford University School of Medicine
Suite 300
1070 Arastradero Road
Palo Alto, CA, 94304
United States
Phone: 1 650 736 0838
Email: lesimons@stanford.edu

Abstract

Background: Virtual reality (VR) and augmented reality (AR) interventions are emerging as promising tools in the treatment of pediatric chronic pain conditions. However, in this young field, there is little consensus to guide the process of engaging in the development and evaluation of targeted VR-based interventions.

Objective: The INOVATE-Pain (Interdisciplinary Network on Virtual and Augmented Technologies for Pain management) consortium aims to advance the field of VR for pediatric chronic pain rehabilitation by providing guidance for best practices in the design, evaluation, and dissemination of VR-based interventions targeting this population.

Methods: An interdisciplinary meeting of 16 academics, clinicians, industry partners, and philanthropy partners was held in January 2020.

Results: Reviewing the state of the field, the consortium identified important directions for research-driven innovation in VR and AR clinical care, highlighted key opportunities and challenges facing the field, and established a consensus on best methodological practices to adopt in future efforts to advance the research and practice of VR and AR in pediatric pain. The consortium also identified important next steps to undertake to continue to advance the work in this promising new area of digital health pain interventions.

Conclusions: To realize the promise of this realm of innovation, key ingredients for success include productive partnerships among industry, academic, and clinical stakeholders; a uniform set of outcome domains and measures for standardized evaluation; and widespread access to the latest opportunities, tools, and resources. The INOVATE-Pain collaborative hopes to promote the creation, rigorous yet efficient evaluation, and dissemination of innovative VR-based interventions to reduce pain and improve quality of life for children.

(J Med Internet Res 2021;23(4):e25916) doi:10.2196/25916

KEYWORDS

virtual reality; pediatric; chronic pain

Introduction

Background

The integration of digital health technologies into the treatment of acute and chronic pain has accelerated in the last decade. Digital health has the potential to increase patient engagement, reduce access barriers, and enhance patient-centered care, with the central goal of alleviating pain and disability in patients with chronic pain. Among these novel technologies are virtual reality (VR) and augmented reality (AR), which allow users to engage completely (VR) or partially (AR) in immersive and interactive digital experiences. VR and AR-based interventions have been found to be effective in reducing acute pain from medical procedures associated with transient pain (eg, distraction) among adults [1-3] and youth [1,4-6]. In this context, the effectiveness of VR arises from its ability to provide multisensorial engagement that can compete with pain signaling while also eliciting enjoyment and decreasing anxiety and negative mood [7]. Youth are particularly well suited to benefit from VR and AR-based pain interventions given their facility with technology and the ease with which they can engage in imaginative experiences [8].

Applying VR and AR interventions to persistent or chronic pain treatment often provides immersive experiences that go beyond redirecting attention away from a discrete painful stimulus to include physical, cognitive, and affective therapeutic targets [9]. There is growing recognition of the potential benefits of VR and AR for persistent pain, but our understanding of the mechanisms of effect and efforts to maximize the potential of these approaches for patients with persistent pain is just beginning to emerge, with those targeting youth with chronic pain being rare. To address this gap, we convened a group of pediatric pain experts to review the state of the field, identify important directions for research-driven innovation in VR and AR clinical care, and establish a consensus on best methodological practices to adopt in future efforts to advance the research and practice of VR and AR in pediatric pain. This paper summarizes the processes and outcomes of this meeting. Specifically, we outline key gaps in the field, examine models of collaboration to advance the development and evaluation of targeted VR technology, and offer recommendations for best practices for future efforts to advance the study and use of VR and AR in the treatment of pediatric chronic pain.

VR and AR interventions for the acute pain context rely primarily on VR's immersive capabilities to distract from the discomfort of the pain experience. In the context of persistent pain, intervention targets must be broad in scope (eg, address pain-related functioning) and enduring in their effects (eg, providing a sustained reduction in pain and/or improvement in functioning lasting beyond the time spent in the VR environment). In a recent systematic review of VR and AR for pain in adults (which included chronic pain cohorts), Mallari et al [2] concluded that VR appears to reduce chronic pain intensity while VR exposure is ongoing, but there is weak evidence for any lasting post-VR exposure effects on pain intensity with existing protocols. Of 10 studies on VR and AR for chronic pain included in the review, only 2 studies [10,11] expanded beyond a focus on pain intensity to measure functional outcomes. Both studies focused on chronic neck pain and demonstrated functional improvements on a standardized self-report measure of neck functioning during or immediately following the VR exposure, with 1 study [10] suggesting continued self-reported functional gains at a 3-month follow-up.

Trost et al [9] provided a timely overview of the existing research on VR applications for pain. The authors highlight the *3 pillars of VR*: presence, immersion, and interactivity, which vary in importance depending on the pain context (acute, experimental to examine mediators or moderators, or chronic). As noted by Trost et al [9], presence refers to the subjective experience of being in one environment while being physically situated in another. Immersion refers to the level of absorption a user experiences in the virtual world, and interactivity refers to the extent to which the user can manipulate the virtual environment. These authors provide a useful heuristic model that outlines technical VR configuration factors (eg, hepatic feedback), user experiential factors (eg, presence), and pain targets (eg, cognitive behavioral therapy skills) that converge to influence outcomes (eg, pain intensity) [9]. Importantly, these dimensions may vary based on the goals of the VR therapy context.

Many chronic pain studies have made the pillars of presence or embodiment and interactivity central to the target. Examples include studies of phantom limb pain [12,13] or spinal cord injury [14,15], with VR potentially amplifying the neuromodulatory effects of movement therapies. Less emphasized in the context of chronic pain is the pillar of immersion, given that distraction is less central in chronic pain management. For example, Villiger et al [16] used 16 to 20 sessions of VR to deliver intensive individual muscle training in patients undergoing neurorehabilitation for spinal cord injuries. The results showed both subjective and objective physical function improvements, with gains remaining stable at 12 and 16 weeks postintervention. Trost et al [17,18] pioneered the use of VR to deliver graded-exposure therapy targeting pain-related fear and disability in adults with chronic pain, recognizing the power of VR to facilitate pain-related movement in the presence of fear and behavioral avoidance. In addition to the recent work in this area specific to chronic pain, we can draw upon more robust literature on the use of VR for exposure-based treatments of other conditions. Several systematic reviews and meta-analyses of virtual reality–based exposure therapy (VRET) for anxiety disorders have shown that VRET is equal or superior to the gold standard of in vivo exposure for anxiety reduction [19-23].

Overall, there is support for using VR and AR as a tool to reduce perceptions of pain intensity in the context of procedural pain, including evidence of effects for children and adolescents. However, a recent review underscores the need for large, well-designed trials to fully evaluate the effects of VR on acute pain in children because of the variability and weaknesses in the study methodology to date [6]. There is growing evidence of the utility of VR and AR for reducing the intensity of chronic pain in the short term and emerging support for using VR in physical rehabilitation and to reduce the fear of pain through exposure-based paradigms in adult populations. The intersection of these areas—that is, the use of VR and AR technology to achieve multiple benefits (reduction of pain, disability, and fear or behavioral avoidance) over a longer period in the context of pediatric chronic pain rehabilitative treatment—is only emerging. To date, published work specifically focused on VR for pediatric chronic pain mainly includes small pilot studies, but these have demonstrated support for using VR to augment established treatments for pediatric chronic pain, including mirror visual feedback therapy for complex regional pain syndrome [24] and biofeedback for pediatric headache [25], with outcomes suggesting increased tolerance of rehabilitative therapies, reduced pain, and improved function and quality of life. Given the increasing accessibility of VR and AR interventions [26], their particular fit for pediatric populations, and their ability to complement and enhance standard approaches for the treatment of complex pediatric pain conditions [27], this is an important and timely area of innovation in our field. As new applications for VR and AR emerge, we need guidelines and best practices to inform the design and evaluation of new potential interventions.

Objectives

To review the current state of the field of VR for the treatment of chronic pediatric pain and to elucidate important directions for future research and innovation efforts, we convened a working group of experts across relevant domains to share lessons learned through their early work in this area and to generate consensus-backed recommendations for advancing the field. Herein, we provide guidance from the newly formed INOVATE-Pain collaborative (Interdisciplinary Network on Virtual and Augmented Technologies for Pain) for key domains: (1) describing models of academic-clinical-industry partnerships in the development and dissemination of immersive pain intervention approaches; (2) outlining current and future opportunities and challenges for evidence-based innovation, with an emphasis on collaboration across diverse sites, settings, and domains of expertise; and (3) identifying best practices in research on the use of VR and AR in children and adolescents with chronic pain, including recommendations for measuring meaningful outcomes of VR and AR-based interventions.

Methods

Participants

The INOVATE-Pain collaborative includes current contributors to science in the field of immersive interventions for pediatric pain rehabilitation, including clinicians, clinical researchers, neuroscientists, and VR software engineers. Potential participants were identified based on their involvement in ongoing work in the realm of VR for pediatric chronic pain treatment. Those first approached by the meeting organizer and chair (DL) were asked to suggest other participants who could represent different perspectives and experiences. Participants included all authors from the following institutions and organizations: Stanford Children's Health network, the Stanford Childhood Anxiety Reduction through Innovation and Technology (CHARIOT) program, Boston Children's Hospital (BCH), Cincinnati Children's Hospital Medical Center (CCHMC), The Hospital for Sick Children (SickKids) in Toronto, University of Southern California or Children's Hospital of Los Angeles (CHLA), Stanford Virtual Human Interaction Lab, and Mighty Immersion, Inc (a company focused on VR device management tools). This was the first meeting of this group. Follow-up meetings were planned but have not been executed because of the COVID-19 pandemic. The group continues to hold monthly Zoom meetings to continue the work that began at the initial meeting. This meeting was supported by The Mayday Fund. A representative from The Mayday Fund was invited to attend the meeting to provide perspective on the funding landscape and report to The Mayday Fund board on the proposed strategic directions of the INOVATE-Pain collaborative.

Setting

The meeting took place over the course of 2 days (January 22-23, 2020) in Palo Alto, California.

Procedures

The impetus for the INOVATE-Pain consensus meeting arose through The Mayday Fund's efforts to connect individuals engaged in funded research on VR for chronic pain with those seeking funding for proposed projects in this area in the hope of promoting greater collaboration and standardization in the

field. The meeting convened with a structured agenda (Multimedia Appendix 1) and planned deliverables for a consortium focused on VR and AR innovations in pediatric pain rehabilitation. Briefly, the meeting began with an overview of the state of the field and moved to reports from each participating academic medical center on their clinical and research-based uses of VR in pediatric chronic pain, followed by a discussion of new technological developments from our industry partner (LW) and his clinical collaborators. Group members representing the funding world and laboratories exploring the use of VR technology in adult populations shared relevant insights. This was followed by a discussion of best practices, with particular emphasis on what outcomes are most relevant to assess when evaluating pediatric chronic pain VR and AR interventions and what tools would be best for such assessments. The second day focused on arriving at a consensus on important new directions in the field and models of collaboration that bring clinicians, researchers, and industry partners together in ways that are mutually beneficial, nonexploitative, and hold patient care as a shared primary goal. Time was devoted to developing INOVATE-Pain's mission and vision statement, along with strategic goals focused on what our collaborative can offer the field. Finally, we outlined plans for moving forward with actionable, collaborative projects and outputs. Detailed notes were recorded throughout the meetings by the nonparticipants.

The meeting relied on consensus decision making in an open discussion format. To begin, we drew on case examples of current and potential projects overseen by group members to develop a shared understanding of where the field stands and to identify gaps. During discussions, the chair and cochair (LS) kept time limits and reflected themes and major points to the group to ensure that members were in agreement on how these were conceptualized. We then engaged in structured brainstorming sessions derived from ideation methods in design thinking, an approach now being applied to health care innovation and education [28]. We undertook this process for several of our key agenda topics, including identifying important outcome domains to assess in VR studies and developing our collaborative's mission, vision, and strategic goals. Given the group's relatively small size, discussions and brainstorming exercises were held with the full group. We adhered to design-thinking brainstorm rules (ie, time limit, stay on topic, defer judgment, encourage wild ideas, aim for quantity, build on each other's ideas, be visual, and one conversation at a time) utilizing a whiteboard and unlimited Post-its. Following the ideation phase, we grouped similar ideas and labeled these larger constructs, which ultimately became the major themes we discuss in our results (see Multimedia Appendix 2 for this exercise's visual depiction).

Results

Describing Models of Collaborative Partnerships to Advance VR Development, Implementation, and Evaluation

As a first step in identifying successful collaboration models, meeting attendees illustrated ways in which such partnerships have succeeded in their own settings to date.

The Stanford CHARIOT Program

The CHARIOT program has led the field in clinical applications and innovations in VR for pediatric procedural anxiety and pain. Having used VR in more than 5000 patients since 2017, the scope of impact and dissemination of VR software tools across 25 institutions worldwide reflects a motivation to rapidly equip hospitals with VR tools for patient care, particularly in the realms of procedural discomfort and anxiety. In addition to the direct distribution of content to these hospitals, the software developed by Stanford CHARIOT has been included on hospital VR platforms used by more than 200 hospitals and clinics in the United States.

Stanford Pediatric Pain Rehabilitation

Since 2018, the Stanford Children's Health Pediatric Rehabilitation Program (PReP) for youth with chronic pain has used VR interventions during physical therapy, occupational therapy, and individual psychology sessions. The program obtained a dedicated VR therapy room for use during the PReP treatment sessions. In collaboration with the Stanford CHARIOT program and through support from The Mayday Fund, a VR platform called Fruity Feet was developed with the PReP interdisciplinary team. This platform was created through an iterative process from PReP provider and patient feedback for improvements and modifications. The overall goal was to improve function and reduce the fear of pain. The initial implementation of Fruity Feet VR tested the acceptability and feasibility of facilitating increased upper and lower extremity engagement [29]. This platform was simultaneously tested with an inpatient population during physical therapy sessions with large effects on movement observed in VR compared with the standard of care physical therapy (Caruso et al, unpublished data, August 2019).

In response to the COVID-19 pandemic, this work involves pilot testing of at-home VR for pain rehabilitation. With PReP operating in a telehealth format for all treatment interventions, youth in PReP were provided with VR equipment for home use during telehealth sessions with providers and independently for their daily home exercise plan. Preliminary data for the at-home VR intervention are forthcoming, but the initial interview feedback is positive.

Cincinnati Children's Hospital Medical Center

Since 2018, the use of VR in patients with pain at CCHMC has followed a -pronged approach: a clinical research arm for managing acute postoperative pain and a clinical application arm for patients with chronic pain undergoing inpatient rehabilitation. Our team has conducted 2 pilot studies in children with moderate to severe pain after surgery, one of which used

distraction-based virtual reality (VR-D) and guided-relaxation virtual reality (VR-GR) using the Mindful Aurora app [30-32]. This pilot study demonstrated that a single, postoperative session of both VR-D and VR-GR was associated with small changes in pain and anxiety lasting up to 30 minutes. This pilot study supports the implementation of VR therapy in managing acute postoperative pain in children and the need for future studies. The CCHMC team worked with Stanford CHARIOT to use VR programs appropriate for pediatric patients. This partnership has afforded the clinical team access to VR content that is highly engaging and motivating for children and adolescents.

Going forward, the VR team continues to explore engaging and innovative ways of using VR for pain management. By adding VR to postoperative pain management, we hope to augment our primary strategy of multimodal analgesia by increasing patient motivation and engagement in alternative therapies (eg, distraction and relaxation). Plans are in place to continue the implementation of VR into chronic pain rehabilitation with an emphasis on increasing social interaction in group settings through cooperative VR activities and improving movement and body awareness by incorporating VR into physical and occupational therapy.

The Hospital for Sick Children

Clinicians and researchers have been implementing VR at SickKids over the past 8 years in diagnostic imaging and emergency departments for intravenous insertions [33], perioperative services to prepare children for anesthesia during the perioperative period, and for subcutaneous port-a-cath access in oncology (Aqua KindVR) [34,35]. In a single-site pilot randomized controlled trial comparing VR (peaceful underwater gamified environment) with iPad (Apple Inc; underwater video with headphones) for oncology procedures, nurses, parents, and children reported the interventions in both groups to be acceptable, with the VR participants reporting significantly higher immersion, an underassessed outcome metric in studies of VR for pediatric pain. In addition, there were trends toward reports of less pain and distress during procedures in the VR group compared with the iPad group [35].

More recently, the SickKids Chronic Pain Clinic rehabilitation team used the Fruity Feet program developed through collaboration with Stanford CHARIOT in a quality improvement (QI) project to implement and evaluate the hospital's rehabilitation department. The QI project involved interviews with 10 adolescents and their physical therapists (PT) and found that participants reported a high acceptance of the program, high immersion and high satisfaction, no adverse events, and lower pain scores after using VR. PTs also reported that VR was easy to use and feasible to implement in the pediatric rehabilitation setting. Given the success and clinical need, SickKids has successfully advocated for a dedicated child life specialist to help with the clinical implementation of VR coupled with a programmatic approach to rolling out new interventions using evidence-based methods.

Boston Children's Hospital

With support from The Mayday Fund, the BCH has played a central role in developing the INOVATE-Pain network. This partnership has afforded BCH with opportunities to gain VR expertise from initiatives undertaken by the leaders of the pediatric pain VR field. In addition, this collaboration has provided BCH with opportunities to creatively explore and expand the utilization of VR in the treatment of pediatric chronic pain and to plan research initiatives to evaluate these novel applications.

Since 2010, the BCH's intensive pain rehabilitation day program has used a basic interactive gaming platform, Xbox Kinect (Microsoft Corporation), to engage youth with chronic pain in their daily treatment. Given this technology's success and popularity, BCH is actively seeking the integration of newly released VR into their pain rehabilitation program. Initial efforts in this process focus on replicating the success established at Stanford, SickKids, and CCHMC in using VR-based physical therapy, occupational therapy, and psychological interventions in the context of our intensive pain rehabilitation day hospital program. In addition, the BCH team is leading the effort in the development of a novel, immersive VR-based intervention targeting successful school reentry in the context of intensive pain rehabilitation. These innovative efforts have successfully garnered the philanthropic support needed to bring them to fruition.

Children's Hospital of Los Angeles

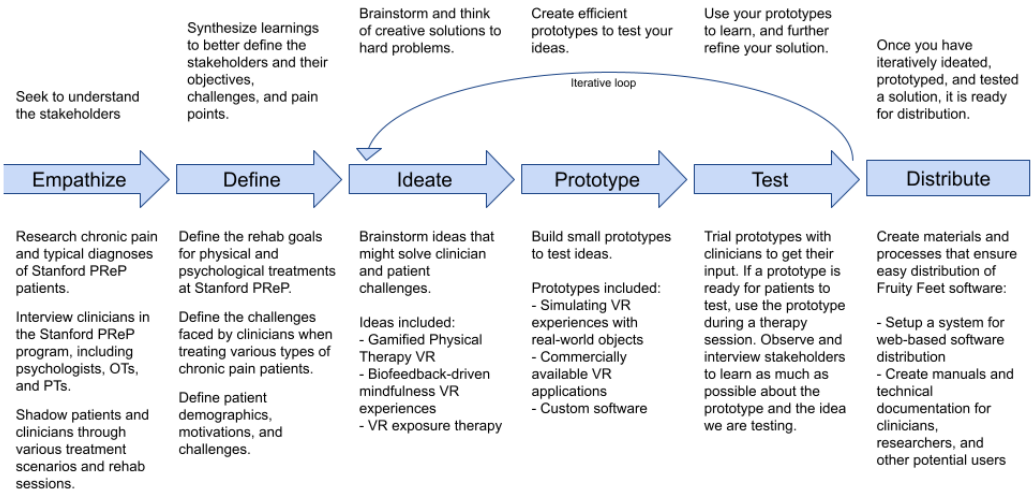
Clinician-scientists at CHLA have been exploring the use of VR for acute pain management for more than 20 years, pioneering its use with children for routine painful procedures [8,36-38]. Since its inception, the BioBehavioral Pain Lab at CHLA has partnered with academic institutions, technology companies, software developers, and other VR collaborators to maximally leverage content expertise in health care, technology advancements, and input from clinician-scientists to design and implement rigorous VR clinical trials to scientifically investigate the feasibility, usability, and efficacy of VR for managing pain with routine medical procedures. Early and recent efforts with technological advancements have fundamentally changed clinical service lines at CHLA in phlebotomy, where children coming for outpatient phlebotomy can now request VR as a standard of clinical care for blood draws. Recently, VR clinical research has focused on VR home-based systems. As virtual care has been catapulted by telehealth, home-based VR has become critical for ongoing VR investigations.

Partnerships in Action: The Fruity Feet Experience

Leveraging our collective experiences thus far, a key deliverable of the meeting was to build from these experiences to define optimal collaborative models to bring clinicians, researchers, industry partners, and funders together in ways that are mutually beneficial, nonexploitative, and hold patient care as a shared primary goal. The industry partner, Luke Wilson of Mighty Immersion Inc [39], provided an overview of the design-thinking framework used to build the VR program developed specifically for pediatric pain rehabilitation treatment centers, *Fruity Feet*, which is currently being implemented at Stanford, CCHMC, and SickKids (Figure 1). In addition to targeted software experiences, we identified additional tools crucial to the successful implementation of VR interventions in patient care settings. First, *device management tools* can facilitate updating

all of an institution's VR tools, track their location in busy clinical settings, and remotely manage devices as needed. Second, *companion applications* enable controlling the VR experience from a second device (eg, provider's phone), thus tweaking the therapy to fit the patient and therapeutic target. Third, *analytics capture tools* collect movement data, create playlists for VR protocols, and display real-time patient movement. Although seamless integration is the goal, having a technically knowledgeable team member is critical to reducing the clinician time burden and ease of implementation. Industry partners can provide valuable training to develop clinician champions who can oversee VR integration in clinical settings.

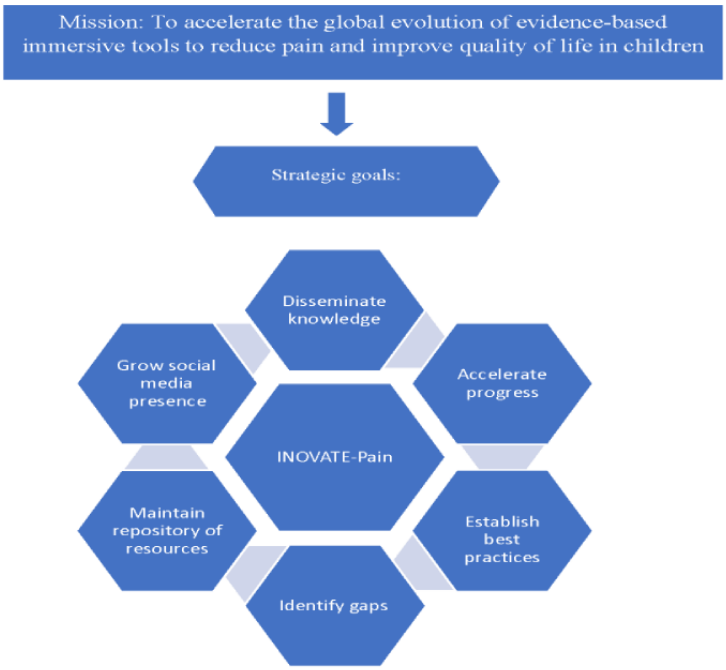
Figure 1. Fruity Feet/Design Thinking Development. PTs: physical therapists; OTs: occupational therapists; VR: virtual reality.



Defining a Mission and Vision for INOVATE-Pain

Building off this discussion of successful industry-academic partnership, the mission statement, vision, and strategic goals for the INOVATE-Pain collaborative were defined through design-thinking exercises and a consensus process. In formulating our mission, vision, and goals, we sought to differentiate our collaborative from existing programs by asking what we could add to the current efforts in the field. For example, how did we differ from the Stanford CHARIOT program, which uses VR in a broader range of populations and emphasizes the global dissemination of technology? Did we want INOVATE-Pain to be a marketplace for ideas, a repository of resources, and/or an established multisite consortium for new research projects? Through discussion, we identified a strong commitment to fostering and facilitating research and evaluation, given the importance of building an evidence base to support the use of VR interventions. As previously noted, a design-thinking process enabled us to identify the themes that were most widely endorsed by the group as important to our mission and vision, and further discussion of these themes led to full consensus. The mission and vision of accelerating the global evolution of evidence-based immersive tools to reduce pain, reduce disability, and improve children's quality of life are supported by our strategic goals (Figure 2).

Figure 2. Mission and strategic goals of the INOVATE-Pain (Interdisciplinary Network on Virtual and Augmented Technologies for Pain management) consortium.



Outlining Current and Future Challenges and Opportunities in the Field

A second key goal of the meeting was to identify both challenges and opportunities to move the field forward, along with ways that a collaborative network such as INOVATE-Pain could serve to capitalize on the opportunities and surmount the challenges. The 4 major factors were identified as follows: (1) industry versus academic approaches to the evaluation process, (2) discrepant resources and endpoints across academic, clinical, and industry collaborators, (3) proliferation of technology products and companies, and (4) adoption of VR in research and clinical applications.

Industry Versus Academic Approaches to the Evaluation Process

The importance of interdisciplinary, multi-institutional, and cross-contextual (industry, academic, and clinical) collaborations toward developing VR software and clinical investigations cannot be understated. However, there are challenges inherent in aligning approaches to the development and evaluation of products in each of these contexts. Navigating logistical, regulatory, and intellectual considerations is a challenge when collaborating with similar institutions (eg, academia), and they are amplified when hospitals and universities attempt to work with industry and commercial partners. Common hurdles that arise include the differing pace of development, the need to navigate multiple regulatory systems, and distinctions in the desired scope or target of the intervention. Start-ups and industries tend to move very quickly with innovation efforts and face fewer regulatory barriers than academic medicine. Large academic and clinical institutions have many layers of regulation, such as intellectual property (IP) considerations when developing products and licensing and billing issues when delivering these interventions in the clinical setting. This raises particular challenges in the realm of digital health and emerging technologies. After a rigorous randomized clinical trial is funded, institutional review is granted, and data are collected and analyzed (a course of action often measured in years), the technology under investigation may be rendered obsolete. However, opportunities exist within these challenges when these 2 approaches can work in a complementary fashion to pursue more nimble approaches to evaluation and dissemination. For example, clinical settings can provide real-world testing grounds and patient populations that are difficult for the industry to access alone. Meanwhile, performing some of the product development in the industry setting may allow the process to move more quickly than projects housed entirely within the academic environment.

In addition to working out issues of project pace and addressing regulatory requirements, partners may differ in their views of project scope. There are advantages and disadvantages to developing interventions tailored to a specific clinical population versus creating widely applicable tools, but potentially less targeted to a particular population (eg, disease condition and age group). This may be particularly true of VR, where there is a huge range of potential software applications and a variety of potential mechanisms of effectiveness that can be harnessed. There is tremendous opportunity to be leveraged when industry and academic partners come together early in the design process so that these advantages and disadvantages can be considered together and interventions can be designed to fit real-world clinical needs, rather than trying to retrofit existing products to use with specific patient populations. Fruity Feet is an excellent example of a collaboration that capitalized on the collaborative process to develop a project that held broad enough appeal to be useful in multiple contexts while also meeting specific needs of children with chronic extremity pain undergoing rehabilitative pain treatment.

Academic, Clinical, and Industry Collaborators May Have Different Resources and Endpoints

Related to, but in some ways separate from, the challenges around evaluation design, resources are a major point of negotiation in developing productive partnerships for VR innovation. Resource sharing between collaborators can offer huge benefits in evaluating new interventions, but resource sharing can also multiply hurdles. Given the obligations one assumes when undertaking this type of endeavor and the time and effort required to succeed, it is critical for academic partners to enter carefully into these partnerships to avoid becoming an open testing ground for too many industry products.

For academic and clinical researchers, opportunities lie in the fact that industry partners can bring specific financial benefits to fuel the research endeavor, with tech-savvy personnel and computer science skills exceeding those typically found within academia. In turn, academic partners bring expertise and access to some funding opportunities (eg, National Institutes of Health [NIH] Small Business Innovation Research and similar mechanisms, seed funding to pilot new ideas), along with the ability to gather data for extra- or intramural grant applications, culminating in peer-review publications contributing to the evidence base in VR. Other benefits to academic partners include the possibility of IP or equity in a start-up that could also contribute to academic success or contribute to building a financially secure lab for ongoing research and future collaborations. The industry tends to focus on validating their product through institutional acceptance and the association with large prestigious medical and/or university settings. Conversely, some industry partners want to engage with content expertise to iterate and improve their products to increase efficacy and impact. Ideally, similar-minded academic-industry collaborations focus on creating an ecologically sound product that delivers on the idea that technology can assist in solving complicated health care problems.

Proliferation of Technology Products and Companies

The major challenge in the vast immersive technology marketplace is finding effective products that arise from sound research-based development. Over time, the cost of VR hardware has significantly reduced. Once an experimental scientific *ivory tower* effort with limited clinical application, it has evolved into affordable, off-the-shelf products that operate with minimum components to increase ease of use. The accessibility of VR products provides new opportunities for integrating these interventions into clinical care. VR hardware has become highly sophisticated and affordable such that most clinician-scientists can now afford to add it to their toolkit or even work with equipment owned by patients. In addition, as the industry expanded, the inventory has grown from very few available virtual experiences (VEs) to hundreds. Even so, very few VEs have been scientifically well established as gold standard interventions for clinical problems.

Adoption of VR in Research and Clinical Applications

Although there are typical barriers to the adoption of new technologies in pediatric health care, VR research and adoption are particularly challenging. First, given the lack of VR safety research in pediatrics, most hardware manufacturers do not recommend VR for children before adolescence. This can further delay institutional review board approval and require additional safety monitoring that is not required with widely accepted hardware, such as tablets. Second, given the typical age of users, the headsets often require custom modifications to appropriately fit children's head circumference to facilitate such safety and efficacy research. Third, because VR is applied directly to the patient, near the nose and mouth, infection control practices require close attention. Many consumers facing VR products include cloth straps and face pads, which present challenges to sanitation among patients. Finally, although reduced costs have made VR headsets more affordable, they are still a relatively costly adjunct given the lack of reimbursement offered by insurance for their utilization. In the future, appropriate reimbursements will mitigate the cost of implementation.

Define and Refine Best Practices

The final major goal of our consensus meeting was to begin to develop a set of best practices for conducting methodologically sound research in VR in the context of pediatric chronic pain treatment. The best practices for VR and chronic pain management have not yet been established. Determining best practices is complex, as this academic-industry collaboration has many facets that include but are not limited to hardware products; software products; building an evidence base; and incorporating heterogeneous disease processes; and age, gender, linguistic, and cultural considerations. This section addresses 4 primary areas, recognizing that each area requires specific customizability for the targeted project: (1) assembling the right team, (2) developing or applying the best technology for the target project or population, (3) designing and executing a sound research design and methodology, and (4) disseminating the collaborative work process and study findings.

Assembling the Right Team

This is not an easy task, as industry start-ups are often quite eager to partner with academic VR leaders and to gain access to clinical populations to pilot and test their VR experiences. It is critical as a clinician-scientist to thoroughly vet potential partners. In these initial meetings, it is also critical to *right-size* the project and the timeline, that is, determine goals and benchmarks that are agreeable to all partners and appropriate to the desired outcomes. Aligning project scope, timeline, and financial concerns; assembling a team that includes all necessary domains of expertise; and clearly defining each team member's role are vital preliminary steps to determining the *match* between academia and industry (Textbox 1).

Textbox 1. Guidelines for establishing positive academic-industry partnerships in virtual reality innovation.

Partner assessment

- What is the reputation and track record of the industry partner?
- Are they a for-profit or nonprofit entity?
- What is their understanding of the research process?
- What are their expectations for what the academic or clinical partner(s) will provide?

Project validation

- What is the project scope?
- What is the expected timeline for deliverables?
- What are the deliverables or intended outcomes?

Understanding funding structure

- Who underwrites what aspects of the project, both in terms of money and effort?
- Is work done for compensation or in-kind arrangement?
- Discuss intellectual property arrangements—include academic or hospital legal representatives

Developing or Applying the Best Technology for the Targeted Project or Population

It is essential to carefully define the target population in the early phases of designing a VR and AR-based intervention. There are advantages and disadvantages to developing highly tailored products. On the one hand, clinical utility is paramount, and having an intervention that clearly targets a specific clinical problem increases the effectiveness of the intervention. On the other hand, the time and effort needed to create and evaluate such interventions steer away from creating products that appeal to a broader population in terms of clinical condition, age, and gender. A product that is not sufficiently specific in its intended application may be more challenging to validate through research trials. Ideally, the goal should be to create a suite of tools with options for customizing the experience to achieve a balance between specificity and broad application to a range of patients who can benefit from the intervention.

Designing and Executing the Right Research Design and Methodology

After establishing a partnership and aligning the direction and goals of the collaboration, a well-designed study to evaluate the project is vital. As with any clinical research, the goals, available sample, resources, and timeline should drive the study design. Studies to evaluate VR-based interventions can take a variety of scientifically rigorous forms before or in lieu of traditional clinical trials. Depending on the scope and goals, feasibility or usability studies, QI studies to evaluate clinical implementation, pilot studies to inform more definitive trials, and adaptive trial designs to reduce time to results may be appropriate. Ultimately, the use of rigorous design, methods, and standardized patient-reported outcomes will be critical to both the academic and the start-up and industry. The use of flexible and agnostic technology is critical for testing new software programs and moving toward disseminating new information and evidence-based results. It is necessary to set realistic shared visions, missions and objectives and to develop task-oriented timelines with deliverables. There is a need to move away from the traditional, cumbersome randomized controlled trial approach to evaluating VR interventions and toward more adaptive and efficient evaluation methods [40].

A critical element of research design is the thoughtful selection of measures to assess processes and outcomes. To date, there is no clear consensus on how to evaluate a VR experience appropriately or how to tailor the evaluation to the specific context of the intervention. As described earlier, immersion may be central in the context of acute pain distraction, whereas interactivity is paramount for pain rehabilitation and exposure-based VR protocols. Unfortunately, even in the VR field more broadly, there are no gold standard measures of embodiment, presence, and immersion, and more research is needed (see Multimedia Appendix 3 for an example measure of some of these constructs used in pediatric research). In clinical settings, it may also be important to assess provider satisfaction and feasibility in terms of how the intervention fits into the overall workload and workflow. Failure to understand and address these aspects can lead to lack of buy-in from clinical partners (see Multimedia Appendix 4 for an example measure of these constructs).

Although the factors that influence outcomes may vary, the outcomes of interest should remain uniform and consistent across studies and clinical contexts. On the basis of the discussion and consensus at our meeting, we present recommended domains of measurement to evaluate VR interventions for pediatric chronic pain (Textbox 2). Some of these domains have been deemed critical for pediatric chronic pain within the NIH common data elements (CDEs [41]). These CDEs include pain intensity, pain interference, functioning, pain catastrophizing, and treatment satisfaction [41]. Additional general domains considered particularly important for VR include affect and fear of pain or movement. Moreover, we highlight the VR-specific metrics to consider. These include physical movement, energy expenditure, physiology, and immersion. In the context of pain rehabilitation, increasing

movement is a highly relevant outcome that can be measured in real time [29]. An added consideration is the frequency of assessment, as it is likely that data collection may range from continual, session-contingent, daily, or milestone-based (eg, start of treatment and discharge). Ideally, a thorough measurement of an intervention will consist of a combination of these measurement timelines. For example, a study could include objective, continuously collected measures of physical activity and movement metrics cataloged during VR sessions, brief questionnaires deployed daily to capture gradual changes and enable single case experimental design analyses [42], and a longer battery of questionnaires completed at specified time points (eg, baseline and 3-month follow-up) to assess changes over time from repeated VR exposure.

Textbox 2. Recommended domains of measurement to evaluate virtual reality interventions for chronic pain.

Pediatric chronic pain

- Pain intensity (common data element for pediatric chronic pain from the National Institutes of Health recommendations [41]) and unpleasantness
  - Visual Analogue Scale [43]
  - Numerical Rating Scale [44]
- Pain interference (common data element for pediatric chronic pain from the National Institutes of Health recommendations [41])
  - Brief Pain Inventory [45]
  - PROMIS-Pain Interference [46]
- Functioning
  - Functional Disability Inventory [47]
  - Lower Extremity Function Scale [48]
  - Upper Extremity Functional Index [49]
  - Canadian Occupational Performance Measure [50]
- Fear of pain and movement
  - Fear of Pain Questionnaire-Short Form [51]
  - Tampa Scale of Kinesiophobia [52]
- Pain catastrophizing (common data element for pediatric chronic pain from the National Institutes of Health recommendations [41])
  - Pain Catastrophizing Scale [53]
- Affect
  - Positive and Negative Affect Schedule 10 item [54]
  - Childhood Anxiety Sensitivity Index [55]
- Satisfaction with treatment (common data element for pediatric chronic pain from the National Institutes of Health recommendations [41])
  - Patient Global Impression of Change [56]

Virtual reality specific

- Physical movement
  - Motion capture [57]
  - Actigraphy [58]
  - Wearables
- Energy expenditure
  - YUR Fit app
  - Wearables
- Physiology
  - Heart rate
  - Galvanic skin response
  - Respiration
  - Functional magnetic resonance imaging
- Immersion or presence
  - Child presence measure (Multimedia Appendix 2)

Dissemination of the Academic-Industry Collaborative Work Group and Study Findings

Not all collaborative efforts were the same. From the onset, it is critical to set mutually agreed-upon deliverables. As previously noted, academic goals are often quite different from industry objectives. Some of these discrepancies need to be addressed in the first phase of the academic-industry collaborative, establishing clear goals or deliverables. The framework (*for-profit* or *nonprofit*) can greatly influence the end goal. Ultimately, both the collaborative process and the findings derived from the academic-industry collaborative need to be published in peer-reviewed journals and disseminated widely to promote the innovation and proliferation of evidence-based technology in health care. To that end, healthy, transparent, well-communicated projects can lead to fruitful and rewarding collaborations for the academic and medical partners, industry, patient and their family, and ultimately, society. INOVATE-Pain is working to create a repository for protocols, products, resources, and recommendations to guide study design and clinical implementation that can be openly accessed in an effort to advance dissemination of the work in this field and increase opportunities for collaboration.

Discussion

Principal Findings and Next Steps

In an area as rapidly evolving and complex as digital health, there is a need for multisite efforts and cross-disciplinary collaboration to keep pace with emerging technology and develop sophisticated studies that build a sound and useful evidence base. Our consortium brought together expertise in software development, clinical applications, experimental work in VR, child psychology, physical or occupational therapy, experience in navigating IP issues and bringing industry partners into the hospital setting, and funders who provided insight into how projects can be competitive for financial support. This represents a diverse group that does not meet often to think of the challenges and opportunities in the field. Our guidance and recommendations are aimed primarily at academic and clinical partners, and we hope that further work by our group can also provide more guidance targeting best practices on the industry side of these partnerships. Furthermore, given the nascent nature of this area of innovation, our current focus is primarily on research and evaluation of new interventions. However, the ultimate goal is for VR interventions to become an evidence-based, widely adopted routine component of clinical care in pediatric chronic pain treatment.

Work to date highlights tremendous opportunities in this area of digital health innovation, with immense promise for improving the treatment of pediatric chronic pain. However, this remains a challenging field to navigate given the number of outstanding issues regarding how to identify and form productive collaborations across academic, clinical, and industry partners; how to design and obtain resources to support solid research protocols to evaluate potential interventions; and how to bring interventions from conceptualization into clinical use in large, complex institutional settings where they can be accessible to patients with the greatest needs.

We identified several important next steps to advance the field toward our mission of accelerating the global evolution of evidence-based immersive tools to reduce pain and improve the quality of life in children with pain. These include the following:

1. *Developing and maintaining a repository of resources.* We are working to collect and curate protocols, publications, available software products, information on funding mechanisms, and other resources to serve as a clearinghouse for researchers and academically oriented VR experience designers. Vetting information and opportunities and housing this information in a centralized, accessible, web-based location may help to lower the barrier to engaging in this field and better standardize the approach to evaluating new interventions. We plan to work synergistically with other like-minded groups in the field, such as the Invincikids nonprofit consortium [59].
2. *Continue our efforts to establish a gold standard set of outcomes to be measured in pediatric pain VR research and specific recommended measurement tools.* Well-validated measures exist in some of the recommended outcome assessment domains we describe but are notably lacking in domains including presence, immersion, interactivity, impact on clinical workload or workflow, and the subjective experience of both the clinician and patient users. Imminent steps toward this goal include undertaking a Delphi process to determine a minimal core data set for pediatric pain VR research and, where gaps are identified, working to develop tools to fill these gaps.
3. *Offer opportunities for education and connections in the field.* Through training opportunities and symposia, we hope our group can increase exposure to the work that has been done to date, disseminate current best practices, and facilitate connections among additional potential collaborators and industry partners who can bring new energy and ideas to advance the INOVATE-Pain mission.
4. *Work together to advance current projects and launch new interventions in a rigorous, appropriately resourced environment.* For example, we have begun planning a trial to develop and evaluate a VR intervention that exposes patients in pediatric intensive pain rehabilitation treatment to the challenges they face in returning to school settings. This intervention would incorporate physical therapy, occupational therapy, and psychological aspects and would be tailored to the specific fears and barriers each patient identifies, making returning to school such a challenging goal. We look forward to designing this project in the context of the INOVATE-Pain collaborative, where we can seamlessly access the full range of expertise needed to develop this type of patient experience and access to a large patient population by scaling the project up to multisite data collection. This is an example of the goals we can attain through the partnerships we have sought to create in this collaborative. Through specific projects, we also hope to evolve the approach to studying VR interventions in the pediatric pain environment, with the goal of balancing the rigor of traditional clinical trial design with the efficiency and creativity needed to deliver promising VR products and experiences to pediatric chronic pain patients in a

timeframe that keeps pace with digital health technology innovation.

Conclusions

In summary, VR is an exciting and promising digital health tool whose applications for reducing pediatric chronic pain are just emerging. To realize the promise of this realm of innovation, key ingredients for success include productive partnerships among industry, academic, and clinical stakeholders; a uniform set of outcome domains and measures for standardized evaluation; and easy, widespread access to the latest opportunities, tools, and resources. By exploring the current opportunities, challenges, best practices, and important next steps in VR for pediatric chronic pain, the INOVATE-Pain collaborative hopes to promote the creation, rigorous yet efficient evaluation, and dissemination of innovative VR-based interventions to reduce pain and improve quality of life for children.

Acknowledgments

The authors acknowledge funding from The Mayday Fund supporting the January 2020 planning meeting and some of the ongoing work of the INOVATE-Pain collaborative. The authors would like to thank Gillian Rush, Deborah Shear, and Marissa Heirich for taking detailed minutes during the summit meeting.

Conflicts of Interest

TC, SR, and EW are on the board of directors of Invincikids, a nonprofit company that seeks to distribute immersive technology to pediatric healthcare centers. They receive zero financial compensation for this role.

Multimedia Appendix 1
Meeting agenda.
[DOCX File , 16 KB - jmir_v23i4e25916_app1.docx]

Multimedia Appendix 2
Visual depiction of ideate brainstorming and consensus building exercise.
[PNG File , 1124 KB - jmir_v23i4e25916_app2.png]

Multimedia Appendix 3
Sample measure of child presence in the virtual reality experience.
[PNG File , 126 KB - jmir_v23i4e25916_app3.png]

Multimedia Appendix 4
Sample measure of virtual reality acceptability and impact on workflow in the clinical setting.
[DOC File , 56 KB - jmir_v23i4e25916_app4.doc]

References

1. Kenney MP, Milling LS. The effectiveness of virtual reality distraction for reducing pain: a meta-analysis. Psychol Consc: Theory, Res, Pract 2016;3(3):199-210. [doi: 10.1037/cns0000084]
2. Mallari B, Spaeth EK, Goh H, Boyd BS. Virtual reality as an analgesic for acute and chronic pain in adults: a systematic review and meta-analysis. J Pain Res 2019 Jul;Volume 12:2053-2085. [doi: 10.2147/jpr.s200498]
3. Malloy KM, Milling LS. The effectiveness of virtual reality distraction for pain reduction: a systematic review. Clin Psychol Rev 2010 Dec;30(8):1011-1018. [doi: 10.1016/j.cpr.2010.07.001] [Medline: 20691523]
4. Gates M, Hartling L, Shulhan-Kilroy J, MacGregor T, Guitard S, Wingert A, et al. Digital technology distraction for acute pain in children: a meta-analysis. Pediatrics 2020 Feb 22;145(2):- [FREE Full text] [doi: 10.1542/peds.2019-1139] [Medline: 31969473]
5. Won AS, Bailey J, Bailenson J, Tataru C, Yoon I, Golianu B. Immersive virtual reality for pediatric pain. Children (Basel) 2017 Jun 23;4(7):52 [FREE Full text] [doi: 10.3390/children4070052] [Medline: 28644422]
6. Lambert V, Boylan P, Boran L, Hicks P, Kirubakaran R, Devane D, et al. Virtual reality distraction for acute pain in children. Cochrane Database Syst Rev 2020 Oct 22;10:-. [doi: 10.1002/14651858.CD010686.pub2] [Medline: 33089901]
7. Colloca L, Raghuraman N, Wang Y, Akintola T, Brawn-Cinani B, Colloca G, et al. Virtual reality: physiological and behavioral mechanisms to increase individual pain tolerance limits. Pain 2020 Sep 01;161(9):2010-2021. [doi: 10.1097/j.pain.0000000000001900] [Medline: 32345915]
8. Gold JI, Mahrer NE. Is virtual reality ready for prime time in the medical space? A randomized control trial of pediatric virtual reality for acute procedural pain management. J Pediatr Psychol 2018 Apr 01;43(3):266-275. [doi: 10.1093/jpepsy/jsx129] [Medline: 29053848]

9. Trost Z, France C, Anam M, Shum C. Virtual reality approaches to pain: toward a state of the science. Pain 2021 Feb 01;162(2):325-331. [doi: 10.1097/j.pain.0000000000002060] [Medline: 32868750]
10. Bahat HS, Takasaki H, Chen X, Bet-Or Y, Treleaven J. Cervical kinematic training with and without interactive VR training for chronic neck pain - a randomized clinical trial. Man Ther 2015 Feb;20(1):68-78. [doi: 10.1016/j.math.2014.06.008] [Medline: 25066503]
11. Harvie DS, Broecker M, Smith RT, Meulders A, Madden VJ, Moseley GL. Bogus visual feedback alters onset of movement-evoked pain in people with neck pain. Psychol Sci 2015 Apr 17;26(4):385-392. [doi: 10.1177/0956797614563339] [Medline: 25691362]
12. Kulkarni J, Pettifer S, Turner S, Richardson C. An investigation into the effects of a virtual reality system on phantom limb pain: a pilot study. Br J Pain 2020 May 02;14(2):92-97. [doi: 10.1177/2049463719859913] [Medline: 32537147]
13. Thøgersen M, Andoh J, Milde C, Graven-Nielsen T, Flor H, Petrini L. Individualized augmented reality training reduces phantom pain and cortical reorganization in amputees: a proof of concept study. J Pain 2020 Nov;21(11-12):1257-1269. [doi: 10.1016/j.jpain.2020.06.002] [Medline: 32574786]
14. Austin PD, Craig A, Middleton JW, Tran Y, Costa DSJ, Wrigley PJ, et al. The short-term effects of head-mounted virtual-reality on neuropathic pain intensity in people with spinal cord injury pain: a randomised cross-over pilot study. Spinal Cord 2020 Oct 19:-. [doi: 10.1038/s41393-020-00569-2] [Medline: 33077900]
15. Richardson EJ, McKinley EC, Rahman AKMF, Klebine P, Redden DT, Richards JS. Effects of virtual walking on spinal cord injury-related neuropathic pain: a randomized, controlled trial. Rehabil Psychol 2019 Feb;64(1):13-24. [doi: 10.1037/rep0000246] [Medline: 30407030]
16. Villiger M, Bohli D, Kiper D, Pyk P, Spillmann J, Meilick B, et al. Virtual reality-augmented neurorehabilitation improves motor function and reduces neuropathic pain in patients with incomplete spinal cord injury. Neurorehabil Neural Repair 2013 Oct 10;27(8):675-683. [doi: 10.1177/1545968313490999] [Medline: 23757298]
17. Trost Z, Parsons TD. Beyond distraction: virtual reality graded exposure therapy as treatment for pain-related fear and disability in chronic pain. J Appl Biobehav Res 2014 Jun 10;19(2):106-126. [doi: 10.1111/jabr.12021]
18. Trost Z, Zielke M, Guck A, Nowlin L, Zakhidov D, France CR, et al. The promise and challenge of virtual gaming technologies for chronic pain: the case of graded exposure for low back pain. Pain Manag 2015 May;5(3):197-206. [doi: 10.2217/pmt.15.6] [Medline: 25971643]
19. Carl E, Stein AT, Levihn-Coon A, Pogue JR, Rothbaum B, Emmelkamp P, et al. Virtual reality exposure therapy for anxiety and related disorders: ameta-analysis of randomized controlled trials. J Anxiety Disord 2019 Jan;61:27-36. [doi: 10.1016/j.janxdis.2018.08.003] [Medline: 30287083]
20. Kothgassner OD, Felnhofer A. Lack of research on efficacy of virtual reality exposure therapy (VRET) for anxiety disorders in children and adolescents : a systematic review. Neuropsychiatr 2020 May 05:-. [doi: 10.1007/s40211-020-00349-7] [Medline: 32372291]
21. Krijn M, Emmelkamp P, Olafsson R, Biemond R. Virtual reality exposure therapy of anxiety disorders: a review. Clin Psychol Rev 2004 Jul;24(3):259-281. [doi: 10.1016/j.cpr.2004.04.001] [Medline: 15245832]
22. Parsons TD, Rizzo AA. Affective outcomes of virtual reality exposure therapy for anxiety and specific phobias: a meta-analysis. J Behav Ther Exp Psychiatry 2008 Sep;39(3):250-261. [doi: 10.1016/j.jbtep.2007.07.007] [Medline: 17720136]
23. Powers MB, Emmelkamp PM. Virtual reality exposure therapy for anxiety disorders: a meta-analysis. J Anxiety Disord 2008 Apr;22(3):561-569. [doi: 10.1016/j.janxdis.2007.04.006] [Medline: 17544252]
24. Won AS, Tataru CA, Cojocaru CM, Krane EJ, Bailenson JN, Niswonger S, et al. Two virtual reality pilot studies for the treatment of pediatric CRPS. Pain Med 2015 Aug 01;16(8):1644-1647. [doi: 10.1111/pme.12755] [Medline: 25930099]
25. Shiri S, Feintuch U, Weiss N, Pustilnik A, Geffen T, Kay B, et al. A virtual reality system combined with biofeedback for treating pediatric chronic headache-a pilot study. Pain Med 2013 May 01;14(5):621-627. [doi: 10.1111/pme.12083] [Medline: 23659372]
26. Darnall BD, Krishnamurthy P, Tsuei J, Minor JD. Self-administered skills-based virtual reality intervention for chronic pain: randomized controlled pilot study. JMIR Form Res 2020 Jul 07;4(7):- [FREE Full text] [doi: 10.2196/17293] [Medline: 32374272]
27. Harrison LE, Pate JW, Richardson PA, Ickmans K, Wicksell RK, Simons LE. Best-evidence for the rehabilitation of chronic pain Part 1: pediatric pain. J Clin Med 2019 Aug 21;8(9):1267 [FREE Full text] [doi: 10.3390/jcm8091267] [Medline: 31438483]
28. McLaughlin JE, Wolcott MD, Hubbard D, Umstead K, Rider TR. A qualitative review of the design thinking framework in health professions education. BMC Med Educ 2019 Apr 04;19(1):98 [FREE Full text] [doi: 10.1186/s12909-019-1528-8] [Medline: 30947748]
29. Griffin A, Wilson L, Feinstein AB, Bortz A, Heirich MS, Gilkerson R, et al. Virtual reality in pain rehabilitation for youth with chronic pain: pilot feasibility study. JMIR Rehabil Assist Technol 2020 Nov 23;7(2):- [FREE Full text] [doi: 10.2196/22620] [Medline: 33226346]
30. Olbrecht VA, Williams SE, O'Conor KT, Boehmer CO, Marchant GW, Glynn SM, et al. Guided relaxation-based virtual reality versus distraction-based virtual reality or passive control for postoperative pain management in children and

adolescents undergoing Nuss repair of pectus excavatum: protocol for a prospective, randomized, controlled trial (FOREVR Peds trial). Br Med J 2020. [doi: 10.1101/2020.09.11.20192708]
31. Olbrecht VA, O'Conor KT, Williams SE, Boehmer CO, Marchant GW, Glynn SM, et al. Transient reductions in postoperative pain and anxiety using virtual reality in children. Br Med J 2019:2724. [doi: 10.1101/2020.09.18.20192724]
32. Olbrecht VA, O'Conor KT, Williams SE, Boehmer CO, Marchant GW, Glynn SM, et al. Guided relaxation-based virtual reality transiently reduces acute postoperative pain and anxiety in a pediatric population. Br Med J 2019:2740. [doi: 10.1101/2020.09.18.20192740]
33. Litwin S, Nguyen C, Hundert A, Stuart S, Liu D, Maguire B, et al. Virtual reality to reduce procedural pain during IV insertion in the pediatric emergency department: a pilot randomized controlled trial. Clin J Pain 2021 Feb 01;37(2):94-101. [doi: 10.1097/AJP.0000000000000894] [Medline: 33177370]
34. Birnie KA, Kulandaivelu Y, Jibb L, Hroch P, Positano K, Robertson S, et al. Usability testing of an interactive virtual reality distraction intervention to reduce procedural pain in children and adolescents with cancer [formula: see text]. J Pediatr Oncol Nurs 2018 Jun 27;35(6):406-416. [doi: 10.1177/1043454218782138] [Medline: 29950139]
35. Stinson JBK, Hundert A, Abla O, Positano K, Cassiani C, Lloyd S, et al. Using virtual reality to reduce procedural pain in children with cancer: a pilot randomized controlled trial. SIOP (Virtual Conference); October 14-17. 2020. URL: https://www.canadiancancertrials.ca/Trial/Default.aspx?TrialId=NCT02929771&lang=en [accessed 2021-03-12]
36. Gold JI, Belmont KA, Thomas DA. The neurobiology of virtual reality pain attenuation. Cyberpsychol Behav 2007 Aug;10(4):536-544. [doi: 10.1089/cpb.2007.9993] [Medline: 17711362]
37. Gold JI, Kant AJ, Kim SH, Rizzo AS. Virtual anesthesia: the use of virtual reality for pain distraction during acute medical interventions. Seminars in Anesthesia, Perioperative Medicine and Pain 2005 Dec;24(4):203-210. [doi: 10.1053/j.sane.2005.10.005]
38. Li A, Montaño Z, Chen VJ, Gold JI. Virtual reality and pain management: current trends and future directions. Pain Manag 2011 Mar;1(2):147-157 [FREE Full text] [doi: 10.2217/pmt.10.15] [Medline: 21779307]
39. Mighty MDM. Mighty Immersion Inc. URL: https://www.mightymdm.com/ [accessed 2021-03-12]
40. Pham Q, Wiljer D, Cafazzo JA. Beyond the randomized controlled trial: a review of alternatives in mhealth clinical trial methods. JMIR Mhealth Uhealth 2016 Sep 09;4(3):e107 [FREE Full text] [doi: 10.2196/mhealth.5720] [Medline: 27613084]
41. HEAL public access and data sharing 2020. Health NIo. URL: https://heal.nih.gov/about/public-access-data [accessed 2021-03-12]
42. Vlaeyen JWS, Wicksell RK, Simons LE, Gentili C, De TK, Tate RL, et al. From boulder to stockholm in 70 years: single case experimental designs in clinical research. Psychol Rec 2020 Jun 03;70(4):659-670. [doi: 10.1007/s40732-020-00402-5]
43. Carlsson AM. Assessment of chronic pain. I. Aspects of the reliability and validity of the visual analogue scale. Pain 1983;16(1):87-101. [doi: 10.1016/0304-3959(83)90088-x]
44. Castarlenas E, Jensen MP, von Baeyer CL, Miró J. Psychometric properties of the numerical rating scale to assess self-reported pain intensity in children and adolescents: a systematic review. Clin J Pain 2017 Apr;33(4):376-383. [doi: 10.1097/AJP.0000000000000406] [Medline: 27518484]
45. Miettinen T, Kautiainen H, Mäntyselkä P, Linton SJ, Kalso E. Pain interference type and level guide the assessment process in chronic pain: categorizing pain patients entering tertiary pain treatment with the Brief Pain Inventory. PLoS One 2019 Aug 20;14(8):- [FREE Full text] [doi: 10.1371/journal.pone.0221437] [Medline: 31430355]
46. Askew RL, Cook KF, Revicki DA, Cella D, Amtmann D. Evidence from diverse clinical populations supported clinical validity of PROMIS pain interference and pain behavior. J Clin Epidemiol 2016 May;73:103-111 [FREE Full text] [doi: 10.1016/j.jclinepi.2015.08.035] [Medline: 26931296]
47. Kashikar-Zuck S, Flowers S, Claar R, Guite J, Logan D, Lynch-Jordan A, et al. Clinical utility and validity of the Functional Disability Inventory among a multicenter sample of youth with chronic pain. Pain 2011 Jul;152(7):1600-1607 [FREE Full text] [doi: 10.1016/j.pain.2011.02.050] [Medline: 21458162]
48. Binkley JM, Stratford PW, Lott SA, Riddle DL. The Lower Extremity Functional Scale (LEFS): scale development, measurement properties, and clinical application. North American Orthopaedic Rehabilitation Research Network. Phys Ther 1999 Apr;79(4):371-383. [Medline: 10201543]
49. Chesworth BM, Hamilton CB, Walton DM, Benoit M, Blake TA, Bredy H, et al. Reliability and validity of two versions of the upper extremity functional index. Physiother Can 2014 Jul;66(3):243-253 [FREE Full text] [doi: 10.3138/ptc.2013-45] [Medline: 25125777]
50. Law M, Baptiste S, McColl M, Opzoomer A, Polatajko H, Pollock N. The Canadian occupational performance measure: an outcome measure for occupational therapy. Can J Occup Ther 1990 Apr 22;57(2):82-87. [doi: 10.1177/000841749005700207] [Medline: 10104738]
51. Heathcote LC, Bhandari RP, Timmers I, Harrison LE, Simons LE. Rapid identification and clinical indices of fear-avoidance in youth with chronic pain. Pain 2020 Mar 29;161(3):565-573. [doi: 10.1097/j.pain.0000000000001742] [Medline: 31688496]
52. Lundberg MKE, Styf J, Carlsson SG. A psychometric evaluation of the Tampa Scale for Kinesiophobia — from a physiotherapeutic perspective. Physiother Theory Pract 2009 Jul 10;20(2):121-133. [doi: 10.1080/09593980490453002]

53. Crombez G, Bijttebier P, Eccleston C, Mascagni T, Mertens G, Goubert L, et al. The child version of the pain catastrophizing scale (PCS-C): a preliminary validation. Pain 2003 Aug;104(3):639-646. [doi: 10.1016/S0304-3959(03)00121-0] [Medline: 12927636]
54. Thompson ER. Development and validation of an internationally reliable short-form of the Positive and Negative Affect Schedule (PANAS). J Cross Cult Psychol 2016 Jul 26;38(2):227-242. [doi: 10.1177/0022022106297301]
55. Silverman WK, Fleisig W, Rabian B, Peterson RA. Childhood Anxiety Sensitivity Index. J Clin Child Psycho 1991 Jun;20(2):162-168. [doi: 10.1207/s15374424jccp2002_7]
56. Ferguson L, Scheman J. Patient global impression of change scores within the context of a chronic pain rehabilitation program. J Pain 2009 Apr;10(4):S73. [doi: 10.1016/j.jpain.2009.01.258]
57. Black WR, DiCesare CA, Thomas S, Pfeiffer M, Williams SE, Kitchen K, et al. Preliminary evidence for the Fibromyalgia Integrative Training Program (FIT Teens) improving strength and movement biomechanics in juvenile fibromyalgia: secondary analysis and results from a pilot randomized clinical trial. Clin J Pain 2021 Jan;37(1):51-60. [doi: 10.1097/AJP.0000000000000888] [Medline: 33093338]
58. Wilson AC, Palermo TM. Physical activity and function in adolescents with chronic pain: a controlled study using actigraphy. J Pain 2012 Feb;13(2):121-130 [FREE Full text] [doi: 10.1016/j.jpain.2011.08.008] [Medline: 22099608]
59. Welcome to Invincikids. Invincikids. URL: https://www.invincikids.org/ [accessed 2021-03-12]

Abbreviations

AR: augmented reality
BCH: Boston Children's Hospital
CCHMC: Cincinnati Children's Hospital Medical Center
CDE: common data element
CHARIOT: Childhood Anxiety Reduction through Innovation and Technology
CHLA: Children's Hospital of Los Angeles
INOVATE-Pain: Interdisciplinary Network on Virtual and Augmented Technologies for Pain management
IP: intellectual property
NIH: National Institutes of Health
PReP: Pediatric Rehabilitation Program
PT: physical therapist
QI: quality improvement
SickKids: The Hospital for Sick Children
VE: virtual experience
VR: virtual reality
VR-D: distraction-based virtual reality
VRET: virtual reality–based exposure therapy
VR-GR: guided-relaxation virtual reality

Edited by R Kukafka; submitted 23.11.20; peer-reviewed by T Jones, R Lee; comments to author 02.12.20; revised version received 14.12.20; accepted 04.03.21; published 26.04.21.

Please cite as:
Logan DE, Simons LE, Caruso TJ, Gold JI, Greenleaf W, Griffin A, King CD, Menendez M, Olbrecht VA, Rodriguez S, Silvia M, Stinson JN, Wang E, Williams SE, Wilson L
Leveraging Virtual Reality and Augmented Reality to Combat Chronic Pain in Youth: Position Paper From the Interdisciplinary Network on Virtual and Augmented Technologies for Pain Management
J Med Internet Res 2021;23(4):e25916
URL: https://www.jmir.org/2021/4/e25916
doi:10.2196/25916
PMID:33667177

Tutorial

The Healing Hearts Together Randomized Controlled Trial and the COVID-19 Pandemic: A Tutorial for Transitioning From an In-Person to a Web-Based Intervention

Kathleen Lalande[1], PhD; Paul S Greenman[2,3,4], PhD; Karen Bouchard[1], PhD; Susan M Johnson[4], EdD; Heather Tulloch[1], PhD

[1]Division of Cardiac Prevention and Rehabilitation, University of Ottawa Heart Institute, Ottawa, ON, Canada

[2]Université du Québec en Outaouais, Gatineau, QC, Canada

[3]Institut du Savoir Montfort, Ottawa, ON, Canada

[4]International Centre for Excellence in Emotionally Focused Therapy, Ottawa, ON, Canada

Corresponding Author:
Heather Tulloch, PhD
Division of Cardiac Prevention and Rehabilitation
University of Ottawa Heart Institute
40 Ruskin St
Ottawa, ON, K1Y 4W7
Canada
Phone: 1 613 696 7000 ext 19705
Email: hetulloch@ottawaheart.ca

Abstract

Supportive couple relationships are associated with reduced risk of chronic illness development, such as cardiovascular disease, as well as improved secondary prevention. Healing Hearts Together (HHT) is an 8-week couples-based intervention designed to improve relationship quality, mental health, quality of life, and cardiovascular health among couples in which one partner has experienced a cardiac event. A randomized controlled trial began in October 2019 to test the efficacy of the in-person, group-based HHT program as compared to usual care. In March of 2020, all recruitment, assessments, and interventions halted due to the COVID-19 pandemic. Guided by optimal virtual care principles, as well as by Hom and colleagues' four-stage framework—consultation, adaptation, pilot-testing, and test launch—this paper is a tutorial for the step-by-step transition planning and implementation of a clinical research intervention from an in-person to a web-based format, using the HHT program as an example. Clinical and research considerations are reviewed, including (1) privacy, (2) therapeutic aspects of the intervention, (3) group cohesion, (4) research ethics, (5) participant recruitment, (6) assessment measures, (7) data collection, and (8) data analyses. This tutorial can assist clinical researchers in transitioning their research programs to a web-based format during the pandemic and beyond.

(J Med Internet Res 2021;23(4):e25502) doi:10.2196/25502

KEYWORDS

web-based intervention; internet-based intervention; randomized controlled trial; COVID-19; research; tutorial; digital medicine; behavioral medicine; telehealth; telemedicine; cardiovascular rehabilitation

Introduction

This tutorial aims to help clinical researchers transition their in-person research programs to a web-based format during the pandemic and beyond. To our knowledge, no literature exists on how to navigate the steps required to transition a randomized controlled trial (RCT), designed for in-person delivery, to an online world. No one has outlined, for example, contacts with sponsors and research ethics boards, recruitment procedures, or informed consent changes. This paper expands the previous transitional work in a clinical-only environment by accounting for the unique demands of clinical research and RCTs. The Healing Hearts Together (HHT) RCT provides researchers a current example of this challenging transition. We begin by briefly describing the original in-person HHT intervention and research protocol. We then guide clinical researchers through five important stages of transition: (1) consultation and assessment of needs, (2) adaptation of procedures and materials

for web-based delivery, (3) adaptation of procedures and materials for the research protocol, (4) pilot sessions, and (5) the final launch of the web-based intervention and research protocol. We have also included a useful checklist of all the critical clinical and research elements that must be addressed so that the transition can be done in an efficient and ethical manner (see Multimedia Appendix 1 for checklist).

Original In-Person HHT Program

A healthy couple relationship (ie, one in which partners feel loved, emotionally supported, respected, and cared for) is a significant protective factor, particularly against cardiac disease incidence and outcomes [1-5]. Healthy relationships can influence positive cardiac outcomes in direct and indirect ways. Direct examples include improved high-frequency heart rate variability and decreased diastolic blood pressure in supportive relationships [6-8]. Indirect links to cardiovascular health include adaptive behavioral pathways, such as partners influencing health behaviors (eg, prepare low sodium meals), modeling healthy behavior (eg, exercise), or assisting in the management of disease (eg, medication management) [9-11]. In contrast to their more distressed peers, happily married cardiac patients demonstrate stronger adherence to blood pressure medication regimens and cardiac rehabilitation programs [12,13].

Unfortunately, the reverse is also true. Relationship distress can have a negative impact on heart disease. A recent meta-analysis showed that poor social relationships were associated with a 29% increase in risk of incident coronary artery disease [14]. Uchino and colleagues found that coronary artery calcification scores were higher for individuals who expressed ambivalence rather than positivity about their couple relationship [15]. Other researchers have highlighted the indirect pathways by which relationship distress influences cardiac outcomes, including smoking and alcohol use in response to relationship problems [16].

Despite burgeoning evidence indicating that healthy relationships are vital for reducing chronic disease incidence and management, current secondary prevention programming inadvertently neglects a crucial resource for disease management: the patient's partner. In order to address this gap in clinical care, the HHT program [17] was created with the aim of helping both cardiac patients and their partners better manage cardiac disease by strengthening the emotional bond between them. The goals of HHT are to help couples improve their relationship quality, mental health, quality of life, and cardiovascular health. The HHT couples-based intervention is adapted from the Hold Me Tight: Conversations for Connection (HMT) program [18], an intervention based on emotionally focused therapy (EFT) for couples, which is an empirically supported treatment for relationship distress [19,20]. EFT, HMT, and HHT are based on attachment theory, which states that humans have an innate need for close emotional bonds to significant others [10,11]. This need becomes especially pertinent when faced with a threat or stress (eg, a cardiac event); it triggers proximity seeking to "attachment figures" in order to regulate affect. EFT interventions help couples identify and articulate their vulnerability (eg, "I almost died/I almost lost you") and respond to these feelings in comforting ways, thereby solidifying their sense of security and emotional connection [19,21]. Couples who achieve this connection are happier in their relationships and are more effective problem solvers (eg, managing cardiovascular disease) [22-24]. Inspired by HMT, the HHT program guides couples through seven conversations, based on EFT principles, in which they learn to clearly communicate their need for connection and reassurance, with a focus on heart disease and healthy coping.

Preliminary results from a proof-of-concept study indicated that couples who participated in the HHT program reported significant improvements in relationship quality, mental health, and quality of life [25]. Based on these promising results and to assess the efficacy of the HHT program on a larger sample with additional cardiovascular outcomes, an RCT was initiated. Eligible patients and their partners who consent to participation undergo a baseline assessment and then are randomly assigned to either the HHT program or usual care at the hospital. Usual care participants are followed by their physician for clinical assessment and are referred to the standard programming at the center. All participants are reassessed at 8 and 24 weeks.

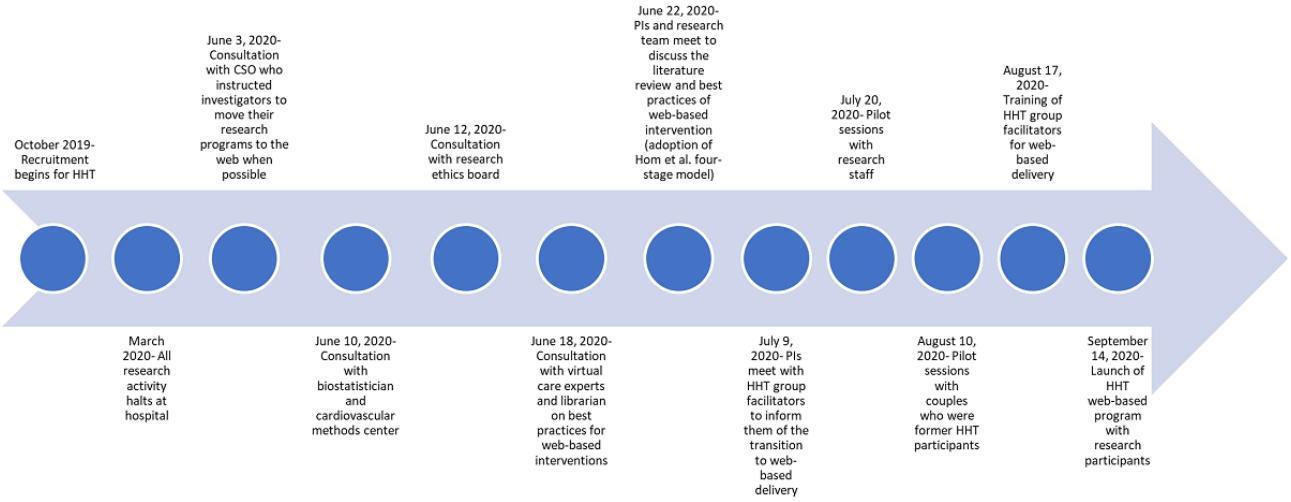
Prior to the COVID-19 pandemic, participants who were randomly assigned to the HHT group attended, in person, eight weekly 2-hour sessions led by two facilitators. The eight sessions focus on understanding love, attachment, and their relationship to heart health; provide an opportunity to share experiences related to cardiovascular disease with partners and peers; and assist in identifying and improving communication patterns that may inhibit positive interactions and healthy behaviors. Throughout each 2-hour session, participants are introduced to concepts through didactic presentations, videos, group and couple discussions, and homework exercises. Additional study procedures and intervention content will be reviewed below as the in-person and web-based HHT programs are compared.

Stage I: Consultation and Assessment of Needs for the Intervention and Research Protocol

Overview

During the first few months of the COVID-19 lockdown in March, April, and May of 2020, the chief scientific officer stopped all research in the hospital. At that time, it was not clear that HHT would have to transition to a web-based platform; many researchers were hoping that the situation with COVID-19 would stabilize within a few months. However, in June 2020, researchers at the hospital were asked to adapt their research protocols to allow for as much virtual care as possible. Following these directives, the HHT research team began to focus on how to do this while following best practices and preserving the integrity of the trial (see Figure 1 for a timeline of the transition to a web-based intervention during COVID-19).

Figure 1. Timeline of the transition to a web-based intervention during COVID-19. CSO: chief scientific officer; HHT: Healing Hearts Together; PI: principal investigator.



Literature Review

The principal investigators of HHT met at the beginning of June 2020 to outline a plan for transitioning to a web-based intervention as well as to review all study procedures to identify areas of required revision (eg, recruitment procedures and presentation of intervention content online). An examination of the current literature to aid in this transition was completed. We were able to identify best practices for virtual care (1) by seeking the help of a research librarian, who assisted with a literature review of resources on how to transition from an in-person psychological intervention to a web-based format, and (2) by consulting with internationally renowned experts in virtual care, who also directed our attention to pertinent literature.

The literature review yielded a small number of resources that provided a framework for the transition [26-33]. More specifically, these articles highlighted key areas of telehealth development, including patient involvement [29], technological support and training for the group facilitators [27], a strong working alliance during web-based therapy [33], ethical considerations [30], and the importance of comprehensive information technology (IT) training and support [28]. A systematic review of telehealth interventions delivering home-based support with group videoconferencing highlighted several benefits of web-based interventions, including improved accessibility, engagement with others facing similar challenges (eg, cardiovascular disease), development of health knowledge and skills, and improved mental health outcomes similar to face-to-face groups [28]. Finally, Hom and colleagues [26] provided a useful framework for modifying a clinical psychiatric program for a web-based platform using four stages: consultation, adaptation, pilot-testing, and test launch.

Research Staff Communication

Following the literature review and consultation with virtual care experts, a focus on ensuring open and clear communication among the members of the research team, who were all working from home, was needed. We chose a web-based collaborative workspace (eg, Trello and Slack) to streamline staff communication and to assign tasks and deadlines to research staff. This avoided the hassle and confusion of sending multiple emails regarding project updates. Instead, we simply log in to our web-based communication board to view the task list to be completed and by whom. Regular weekly lab meetings were also transitioned to a videoconferencing platform. These meetings assisted in the organization of deliverables and facilitated lab staff cohesion, morale, and motivation during these unprecedented times.

Updating the Funding Agency and Ethics Board

Shortly after research was halted at the hospital, the study's federal funding agency reached out to the nominated principal investigator proactively with an update reassuring all researchers that the funding window would be extended due to the COVID-19 crisis, and that no extra report or documentation would be required. The research coordinator updated the study registry, ClinicalTrials.gov, and the hospital's electronic medical record, noting that the study was no longer actively recruiting. Researchers may need to contact their sponsors or funding agencies to inquire about their specific requirements.

Consultation with our institutional ethics board regarding required study updates for the transition from a predominantly in-person-based study to one with virtual methods was then conducted. Within days of the shutdown, we informed the research ethics board of all changes in research activities (eg, phone contact to inform patients, no recruitment, and temporary follow-up by mail). Although the move to a web-based HHT program does lower the risk of COVID-19 transmission to our participants, there is also an increased risk of privacy breaches with an online format. In response to privacy concerns, a form outlining the changes to our methods and protocol, noting strategies to protect personal health information and privacy, was submitted to the research ethics board for approval. Once the approval was received, a form requesting permission to restart our research study was submitted to our research services department lead by the chief scientific officer of our institution. In this form, we outlined the safety protocols for the in-person baseline and follow-up assessments at the hospital, including appropriate personal protective equipment, cleaning the

workspace once participants have left, physical distancing when possible, and equipment changes to ensure that COVID-19 is not transmitted (eg, new valve in the carbon monoxide monitor). Our research staff attended a consenting and documentation seminar for studies conducting virtual recruitment, consent, and interventions at our facility. The principal investigator completed an online training seminar on how to use the hospital-approved secure videoconferencing platform linked with the hospital's secure electronic medical record software. This information was subsequently shared with staff.

Stage II: Adaptation of Procedures and Materials for Web-Based Delivery

Optimal Web-Based Care Principles

To ensure that we were following best practices in the provision of psychological services via telepsychology, we referred to the guidelines provided by the American Psychological Association [34], the Canadian Psychological Association, and the Ontario Psychological Association, in addition to the previously mentioned resources from the literature review [26-32]. For the purposes of this tutorial, we will highlight the best practices that were most relevant to our study and the adjustments that were required for the research protocol.

Provincial Jurisdiction of Psychological Services

Patients from the neighboring province of Quebec frequently seek services at our Ontario hospital. Because Quebec patients would no longer be able to come to the hospital to participate in the HHT program, the principal investigator contacted the Order of Psychologists of Quebec to obtain their permission to provide services to Quebec residents via a web-based platform. Due to the unusual pandemic circumstances, this permission was swiftly granted. It is recommended that clinician researchers ensure they have the appropriate credentials to provide care in the jurisdiction of the patient.

Security and Transmission of Data and Information

As per the directives at our institution, the hospital-approved videoconferencing platform was employed to run the HHT groups online. The principal investigators reviewed every session of the HHT program and determined that the intervention elements could be delivered securely via videoconferencing.

Appropriate Medium of Delivery

Another crucial factor in successful web-based therapy is the existence of adequate technology, equipment, and usability for research staff and participants. Previous research has shown that web-based interventions face common IT and visual issues, such as audio difficulties, delays, dropouts, background noise, difficulties downloading software, and poor lighting [28]. In response to these common challenges, a checklist for facilitators and potential participants was created to confirm that they had the necessary equipment, technology, and space to do HHT in a web-based format in a secure and private manner. It includes questions that assess whether they have (1) internet access (ie, Wi-Fi or ethernet) at home; (2) two devices that allow for videoconferencing (ie, camera and microphone), such as a desktop computer or laptop, a tablet or iPad, a Chromebook, or a smartphone; and (3) a quiet and private room in their residence.

Stage III: Adaptation of Procedures and Materials for the Research Protocol

Recruitment

Before the pandemic, recruitment was conducted in person at the hospital when patients came to see their cardiologists for follow-up appointments. Members of the research staff spoke with patients after their appointments and presented an overview of the HHT program. Interested patients agreed to be contacted by study staff, who described the study in greater detail and answered any questions. If still interested, an initial assessment appointment was scheduled. With the new COVID-19 restrictions, most clinical care is done virtually (ie, phone or video appointments). A new standard operating procedure for HHT recruitment was developed outlining the steps to call patients who have already given general consent to be contacted for research purposes at our center. With no in-person clinical visits, recruitment is now completed entirely over the phone. Members of the research staff call potential participants shortly after their phone consultation with their cardiologist and review the technology checklist before proceeding with the recruitment script. If patients do not have the necessary computer equipment, they are not eligible for the study at this time. As the pandemic restrictions ease, patients without the necessary technology will be provided the option to engage in person if randomized to the HHT condition.

Informed Consent

As our study is nonregulated and low risk, our institutional ethics review board informed us that we could adopt a well-documented verbal consent process that permits the research coordinator to read the full written informed consent form to potential participants via telephone. Participants can then ask the research coordinator any questions they have about the study before granting verbal consent. The research coordinator must document the consent process in the participant's research file and electronic medical record, with the help of a verbal consent checklist. If requested by the participant, the consent process may also be completed in person at the time of the assessment at the hospital. In this case, the in-person consent process will still be completed verbally with no transfer of paper between staff and participants.

Other secure methods of obtaining consent via electronic signature software and electronic consent platforms (eg, Research Electronic Data Capture [REDCap] e-consent framework, DocuSign, and Adobe Sign) may also be considered. The advantage of these systems is that they provide a user-friendly option for individuals to personally sign the consent form. However, there are potential disadvantages that include cost, the need for additional institutional approval, and complexity barriers for less tech-savvy individuals. Finally, there is the traditional method of obtaining informed consent by which the study team establishes a process, in alignment with research regulations, whereby the paper consent form is mailed to the participant and, following the consent discussion,

the participant signs the paper consent form and mails it back to the institution. Researchers who are working with external sponsors need to verify with their institution and their sponsor the appropriate informed consent process to use during the pandemic.

Assessment

In the original protocol, patients are asked to come to the hospital for a baseline assessment and two follow-up assessments. Physiological measurements are taken, including heart rate variability, blood pressure, height, weight, waist circumference, carbon monoxide concentration, and salivary samples. In addition, all the couples are asked to participate in a conflict resolution task (ie, a discussion of topics that engender conflict, such as finances, in-laws, or jealousy) that is video-recorded for analysis. The research team initially explored ways to complete these measures remotely, but realized it was not possible for many reasons (eg, cost, coordination of assessment, and variability in measurement). To facilitate coordination of staff on site, the hospital research support office implemented a shared calendar for staff to indicate when they would be on site so as to not exceed maximum pandemic capacity and to allow the coordination of clinical research activity among staff.

Stage IV: Pilot-Testing and Streamlining of Procedures

Pilot Tests

Once the HHT in-person intervention was adapted for web-based administration, the principal investigators (ie, trained group facilitators) ran pilot web-based HHT group sessions with volunteers from the research staff and their partners to identify any issues or problems. During these sessions, features of the videoconferencing platform were explored and problems addressed (eg, chat function and breakout rooms). During the traditional in-person format, the facilitators can quietly and respectfully check in with couples who are working privately on an in-class exercise together. The facilitators realized, however, that without the in-person visual cues, it would prove difficult to identify couples in need of assistance during the breakout sessions or to enter these sessions unannounced. Couples also have the option of notifying their hosts that they are in need of assistance. In addition, to prevent an intrusive interruption, facilitators now provide participants with a screen announcement—audio and video off—letting them know that a facilitator would like to join their breakout room. Next, permission to join is requested via audio, before their video image is shown. It is important to note that videoconferencing services now have excellent online support centers with helpful videos and tutorials where researchers can learn and explore the features and strategies that exist to enhance their web-based intervention. The lessons learned from the pilot tests were used to create separate tutorials for group facilitators and participants.

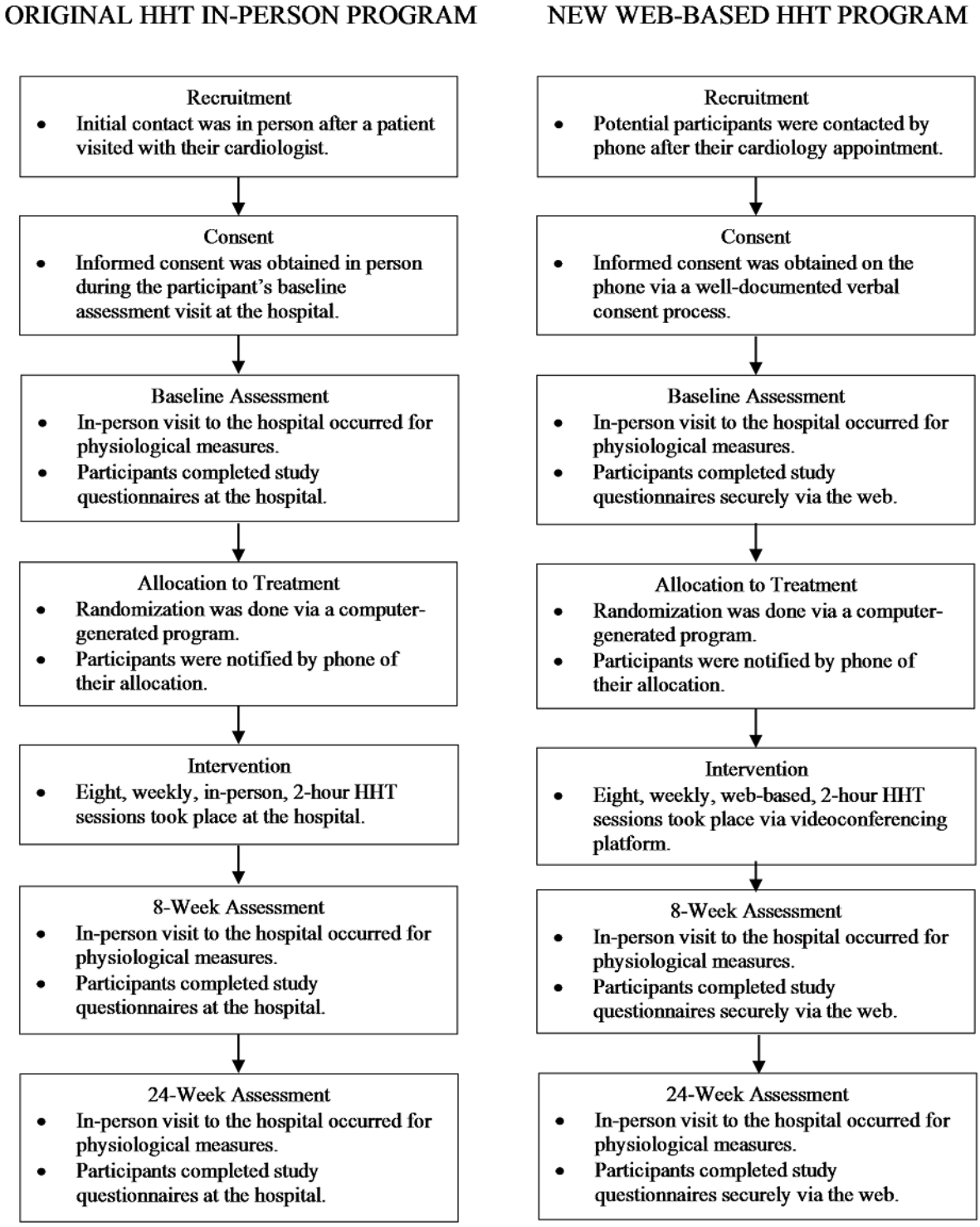
Test Launch With Previous HHT Participants

After running the pilot sessions with volunteer research staff and their partners, the principal investigators conducted a pilot HHT session with couples who were previous participants of the in-person HHT program. The research coordinator contacted these individuals in advance to review the technology checklist and to provide a brief tutorial on how to use the videoconferencing platform. By consulting previous participants of the traditional in-person HHT group, the facilitators gained instructive feedback that contributed to treatment fidelity and acceptability of the web-based format [29]. In an informal postsession feedback conversation, the couples provided suggestions, including (1) visual format changes (ie, reduce the slide size) in order to make the presenter's face more prominent to facilitate a personal connection, (2) offer participants a choice between receiving the printed session handouts by mail or digitally via email, and (3) purposefully enhance and monitor group cohesion. Methods to enhance group cohesion online have been suggested recently [35].

Lessons Learned From the Pilot Sessions

The most helpful step, of those outlined by Hom and colleagues [26], was pilot-testing the intervention to staff and previous HHT participants. Facilitators learned that to maximize group cohesion, maximize flow of conversation, and allow participants to contribute substantially to meaningful discussions, limiting the group to a maximum of 5 couples was required. During the sessions themselves, facilitators found it easier to manage the group dynamics by disabling the chat feature—except with the hosts—and by requesting that the participants use their mute function during the didactic and presentation portion of the session to minimize distraction and hearing difficulties. Finally, facilitators realized that it will be important to seek ongoing feedback from all the participants regarding IT issues or anything related to the web-based format; to that end, they employed a brief questionnaire for participants to fill out after the completion of the HHT program (see Multimedia Appendix 2 for the HHT Client Satisfaction Survey). With the pilot sessions completed, the new web-based HHT program was ready to launch (see Figure 2 for a comparison of the original in-person HHT program and the web-based HHT program).

Figure 2. A comparison of the original in-person Healing Hearts Together (HHT) program and the web-based HHT program.



Stage V: Launch of the Web-Based HHT Intervention and Research Protocol

Baseline Assessment

The baseline assessment visit is conducted at the hospital. Participants scheduled for the assessment are now screened for COVID-19 over the phone within 24 hours of their visit. Once they arrive at the hospital, they are rescreened at the hospital entrance and provided with a mask. Research staff wear appropriate personal protective equipment. Physical distancing regulations are respected, except when completing measurements such as heart rate variability, blood pressure, height, weight, waist circumference, carbon monoxide concentration, and saliva. Proper hygiene procedures are used at every stage of assessment. Questionnaires are completed by participants at home via a secure web application for building and managing web-based surveys and databases (eg, REDCap).

If participants decline an on-site appointment, questionnaires can still be completed online and physiological measures are not completed.

Allocation to Treatment

The randomization process has not changed with the new web-based format. Participants are randomized to the HHT program or to usual care using a computer-generated program at our cardiovascular methods center. Sequences are placed in sealed, numbered, opaque envelopes to ensure concealment of treatment allocation until after baseline data collection. The research coordinator allocates the next available number on study entry, logs all randomizations, and notifies participants of their allocated group immediately by phone. The research coordinator and participants are aware of group allocation. Research assistants, blinded to the participants' treatment allocation, conduct both follow-up assessments.

Intervention

The participants who are randomly assigned to the experimental intervention participate in the new web-based HHT group program. The research coordinator informs participants of their program start date, provides each couple with a brief tutorial of how to access the videoconferencing software, and sends them the intervention materials (eg, book and handouts) via email and/or post before the group start date. For extra technical support, participants receive an email with the prepared tip sheets to help them navigate the web-based format. The number (eight sessions) and duration (2 hours) of sessions remain the same in the web-based format. In this format, couples continue to benefit from the same didactic presentations, videos, and group and couple discussions that were offered in the traditional in-person format. After each web-based session, participants are asked to fill out a very brief feedback questionnaire.

End-of-Treatment and Follow-Up Assessment

As with the baseline assessment, participants will be permitted to go in person to the hospital for their 8-week follow-up assessments. Nevertheless, it was necessary to create a contingency plan that allows for a more flexible approach. Should more strict hospital restrictions emerge, the research staff has prepared alternative means of gathering assessment data, (eg, recording the conflict resolution task via the videoconferencing platform, using self-reported unstandardized height and weight readings, and using self-reported blood pressure readings).

CONSORT Reporting

The research coordinator updated the CONSORT data to include patients who were no longer eligible due to surpassing the eligibility window as well as patients who were unable to complete follow-up assessments due to the pandemic. In addition, revisions were made to the inclusion and exclusion criteria (eg, technology requirements and broader catchment area).

Data Management and Analysis

As COVID-19 restrictions lift, a choice for the in-person versus the web-based HHT program may be provided to participants. However, this creates a new and potentially confounding variable that will need to be included in the database and explored in statistical analyses. A dichotomous variable indicating participation in the web-based or in-person HHT program will then be used as a covariate to control for the mode of delivery. Exploratory analyses can be used to investigate whether the delivery mode made a difference by using an interaction term between delivery mode and the intervention. Advice from the study statistician regarding the above procedures was sought. Researchers should consult with appropriate statistics advisors when interim analyses are conducted and changes to the data analysis plan are considered.

Advantages of the New Web-Based Format

As we updated our protocol, several advantages of the web-based program became evident, both from a clinical and research perspective. These included the following:

1. Broader recruitment. With the necessity of weekly travel removed, the potential recruitment of patients living outside of a 1-hour radius from the hospital could now be included. Prior to the pandemic, participants who lived longer than an hour's drive from the hospital were excluded, as it was seldom feasible for couples to drive an hour or more for eight weekly group sessions, in addition to the three assessment appointments, especially in the winter months. However, with the web-based format, participants are required to attend only three assessments on site. As such, from a research perspective, it is important to note that the reduced travel burden permits potential recruitment from rural areas, which will enrich the diversity of our sample.
2. Flexible scheduling. Flexibility in scheduling both participants and facilitators for the group sessions and assessments emerged as another benefit of the HHT web-based intervention. With no need to reserve group rooms or to arrange for parking at our busy hospital, finding mutually convenient times for the web-based sessions and for the in-person assessments has been optimized. In addition, all group facilitators can be trained online at their convenience individually or as a group by the research coordinator, a process that allows for streamlined, efficient training and ongoing support for the HHT facilitators [27].
3. Flexible methodology. For several years, the gold standard for the informed consent process has involved face-to-face interactions with potential participants. With the recent increase in web-based surveys and interventions, there has been more openness to online consent forms. However, hospital-based clinical trials have always favored the traditional in-person model for obtaining consent. The COVID-19 pandemic has required hospitals and research institutions to adopt more flexible means to obtain informed consent, while assuring high ethical standards. This flexibility toward informed consent and other aspects of

methodology (eg, recruitment and web-based measures) will provide opportunities for researchers who would like to reach a broader population.

Conclusions

When the principal investigators conceptualized and created the HHT program, they envisioned an in-person group program that would help couples grow closer and work well together to improve cardiac health. The COVID-19 pandemic forced hospitals to rethink the mode of delivery for their services. While these drastic changes undoubtedly prompt stress and invoke challenges, there are opportunities for clinical research to extend the reach of recruitment to hard-to-access populations (eg, patients with chronic conditions or mobility issues and patients who have family and work obligations that do not allow them to leave their homes) and, ultimately, promote patient-centered care. It is understood that researchers around the world will also have to take into consideration the subject matter of their research, as well as the cultural, environmental, occupational, and economic factors in their home countries. Despite the diverse nature of global research, this tutorial aims to serve as a brief, yet comprehensive, framework for clinical researchers facing the challenge to offer flexible and innovative web-based interventions.

Acknowledgments

HT, PG, and SJ contributed to the development of the intervention. HT, PG, and KL coordinated the transition to a web-based intervention. KL and HT drafted the manuscript, and PG and KB critically revised the manuscript. All authors gave final approval and agree to be accountable for all aspects of the work, ensuring integrity and accuracy. This work was supported by an operating grant from the Canadian Institutes of Health Research. KB is supported by a Social Sciences and Humanities Research Council Postdoctoral Fellowship.

Conflicts of Interest

HT, SJ, and PG are coauthors of the Healing Hearts Together Relationship Program and receive royalties from its sales.

Multimedia Appendix 1
Checklist for clinical researchers transitioning their research and intervention to a web-based platform.
[PDF File (Adobe PDF File), 28 KB - jmir_v23i4e25502_app1.pdf]

Multimedia Appendix 2
Healing Hearts Together (HHT) Client Satisfaction Survey.
[PDF File (Adobe PDF File), 35 KB - jmir_v23i4e25502_app2.pdf]

References

1. Greenman P, Tassé V, Tulloch H. Straight to the Heart: Romantic relationships, attachment, and the management of cardiac disease. In: Rennolds A, editor. Psychology of Interpersonal Perception and Relationships. Hauppage, NY: Nova Science Publishers; 2015:157-174.
2. King KB, Reis HT. Marriage and long-term survival after coronary artery bypass grafting. Health Psychol 2012 Jan;31(1):55-62. [doi: 10.1037/a0025061] [Medline: 21859213]
3. Liu H, Waite L. Bad marriage, broken heart? Age and gender differences in the link between marital quality and cardiovascular risks among older adults. J Health Soc Behav 2014 Dec;55(4):403-423 [FREE Full text] [doi: 10.1177/0022146514556893] [Medline: 25413802]
4. Dhindsa DS, Khambhati J, Schultz WM, Tahhan AS, Quyyumi AA. Marital status and outcomes in patients with cardiovascular disease. Trends Cardiovasc Med 2020 May;30(4):215-220 [FREE Full text] [doi: 10.1016/j.tcm.2019.05.012] [Medline: 31204239]
5. Robles TF, Slatcher RB, Trombello JM, McGinn MM. Marital quality and health: A meta-analytic review. Psychol Bull 2014 Jan;140(1):140-187 [FREE Full text] [doi: 10.1037/a0031859] [Medline: 23527470]
6. Baucom BRW, Baucom KJW, Hogan JN, Crenshaw AO, Bourne SV, Crowell SE, et al. Cardiovascular reactivity during marital conflict in laboratory and naturalistic settings: Differential associations with relationship and individual functioning across contexts. Fam Process 2018 Sep;57(3):662-678. [doi: 10.1111/famp.12353] [Medline: 29577270]
7. Donoho CJ, Seeman TE, Sloan RP, Crimmins EM. Marital status, marital quality, and heart rate variability in the MIDUS cohort. J Fam Psychol 2015 Apr;29(2):290-295 [FREE Full text] [doi: 10.1037/fam0000068] [Medline: 25844496]
8. Cundiff JM, Birmingham WC, Uchino BN, Smith TW. Marital quality buffers the association between socioeconomic status and ambulatory blood pressure. Ann Behav Med 2016 Apr;50(2):330-335. [doi: 10.1007/s12160-015-9742-z] [Medline: 26442917]
9. DiMatteo MR. Social support and patient adherence to medical treatment: A meta-analysis. Health Psychol 2004 Mar;23(2):207-218. [doi: 10.1037/0278-6133.23.2.207] [Medline: 15008666]

10. Lewis MA, Butterfield RM. Social control in marital relationships: Effect of one's partner on health behaviors. J Appl Soc Psychol 2007 Feb;37(2):298-319. [doi: 10.1111/j.0021-9029.2007.00161.x]
11. Lewis MA, Rook KS. Social control in personal relationships: Impact on health behaviors and psychological distress. Health Psychol 1999;18(1):63-71. [doi: 10.1037/0278-6133.18.1.63]
12. Trevino DB, Young EH, Groff J, Jono RT. The association between marital adjustment and compliance with antihypertension regimens. J Am Board Fam Pract 1990;3(1):17-25. [Medline: 2305637]
13. Rankin-Esquer LA, Deeter AK, Froeliche E, Taylor CB. Coronary heart disease: Intervention for intimate relationship issues. Cogn Behav Pract 2000 Mar;7(2):212-220. [doi: 10.1016/s1077-7229(00)80034-6]
14. Valtorta NK, Kanaan M, Gilbody S, Ronzi S, Hanratty B. Loneliness and social isolation as risk factors for coronary heart disease and stroke: Systematic review and meta-analysis of longitudinal observational studies. Heart 2016 Jul 01;102(13):1009-1016 [FREE Full text] [doi: 10.1136/heartjnl-2015-308790] [Medline: 27091846]
15. Uchino BN, Smith TW, Berg CA. Spousal relationship quality and cardiovascular risk: Dyadic perceptions of relationship ambivalence are associated with coronary-artery calcification. Psychol Sci 2014 Apr;25(4):1037-1042 [FREE Full text] [doi: 10.1177/0956797613520015] [Medline: 24501110]
16. Skoyen JA, Kogan AV, Novak SA, Butler EA. Health behavior and emotion regulation in couples. In: Newman ML, Roberts NA, editors. Health and Social Relationships: The Good, the Bad, and the Complicated. Washington, DC: American Psychological Association; 2013:121-142.
17. Tulloch H, Greenman P, Demidenko N, Johnson SM. Healing Hearts Together Relationship Education Program: Facilitator's Guide for Small Groups. Ottawa, ON: ICEEFT; 2017.
18. Johnson SM. Facilitator's Guide for Hold Me Tight Groups. 2nd edition. Ottawa, ON: ICEEFT; 2010.
19. Johnson SM. The Practice of Emotionally Focused Marital Therapy: Creating Connection. Philadelphia, PA: Brunner/Mazel; 1996.
20. Wiebe SA, Johnson SM. A review of the research in emotionally focused therapy for couples. Fam Process 2016 Sep;55(3):390-407. [doi: 10.1111/famp.12229] [Medline: 27273169]
21. Johnson S. Hold Me Tight: Seven Conversations for a Lifetime of Love. New York, NY: Little, Brown and Company; 2008.
22. Dalgleish TL, Johnson SM, Burgess Moser M, Lafontaine M, Wiebe SA, Tasca GA. Predicting change in marital satisfaction throughout emotionally focused couple therapy. J Marital Fam Ther 2015 Jul;41(3):276-291. [doi: 10.1111/jmft.12077] [Medline: 24910261]
23. Feeney BC. The dependency paradox in close relationships: Accepting dependence promotes independence. J Pers Soc Psychol 2007 Feb;92(2):268-285. [doi: 10.1037/0022-3514.92.2.268] [Medline: 17279849]
24. Li T, Chan DK. How anxious and avoidant attachment affect romantic relationship quality differently: A meta-analytic review. Eur J Soc Psychol 2012 Jan 20;42(4):406-419. [doi: 10.1002/ejsp.1842]
25. Tulloch H, Johnson S, Demidenko N, Clyde M, Bouchard K, Greenman PS. An attachment-based intervention for patients with cardiovascular disease and their partners: A proof-of-concept study. Health Psychol 2020 Dec 21:1. [doi: 10.1037/hea0001034] [Medline: 33346674]
26. Hom MA, Weiss RB, Millman ZB, Christensen K, Lewis EJ, Cho S, et al. Development of a virtual partial hospital program for an acute psychiatric population: Lessons learned and future directions for telepsychotherapy. J Psychother Integr 2020 Jun;30(2):366-382. [doi: 10.1037/int0000212]
27. Inchausti F, MacBeth A, Hasson-Ohayon I, Dimaggio G. Telepsychotherapy in the age of COVID-19: A commentary. J Psychother Integr 2020 Jun;30(2):394-405. [doi: 10.1037/int0000222]
28. Banbury A, Nancarrow S, Dart J, Gray L, Parkinson L. Telehealth interventions delivering home-based support group videoconferencing: Systematic review. J Med Internet Res 2018 Feb 02;20(2):e25 [FREE Full text] [doi: 10.2196/jmir.8090] [Medline: 29396387]
29. Van Daele T, Karekla M, Kassianos AP, Compare A, Haddouk L, Salgado J, et al. Recommendations for policy and practice of telepsychotherapy and e-mental health in Europe and beyond. J Psychother Integr 2020 Jun;30(2):160-173. [doi: 10.1037/int0000218]
30. Lawlor-Savage L, Prentice JL. Digital cognitive behaviour therapy (CBT) in Canada: Ethical considerations. Can Psychol 2014;55(4):231-239. [doi: 10.1037/a0037861]
31. McDonald A, Eccles JA, Fallahkhair S, Critchley HD. Online psychotherapy: Trailblazing digital healthcare. BJPsych Bull 2020 Apr;44(2):60-66 [FREE Full text] [doi: 10.1192/bjb.2019.66] [Medline: 31685068]
32. Békés V, Aafjes-van Doorn K. Psychotherapists' attitudes toward online therapy during the COVID-19 pandemic. J Psychother Integr 2020 Jun;30(2):238-247. [doi: 10.1037/int0000214]
33. Watts S, Marchand A, Bouchard S, Gosselin P, Langlois F, Belleville G, et al. Telepsychotherapy for generalized anxiety disorder: Impact on the working alliance. J Psychother Integr 2020 Jun;30(2):208-225. [doi: 10.1037/int0000223]
34. Joint Task Force for the Development of Telepsychology Guidelines for Psychologists. Guidelines for the practice of telepsychology. Am Psychol 2013 Dec;68(9):791-800. [doi: 10.1037/a0035001] [Medline: 24341643]
35. Weinberg H. Online group psychotherapy: Challenges and possibilities during COVID-19—A practice review. Group Dyn 2020 Sep;24(3):201-211. [doi: 10.1037/gdn0000140]

Abbreviations

EFT: emotionally focused therapy
HHT: Healing Hearts Together
HMT: Hold Me Tight: Conversations for Connection
IT: information technology
RCT: randomized controlled trial
REDCap: Research Electronic Data Capture

Edited by G Eysenbach; submitted 04.11.20; peer-reviewed by A Hidki,Dr, S Sabarguna, S Lal; comments to author 19.12.20; revised version received 23.01.21; accepted 15.02.21; published 06.04.21.

Please cite as:
Lalande K, Greenman PS, Bouchard K, Johnson SM, Tulloch H
The Healing Hearts Together Randomized Controlled Trial and the COVID-19 Pandemic: A Tutorial for Transitioning From an In-Person to a Web-Based Intervention
J Med Internet Res 2021;23(4):e25502
URL: https://www.jmir.org/2021/4/e25502
doi:10.2196/25502
PMID:33729984

Original Paper

A Mobile App for Self-management of Urgency and Mixed Urinary Incontinence in Women: Randomized Controlled Trial

Towe Wadensten[1], MMSc; Emma Nyström[1], MD, PhD; Karin Franzén[2,3], MD, PhD; Anna Lindam[4], PhD; Elisabet Wasteson[5], PhD; Eva Samuelsson[1], MD, PhD

[1]Family Medicine, Department of Public Health and Clinical Medicine, Umeå University, Umeå, Sweden

[2]Women's Clinic, Örebro University Hospital, Örebro, Sweden

[3]School of Medical Sciences, Örebro University, Örebro, Sweden

[4]Unit of Research, Education, and Development, Östersund Hospital, Department of Public Health and Clinical Medicine, Umeå University, Umeå, Sweden

[5]Department of Psychology and Social Work, Mid Sweden University, Östersund, Sweden

Corresponding Author:
Towe Wadensten, MMSc
Family Medicine, Department of Public Health and Clinical Medicine
Umeå University
Umeå, SE-90187
Sweden
Phone: 46 907855000
Email: towe.wadensten@umu.se

Abstract

Background: Many women experience urgency (UUI) and mixed (MUI) urinary incontinence but commonly hesitate to seek care. Treatment access and self-management for these conditions can be supported through eHealth approaches.

Objective: This study aimed to investigate the efficacy of the mobile app Tät II for self-management of UUI and MUI in women.

Methods: This randomized controlled trial included women ≥18 years old with UUI or MUI and ≥2 leakages per week. Those with red-flag symptoms were excluded. Participants were recruited via analog and digital advertisements and screened for initial selection through a web-based questionnaire. Data were collected using another questionnaire and a 2-day bladder diary. A telephone interview confirmed the symptom diagnosis. Participants were randomized (1:1) to receive access to a treatment app (including pelvic floor muscle training, bladder training, psychoeducation, lifestyle advice, tailored advice, exercise log, reinforcement messages, and reminders) or an information app (control group), with no external treatment guidance provided. The primary outcome was incontinence symptoms at the 15-week follow-up, measured using the International Consultation on Incontinence Questionnaire (ICIQ)−Urinary Incontinence Short Form (ICIQ-UI SF). Urgency symptoms were assessed using the ICIQ−Overactive Bladder Module (ICIQ-OAB) and quality of life using the ICIQ−Lower Urinary Tract Symptoms Quality of Life Module (ICIQ-LUTSqol). Incontinence episode frequency (IEF) was calculated per bladder diary entries. Improvement was measured using the Patient's Global Impression of Improvement. All outcomes were self-reported. Cure was defined as no leakages per the bladder diary. Intention-to-treat analysis was performed.

Results: Between April 2017 and March 2018, 123 women (mean age 58.3, SD 9.6 years) were randomized to the treatment (n=60, 2 lost to follow-up) or information (n=63) group. Of these, 35 (28%) women had UUI, and 88 (72%) had MUI. Mean ICIQ-UI SF score at follow-up was lower in the treatment group than in the information group (estimated difference −3.1, 95% CI −4.8 to −1.3). The estimated between-group difference was −1.8 (95% CI −2.8 to −0.99) for mean ICIQ-OAB score and −6.3 (95% CI −10.5 to −2.1) for the mean ICIQ-LUTSqol score at follow-up. IEF reduction from baseline to follow-up was greater in the treatment group (−10.5, IQR −17.5 to −3.5) than in the information group (P<.001). Improvement was reported by 87% (52/60) of treatment group participants and by 30% (19/63) of information group participants. The cure rate was 32% in the treatment group, and 6% in the information group (odds ratio 5.4, 95% CI 1.9-15.6; P=.002). About 67% (40/60) of the treatment group participants used the app more than thrice a week.

Conclusions: The treatment app was effective for improving urgency and mixed incontinence in women. When self-management is appropriate, this app may be a good alternative to pharmacological treatment or other conservative management, thus increasing access to care.

Trial Registration: ClinicalTrials.gov NCT03097549; https://clinicaltrials.gov/ct2/show/NCT03097549

(J Med Internet Res 2021;23(4):e19439) doi:10.2196/19439

KEYWORDS

eHealth; mHealth; urinary incontinence; urgency urinary incontinence; mixed urinary incontinence; self-management; mobile app; smartphone app; women

Introduction

Urinary incontinence is a common problem that affects many women at some time during their lives, with reported prevalence rates ranging between 25% and 45%, depending on the population and study design [1-4]. There are several types of urinary incontinence. Stress urinary incontinence (SUI) is defined as leakage upon exertion (eg, during coughing or jumping), urgency urinary incontinence (UUI) involves urinary leakage combined with an urge to void, and mixed urinary incontinence (MUI) manifests as a combination of SUI and UUI symptoms [5]. Prevalence rates vary from 1% to 11% for UUI and from 2% to 36% for MUI [1,4]. Overactive bladder is a broad term that includes UUI, and it is defined as the experience of a compelling urgency to void, often combined with more frequent voiding, and sometimes nocturia [5]. These conditions can lead to a sense of shame, social isolation, and lower self-esteem—with a significant impact on health-related quality of life [6-8].

For the three main types of urinary incontinence, the recommended first-line treatment includes pelvic floor muscle training (PFMT) and, where appropriate, lifestyle changes (eg, reduced caffeine intake, modified fluid intake, and weight reduction if overweight) [9,10]. Unsupervised PFMT has been recommended in cases wherein an underlying pathology is absent [9]. For women with urgency-predominant urinary incontinence and small micturition volumes, bladder training is recommended, with scheduled voiding or prolonged voiding intervals [3,9,10]. According to a recent review, PFMT might also be useful in overactive bladder treatment, but more evidence is needed in this regard [11]. As a second line of treatment, pharmacological therapy is recommended and widely used, but it often exhibits only modest effectiveness and commonly leads to side effects [3,10].

Although effective treatments for urinary incontinence are available, they do not cater to all individuals who may benefit from them [9,12]. Some studies describe patients' reluctance to seek help for urinary incontinence, sometimes explained by a sense of embarrassment or mistrust in health care [6,13]. Self-management or treatment options that do not require face-to-face contact might be suitable ways to provide care in some of these cases. Web-based platforms and smartphone apps represent an increasingly common way of supporting self-management or providing treatment for various conditions [14-16]. In the context of the current COVID-19 pandemic, the interest in these kinds of technical solutions has increased even further, and urology is one such field where technology-supported treatment or self-management might be useful [17,18]. However, among the currently available treatment apps related to urinary incontinence, only few have been evaluated with regard to their efficacy [19].

As part of the current research project, we examined the effects of an internet-based treatment program and a smartphone app designed for women with SUI. Both programs were found to be effective treatment options with regard to short-term and long-term improvement of clinically relevant symptoms as well as cost-effectiveness [20-24]. However, evidence regarding app-based treatment for women with UUI or MUI remains scarce. Since urgency-predominant urinary incontinence may be associated with an underlying disease, physical examination is recommended before treatment. An algorithm comprising structured questions combined with dipstick urinalysis has been found to be useful in identifying women who may benefit from management in ways other than the usual care provided [25]. Along with other innovative options for providing non−face-to-face diagnosis and treatment for UUI and MUI, this approach might facilitate patients to seek help regarding urinary incontinence and enable increased access to treatment.

We have developed a new smartphone app featuring a complex, individually tailored treatment program designed to help patients self-manage UUI and MUI. In this study, we aimed to evaluate whether this app was effective for improvement and cure of UUI and MUI in women.

Methods

Study Design and Participants

This 1:1 randomized, controlled, parallel-arm trial was performed in Sweden between April 2017 and September 2018. Community-dwelling adult women were recruited via information broadcasted on TV, radio, and newspapers in Sweden, and via targeted Facebook advertisements. The inclusion criteria were as follows: female sex, age ≥18 years, experiencing UUI or MUI with ≥2 leakages/week and symptoms lasting for ≥12 months, access to a smartphone (at least iOS 8.0 or Android 4.0.3), and the ability to send and receive email. The exclusion criteria were as follows: pregnancy, use of another PFMT app, use of mirabegron or antimuscarinic drugs, and incontinence surgery within the last 5 years. Additional exclusion criteria related to red-flag symptoms and certain medical conditions were also considered, namely, painful urgency; previous pyelonephritis; ≥3 urinary tract infections in the last 12 months; dysuria (burning sensation when voiding); visible hematuria; noninvestigated bladder-emptying difficulties;

metrorrhagia; cancer in the pelvic area, bladder, or bowels; decreased mobility or sensitivity in the legs or pelvic area; history of stroke; neurological disease; or diabetes.

Initial selection was performed using a web-based screening questionnaire that included questions regarding education level, postal code, and inclusion and exclusion criteria, which was available on the eContinence project website. To distinguish sex from gender, the question "Are you a woman?" was followed by the question "Were you assigned female sex at birth?" To identify red flags, an algorithm containing structured questions about the presence of relevant symptoms was integrated into the questionnaire (Multimedia Appendix 1). This algorithm was developed through several workshops with researchers and clinicians, and it was based on the best available evidence and clinical experience. Respondents presenting any red flags or other exclusion criteria were not allowed to proceed with the questionnaire, and they were automatically recommended to seek usual care (ie, contact their ordinary health care provider). After completing the full questionnaire and submitting their email address, eligible respondents received an email with an informed consent form and a printable 2-day bladder diary. Respondents with a maximum voided volume of ≤150 mL were deemed ineligible and were contacted by a physician (ES) who redirected them to their usual health care provider as a precautionary measure. The remaining respondents completed a web-based inclusion questionnaire comprising items regarding background information, medical history and lifestyle, more detailed symptom questions, as well as forms related to the outcome measures. In all web-based questionnaires, respondents were required to answer each question in order to proceed to the next. Nonrespondents who did not submit their informed consent and bladder diary, or those who did not answer the inclusion questionnaire, were sent two reminders via email.

Finally, each respondent was contacted via telephone by a specialist incontinence nurse or general practitioner (ES). A telephone interview was conducted with the objectives of confirming the symptom diagnosis (ie, UUI or MUI), reconfirming the absence of exclusion criteria, and ensuring that the participant was fully informed about the study.

This study was approved by the regional ethical review board of Umeå, Sweden (registry number 2016/523-31) and registered at Clinicaltrials.gov (NCT03097549). Before and during the study, and after completion, on-site monitoring was conducted by an independent monitor. The monitor ensured study performance according to the protocol, and the collection, documentation, and reporting of data following good clinical practice and applicable ethical and regulatory requirements.

Randomization and Blinding

The participants included in the study were randomized to one of two study groups: the treatment group or the information group. An independent administrator generated the allocation sequence and prepared 130 numbered, opaque, sealed envelopes, with assignments equally distributed between the two study groups. The study coordinator opened one envelope for each participant, in the order in which they received an email from the interviewer indicating that they were ready for randomization. The participants were not blinded to their allocation. Each participant received an email informing them of their assigned group and providing instructions on how to access the relevant app. Participants randomized to the information group were notified that they would gain access to the intervention once of their follow-up data for the trial was complete.

Intervention and Procedures

The Tät II mobile app was designed for both Android and iOS devices. The contents of the app were developed based on research and clinical experience and were discussed in 2015-2016 with a multi-professional group comprising researchers and clinicians with expertise in family medicine, urogynecology, urology, specialized incontinence care, and psychology. The app was developed during 2016-2018 by ES and TW, in collaboration with other researchers involved in the project and the technical development division at Umeå University. The development process also incorporated user feedback—both from users of the previous app developed within this research project and from a test group of women outside the medical professions. The Tät II app is focused on four themes: PFMT, bladder training, psychoeducation, and lifestyle advice. It also contains automatic reinforcement messages and an exercise log. Tailored advice, based on information from the user's bladder diary and responses to the inclusion questionnaire, was designed to guide the user to the features of the app that would be most relevant to her symptoms and lifestyle (eg, bladder training was recommended if the user had small micturition portions, or weight reduction was recommended if the user was overweight). The different components of the Tät II app are detailed below (Figure 1, Table 1, and Multimedia Appendices 2 and 3). The PFMT treatment program in the app has been previously described and evaluated as part of a smartphone app developed earlier [20].

Figure 1. Screenshots from the treatment app (Tät II). Upper-left corner: main (home) screen; upper-right corner: active view of an exercise in the bladder training program; lower-left corner: textual description of another bladder training exercise; lower-right corner: information from the lifestyle section. Text has been translated from Swedish to English for illustration purposes.

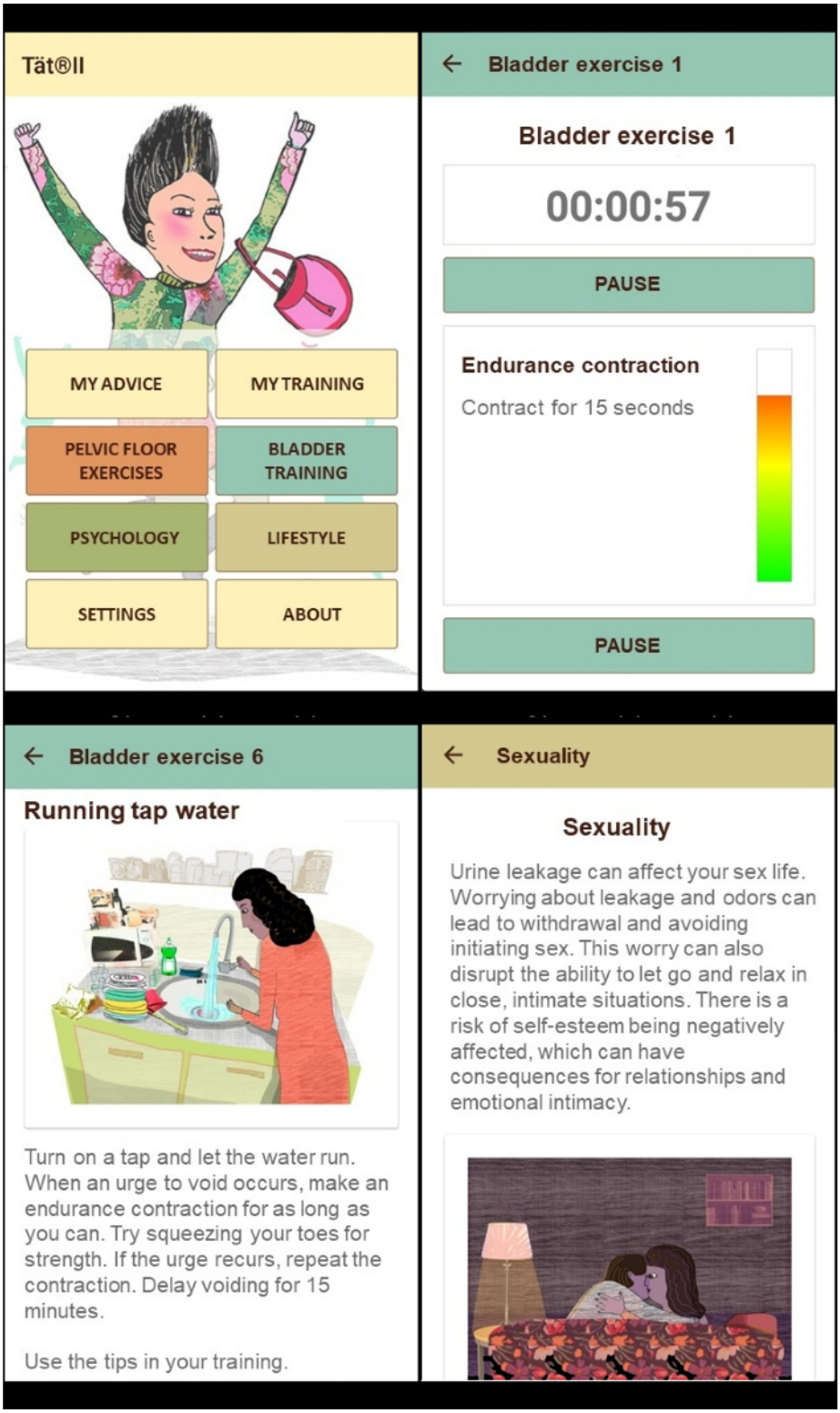


Table 1. Contents of the Tät II treatment app and information app.

Topic	Treatment app	Information app
Pelvic floor muscle training	• Extensive information on anatomy and pelvic floor muscle training • A pelvic floor muscle training program in 11 steps	• Very brief information on pelvic floor muscle training
Bladder training	• Extensive information on bladder physiology and bladder training • A bladder training program in 7 steps[a]	• Very brief information on bladder training
Psychological education	• Extensive information on psychological topics related to urgency symptoms • Three tasks based on cognitive behavioral therapy theory	• Very brief information on psychological topics related to urgency symptoms
Lifestyle advice	• Information on topics of estrogen, fluid intake, physical activity, overweight, smoking, sexuality, constipation, foods, and incontinence aids	• Summarized information on lifestyle advice
Reinforcement	• Recurring questions on degree of bother[b] • Automatic reinforcement messages based on progress • Three customizable notifications for daily reminders	N/A[c]
Other functions	• Exercise log • Tailored advice on what areas to focus on in the app, based on information from bladder diary and questionnaire[d] • Optional 4-digit password protection	N/A

[a]The bladder training program featured exercises for enduring urgency to achieve longer voiding intervals. It did not feature scheduled voiding.

[b]In-app questions about the degree of bother from leakage, urgency, and worry about leaking. The user is asked these questions directly after installing and activating the treatment app, and after 4, 8, and 12 weeks of intervention.

[c]Not applicable.

[d]User is provided 4-10 advices, as relevant, on the following topics: pelvic floor muscle training, bladder training, fluid intake, psychoeducation, local estrogen treatment, obesity, smoking, and constipation.

Activation of the treatment app enabled complete access to all components of the Tät II app. For activation, participants randomized to the treatment group were registered in a database stored on a secure server. A unique one-time activation code was generated and used as a password along with the user ID. The tailored advice was automatically downloaded from the database into the app during the activation process. If the app was not activated within 2 weeks, an email reminder was sent to the participant, and if it was not activated within another week, the participant was contacted via telephone and offered technical guidance. Apart from this, the participants received no guidance from the researchers during the study.

The information app is a limited version of Tät II, which is freely downloadable from app stores. It includes a short summary of lifestyle advice and brief information about the various app components (Table 1).

At the 15-week follow-up, participants were asked to complete a web-based questionnaire and a new 2-day bladder diary. The follow-up questionnaire also included the possibility to add qualitative user feedback. After collection of the follow-up data for the trial, participants in the information group received access to the full treatment app, and participants in the treatment group received information on maintenance training. The different data collection time points of the trial are detailed in Multimedia Appendix 4.

No data were transmitted from the app apart from the voluntary submission of user statistics at follow-up. Participants were encouraged to report any potential side-effects to the research team via email or telephone. Participants were also instructed to seek usual care if any red-flag symptoms appeared.

Outcomes

The primary outcome was the between-group difference in incontinence symptom severity at follow-up, as measured using the Swedish version of a validated questionnaire: the International Consultation on Incontinence Questionnaire−Urinary Incontinence Short Form (ICIQ-UI SF) [26]. The ICIQ-UI SF includes 3 questions about the frequency and amount of urinary leakage and its effect on everyday life. The responses are summed to obtain a total score ranging from 0 to 21 points. The severity of incontinence symptoms was categorized as slight (1-5 points), moderate (6-12 points), severe (13-18 points), or very severe (19-21 points) [27].

Secondary outcomes included urgency symptoms, quality of life, and catastrophizing. The International Consultation on Incontinence Questionnaire−Overactive Bladder Module (ICIQ-OAB) includes 4 items on the frequency of day and night

micturition, urgency, and urgency leakage, and the responses are summed to obtain a total score ranging from 0 to 16 points [28]. The International Consultation on Incontinence Questionnaire−Lower Urinary Tract Symptoms Quality of Life Module (ICIQ-LUTSqol) includes 19 items regarding the impact of urinary leakage on the quality of life, and the responses are summed to obtain an overall score ranging from 19 to 76 points [28]. We also used a nonvalidated score, the Incontinence Catastrophizing (IC) Scale, which was adapted from a short version of the validated Pain Catastrophizing Scale [29]. This scale was translated to Swedish by the research group by using a structured procedure. The IC Scale comprises 7 items regarding fear of leakage and urgency, and the responses are summed to obtain a total score ranging from 0 to 21 points. For all the above-mentioned scores, a reduction in the score indicates an improvement of the symptoms.

Other secondary outcomes included the number of leakages, use of incontinence aids, impression of improvement, and patient satisfaction. Incontinence episode frequency (IEF) was calculated as the number of leakages reported in a 2-day bladder diary multiplied by 3.5 to generate the weekly number of incontinence episodes. Participants were asked about their use of incontinence aids over the last 4 weeks, and they were provided 6 response options ranging from “No, never” to “Yes, more than 1 pad per day.” The Patient Global Impression of Improvement (PGI-I) is a validated questionnaire evaluating improvement, which was only used at follow-up. Participants rated their follow-up condition as compared with their pretreatment condition, using a 7-item scale with answer options ranging from “Very much better” to “Very much worse” [30]. Patient satisfaction was evaluated only in the treatment group at follow-up. This item asked whether the current treatment was perceived as sufficient, with 3 response options regarding satisfaction and intention to seek further care.

We used information from the bladder diary and follow-up questionnaire to assess cure and improvement. *Cure* was defined as no leakage episodes recorded in the bladder diary at follow-up, and *improvement* was defined as any improvement on the PGI-I.

Performance and Adherence

At both the baseline and follow-up, participants were asked whether they perceived themselves as able to correctly perform pelvic floor contractions. At the follow-up, they were also asked to appraise their current ability to contract their pelvic floor muscles as compared with before they had access to the assigned treatment or information app, with responses ranging from “Much better” to “Much worse,” and to appraise the ability to resist urgency through a corresponding question, with similar response options.

Furthermore, the follow-up questionnaire included a question on how often the participants had used their assigned app during the study period. Response options ranged from “Never” to “Daily, three times a day, or more often.” There was also a question on whether the participant had used another incontinence app or participated in another incontinence treatment program during the study period.

Technical Issues and User Feedback

Participants were informed that they could contact the researchers via email in case of technical problems with the app. The follow-up questionnaire allowed participants to provide qualitative feedback via nonmandatory open-ended questions about how they perceived the assigned app, in general, and the specific contents of the app. Participants in the treatment app group were also asked to rank the 6 treatment app components (tailored advice, exercise log, PFMT, bladder training, psychoeducation, and lifestyle advice) from most useful to least useful.

Sample Size

The expected response was based on results from our previous smartphone study, and the findings of Albers-Heitner et al [20,31]. We anticipated ICIQ-UI SF improvements of 2.5 points in the treatment group and 0.9 points in the information group. This level of change has also been found to reflect a clinically important difference in women with stress urinary incontinence after treatment via eHealth [22]. Detecting this difference with 80% power, a two-sided test, and a significance level of $P<.05$ would require a sample size of 49 in each group. With an expected drop-out rate of 20%, we needed approximately 60 participants in each group. Thus, we aimed to recruit 120 women for this study.

Statistical Analysis

For all outcome measures, we performed an intention-to-treat analysis. To analyze the ICIQ-UI SF, ICIQ-LUTSqol, ICIQ-OAB, and IC scale, we used analytical methods that accounted for all available data. We used the last observation carried forward method for IEF and incontinence aid usage, and we applied imputation of values corresponding to no change in the PGI-I.

Baseline data were described in terms of age, BMI, educational level, medication use, and all primary and secondary outcomes that were measured at the baseline. For the primary outcome, we examined the between-group difference in the mean ICIQ-UI SF score at the 15-week follow-up using a linear mixed-model analysis incorporating baseline data. For the secondary outcomes, between-group comparisons were made using a linear mixed-model analysis for the difference in mean values for continuous variables, and the Mann-Whitney *U* test for the distribution of categorical variables and for the difference in median for nonnormally distributed continuous variables. For within-group comparisons (ie, between baseline and follow-up), we used a paired *t* test for continuous variables and a Wilcoxon signed-rank test for nonnormally distributed continuous variables. For IEF, we calculated the difference between baseline and follow-up scores for each individual, presented as median and IQR values. Since IEF data were not normally distributed, they were analyzed using the Wilcoxon signed-rank test. We used a chi-square test with continuity correction to calculate the odds ratios (ORs) for cure and improvement.

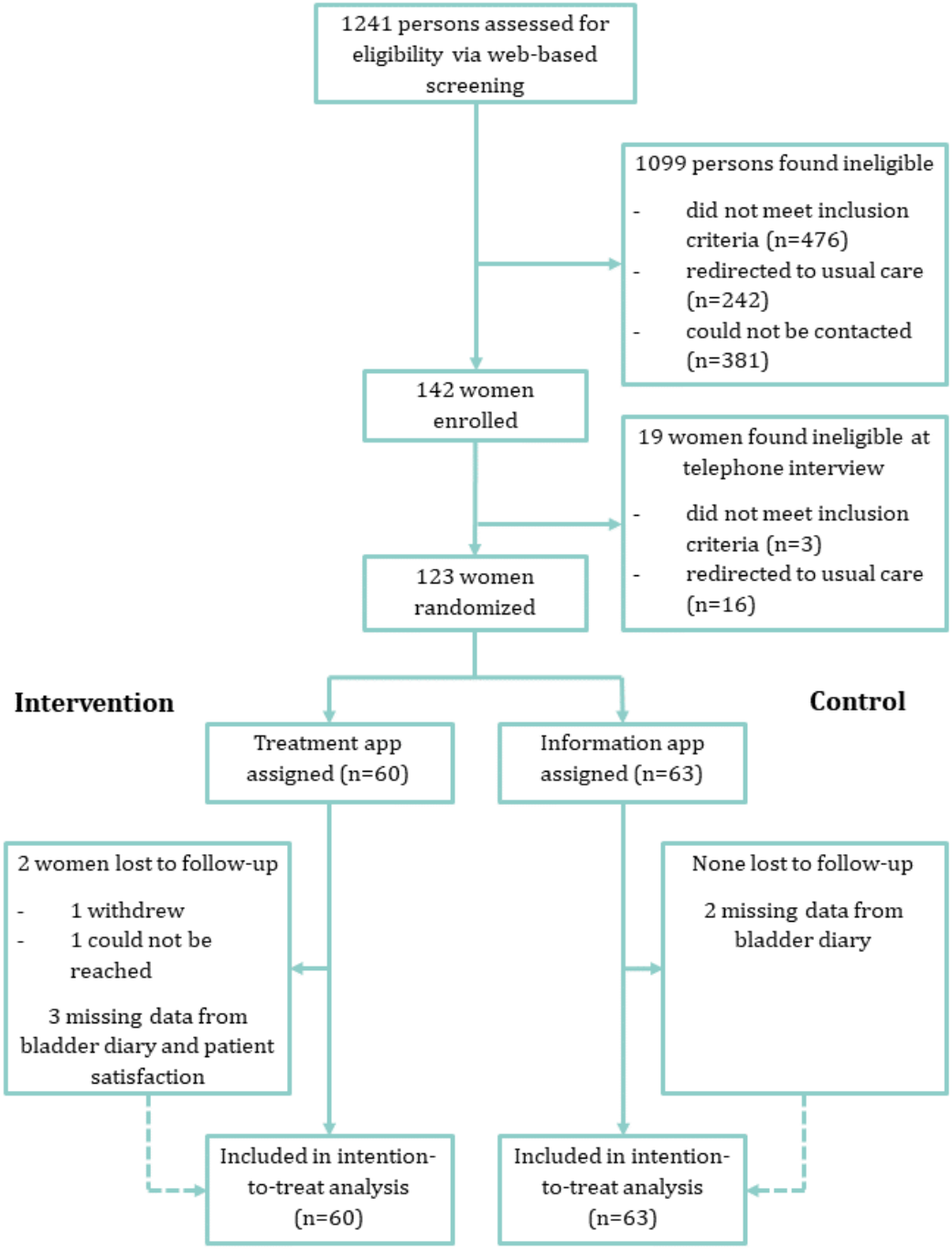
All statistical analyses were performed using SPSS (version 25; IBM Corp).

Results

Study Flow and Participant Characteristics

The web-based screening questionnaire was initiated by 1241 individuals, of whom 1099 were ineligible. A total of 142 women were interviewed, of which 123 were included in the study and randomized to receive either the treatment app (n=60) or the information app (n=63). Two women (both in the treatment group) were lost to follow-up. Five women submitted incomplete follow-up data regarding two secondary outcomes (Figure 2). The median time from randomization to follow-up was 16.1 (IQR 15.0-18.1) weeks in the treatment group and 15.5 (IQR 14.6-17.6) weeks in the information group (*P*=.14).

Figure 2. Trial profile.



Baseline characteristics did not differ between the two groups. The mean age of women included was 58 (range 31-77) years. Half of the women (64/123, 52%) were overweight or obese. The majority of participants had received university-level education, and 95% (117/123) of them classified themselves as being quite or very knowledgeable in using computers or tablets.

Three-quarters of the participants (88/123, 72%) were diagnosed with MUI, and 69% (85/123) had not previously sought care for their incontinence (Table 2). Severe or very severe incontinence was reported by 38% (47/123) of the participants, and the mean symptom severity score was 11.6 (SD 3.3).

Table 2. Baseline characteristics of study participants.

Characteristic	Treatment group (n=60)	Information group (n=63)
General information		
Age in years, mean (SD)	58.9 (9.2)	57.7 (9.9)
BMI in kg/m^2, mean (SD)	26.5 (3.6)	25.9 (5.2)
University education ≥3 years, n (%)	44 (73.3)	35 (55.6)
Self-rated knowledge in using computers or tablets (1-4 points[a]), mean (SD)	3.5 (0.6)	3.4 (0.6)
eHEALS[b] score, mean (SD)	32.9 (5.6)	33.8 (4.5)
Lifestyle		
Physical activity >3 hours/week, n (%)	33 (55.0)	40 (63.5)
Smoker[c], n (%)	1 (1.7)	2 (3.2)
Coffee consumption ≥5 cups/day, n (%)	7 (11.7)	10 (15.9)
Tea consumption ≥5 cups/day, n (%)	4 (6.7)	2 (3.2)
Often or always constipated, n (%)	6 (10)	3 (4.8)
Gynecology		
Parity, n (%)		
0	8 (13.3)	7 (11.1)
1	4 (6.7)	7 (11.1)
≥2	48 (80)	49 (77.8)
Postmenopausal >1 year, n (%)	42 (70)	47 (74.6)
Current estrogen usage, n (%)		
Local estrogen	11 (18.3)	13 (20.6)
Systemic estrogen	1 (1.7)	1 (1.6)
Gynecological surgery, n (%)		
Hysterectomy	4 (6.7)	7 (11.1)
Prolapse surgery	3 (5)	0 (0)
Incontinence surgery[d]	3 (5)	3 (4.8)
Urinary incontinence		
Symptom diagnosis, n (%)		
Mixed urinary incontinence	45 (75)	43 (68.3)
Urgency urinary incontinence	15 (25)	20 (31.7)
Duration of symptoms >5 years, n (%)	37 (61.7)	38 (60.3)
Previous health care contact for incontinence symptoms, n (%)	19 (31.7)	19 (30.2)
ICIQ-UI SF[e] score, mean (SD)	11.7 (3.5)	11.4 (3.2)
IEF[f] per week, mean (SD)[g]	21.8 (16.8)	21.1 (13.7)

[a]A higher score indicates higher self-perceived knowledge.

[b]eHEALS: eHealth Literacy Scale, a self-reported 8-item scale assessing an individual's ability to identify, evaluate, and use eHealth resources.

[c]No daily smokers, only weekly smokers, participated in the study.

[d]Participants who had undergone incontinence surgery in the last 5 years were not included in the study.

[e]ICIQ-UI SF: International Consultation on Incontinence Questionnaire−Urinary Incontinence Short Form.

[f]IEF: incontinence episode frequency.

[g]Mean (SD) values presented to facilitate comparison with other populations.

Outcomes

At the 15-week follow-up, women in the treatment group had significantly lower incontinence symptom scores than those in the information group. The estimated between-group difference in mean in the primary outcome, the ICIQ-UI SF score, was −3.1 (95% CI −4.8 to −1.3). Both groups showed an improvement from the baseline but a larger improvement was noted in the treatment group (Table 3).

Table 3. Continuous outcomes compared between the treatment group (n=60) and information group (n=63) at follow-up.

Outcome measure and group allocation	Baseline, mean (SD)	Follow-up, mean (SD)	Between-group comparison at follow-up	
			Estimated difference (95% CI)[a]	*P* value
Primary outcome				
ICIQ-UI SF[b]			−3.1 (−4.8 to −1.3)	.001
Treatment group (n=60)	11.7 (3.5)	7.0 (3.7)[c]		
Information group (n=63)	11.4 (3.2)	9.8 (3.5)		
Secondary outcomes				
ICIQ-OAB[d]			−1.8 (−2.8 to −0.9)	<.001
Treatment group (n=60)	6.8 (1.8)	4.7 (2.0) [c]		
Information group (n=63)	6.7 (1.8)	6.4 (2.0)		
ICIQ-LUTSqol[e,f]			−6.3 (−10.5 to −2.1)	.004
Treatment group (n=60)	37.6 (8.3)	29.8 (7.8)[c]		
Information group (n=63)	38.0 (8.1)	36.5 (9.0)		
Incontinence Catastrophizing Scale			−1.6 (−2.8 to −0.3)	.016
Treatment group (n=60)	4.4 (2.8)	2.3 (2.1)[c]		
Information group (n=63)	4.7 (2.5)	4.1 (2.5)		

[a]Comparison of mean scores using a linear mixed model.

[b]ICIQ-UI SF: International Consultation on Incontinence Questionnaire−Urinary Incontinence Short Form.

[c]Mean values based on the scores of the 58 treatment app users who completed the follow-up questionnaire.

[d]ICIQ-OAB: International Consultation on Incontinence Questionnaire−Overactive Bladder Module.

[e]ICIQ-LUTSqol: International Consultation on Incontinence Questionnaire−Lower Urinary Tract Symptoms Quality of Life Module.

[f]Three of the items in the ICIQ-LUTSqol included an additional response option, “Not applicable” (these questions concerned partner relations, sex life, and family life). For this study, we set this response option as equal to 1 point, corresponding to the response option “Not at all” (ie, no impact).

Compared with those in the information group, participants in the treatment group also showed significantly greater improvements in the secondary outcomes, with lower scores for urgency symptoms, condition-specific quality of life, and catastrophizing at follow-up. Within-group comparisons revealed statistically significant improvements from baseline to follow-up in all outcomes, except urgency symptoms in the information group (Table 3 and Multimedia Appendix 5).

Participants in both the treatment and information groups exhibited a significant reduction in the number of incontinence episodes (IEF) from the baseline to follow-up. This improvement was greater in the treatment group than in the information group (Table 4). The number of incontinence episodes was reduced by 50% or more in 68% (41/60) of the participants in the treatment group and 30% (19/63) of the participants in the information group. Incontinence aids were used less than once a week at follow-up by the majority of participants (38/60, 63%) in the treatment group compared with those (25/60, 40%) in the information group (Table 5).

Table 4. Differences in incontinence episode frequency from the baseline to follow-up compared between the treatment group (n=60) and the information group (n=63).

Group allocation	Baseline, median (IQR)	Follow-up, median (IQR)	Within-group comparison[a]		Between-group comparison at follow-up[b]
			Difference[c], median (IQR)	*P* value	*P* value
Treatment app	17.5 (10.5 to 27.1)	3.5 (0.0 to 10.5)	−10.5 (−17.5 to −3.5)	<.001	<.001
Information app	21.0 (7.0 to 31.5)	10.5 (7.0 to 21.0)	−3.5 (−14.0 to 3.5)	.003	

[a]For within-group comparisons, we calculated the difference from baseline to follow-up for each individual and performed analyses using a Wilcoxon signed-rank test.

[b]Mann-Whitney *U* test.

[c]5 participants in the treatment group and 2 participants in the information group had a missing value at follow-up, and for those, the difference was set to 0 (ie, no change).

Table 5. Incontinence aid usage by participants in the treatment group (n=60) and information group (n=63), reported at baseline and at follow-up.

Allocation and incontinence aid usage	Participants, n (%)		*P* value	
	Baseline	Follow-up	Within-group comparison[a]	Between-group comparison at follow-up[b]
Treatment app			<.001	.01
Never	11 (18.3)	26 (43.3)[c]		
<1/week	10 (16.7)	12 (20)		
1-3/week	12 (20)	4 (6.7)		
>3/week, not daily	4 (6.7)	2 (3.3)		
1/day	12 (20.0)	10 (16.7)		
>1/day	11 (18.3)	6 (10)[c]		
Information app			.15	
Never	14 (22.2)	14 (22.2)		
<1/week	11 (17.5)	11 (17.5)		
1-3/week	7 (11.1)	8 (12.7)		
>3/week, not daily	4 (6.3)	6 (9.5)		
1/day	11 (17.5)	13 (20.6)		
>1/day	16 (25.4)	11 (17.5)		

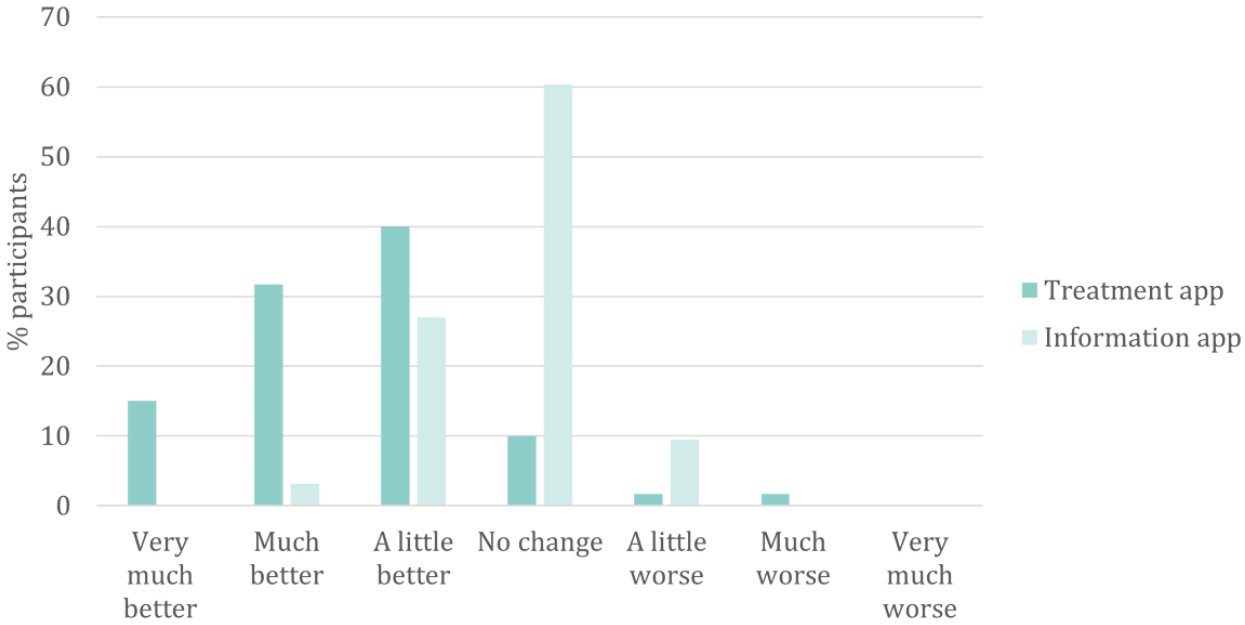
[a]Wilcoxon signed-rank test.

[b]Mann-Whitney *U* test.

[c]Imputed baseline value for 1 participant lost to follow-up.

PGI-I scores indicated that 87% (52/60) of the participants in the treatment group reported an improvement as compared with 30% (19/63) in the information group. Figure 3 details the distribution of the responses. Cure was reported by 32% (19/60) of the participants in the treatment group and by 8% (5/63) of the participants in the information group. The OR for cure was 5.4 (95% CI 1.9-15.6, *P*=.002) for the treatment group with the information group as reference. Moreover, the OR for improvement was 15.1 (95% CI 6.0-37.7, *P*<.001) for the treatment group with the information group as reference.

Figure 3. Patient's Global Impression of Improvement responses reported by the participants at follow-up. Comparison between the treatment group (n=60) and the information group (n=63). *P<.001 (Mann-Whitney U test). Two participants were lost to follow-up in the treatment group and imputed as "no change" for this analysis.



In the treatment group, 7% (4/60) of the participants reported their current satisfaction with the treatment and that they were completely free from urinary leakage and urgency symptoms, whereas 52% (31/60) of the women reported their satisfaction with the treatment despite some remaining symptoms. The remaining 33% (20/60) of the participants reported that they were not satisfied with the treatment, but only 7 of these 20 considered seeking additional care.

Performance and Adherence

At baseline, 40% (49/123) of the participants were confident that they correctly performed pelvic floor contractions, with no between-group difference. At follow-up, this rate was increased to 62% (37/60) of the participants in the treatment group, and 52% (33/63) of the participants in the information group (P=.39). At follow-up, 85% (51/60) of the participants in the treatment group stated that their ability to contract their pelvic floor muscles was slightly or much improved, compared with 16% (10/63) of the participants in the information group (P<.001). Similarly, the ability to resist an urge to void was slightly or much improved for 80% (48/60) of the participants in the treatment group, compared with 27% (17/63) of the participants in the information group (P<.001).

During the 15-week treatment period, 40 (67%) of the 60 participants in the treatment group used the app more than three times per week, and 6 (10%) participants used it at least three times per day. None of the treatment group participants used other incontinence apps or treatment programs during the study period. In the information group, 2 of the 63 (3%) participants used another PFMT app, and 1 (2%) participant practiced PFMT and tried to resist urgency. Moreover, 1 (2%) participant in the information group sought help from usual care for incontinence-related symptoms during the study period and received incontinence aids, advice on PFMT, and treatment with intravesical hyaluronic acid-chondroitin sulphate. The adherence to the tailored advice is described in Multimedia Appendix 3.

Technical Issues and User Feedback

No participants reported technical problems with the apps, and no privacy breaches occurred. Many participants in the information group thought that the information was too brief and, therefore, they rarely used the app. Most participants in the treatment group were satisfied with their experience of the app in terms of contents and usability. The users ranked the various treatment app components, with the PFMT portion deemed most useful (ranked 1, 2, or 3 by 95% of all users), followed by the exercise log (ranked 1, 2, or 3 by 80% of all users) and tailored advice (ranked 1, 2, or 3 by 66% of all users).

Adverse Events

Two participants (both in the treatment group) reported potential adverse events: one of them reported the development of an inguinal hernia during the treatment period, which later required surgery. Discussion of this case with the specialists in the research group, an independent specialist in hernia surgery, and an official at Medical Products Agency, Sweden, led to the conclusion that the hernia was likely not related with the use of the treatment app. The other participant reported altered incontinence symptoms, with decreased urgency but increased episodes of spontaneous urinary leakage.

Discussion

Principal Results

In the present randomized controlled trial among women with mixed and urgency urinary incontinence, we found that the treatment app was effective in reducing incontinence symptoms (our primary outcome), with lower scores observed at the 15-week follow-up in the treatment group compared with those in the information group. We also found that treatment app users

showed greater improvements regarding quality of life, urgency symptoms, number of incontinence episodes, use of incontinence aids, and catastrophizing, compared with the information app users.

Clinical Relevance

About 87% (52/60) of the treatment app users reported an improvement, and half of them (28/60, 47%) reported *much* or *very much* improvement. A 50% reduction in the frequency of incontinence episodes is considered clinically relevant, and this degree of reduction was observed for the majority of the treatment app users in our study [32]. For women with SUI, reductions of 2.5 points in the ICIQ-UI SF score and 3.7 points in the ICIQ-LUTSqol score has been considered to reflect clinically relevant improvement after pelvic floor muscle training via an eHealth approach [22]. The within-group reductions observed in our study were larger than these thresholds and also larger than the 4- and 6-point reduction in the ICIQ-UI SF and ICIQ-LUTSqol scores established by Lim et al as minimum important differences for other conservative management [33]. To our knowledge, there are no similar studies of women with UUI and MUI, but minimum important differences are likely to be at a similar level compared with the studies mentioned above. The PGI-I and comparison with minimum important differences indicate that the changes in symptoms and quality of life for the treatment group observed in our study are clinically relevant.

Strengths and Limitations

One of the strengths of our study was that we used clinically relevant outcomes that were carefully selected to cover different aspects of UUI and MUI, including symptoms, quantification, quality of life, and subjective improvement. The three ICIQ scores, including the one used as the primary outcome (ICIQ-UI SF), are all Grade A or A+ recommended outcome measures according to the International Continence Society [3]. These questionnaires are also validated for electronic use [34,35]. Additionally, our study was adequately sized and conducted with external monitoring, as a mark of quality. There were very small losses to follow-up and no internal losses, except for the secondary outcome of IEF. We developed and utilized an algorithm to identify suitable participants, and all participants experienced extensive incontinence symptoms at the baseline, strengthening the need for treatment. Another strength of our study was that the app was thoroughly designed and stable and required no updates during the study period. No users experienced technical issues. The treatment app included a patient-centered design, such that the user herself decided which parts of the app to use. Tailored advice, based on information about the user's lifestyle and incontinence symptoms, offered guidance on what might be the most beneficial component or feature of the app for her to focus on. All participants in the treatment group downloaded and activated the app, and most of them regularly used the app. The risk for contamination between the two groups was negligible since activation of the treatment app required a unique one-time authorization code. The treatment app featured information and exercises covering multiple topics related to UUI and MUI, resembling a clinical reality with a multi-faceted intervention. Therefore, this study cannot discriminate how different parts of the app contributed to the various effects.

A potential limitation of this study is that we cannot yet assess the long-term effects of the app. However, we previously reported that our smartphone app targeted at women with SUI had a long-lasting effect at the 2-year follow-up [21]. Similar to most other investigations of eHealth or behavioral therapy interventions, another limitation of our study was that the participants could not be blinded to their group allocation. Our choice to not use care-as-usual as a control group could also be viewed as a limitation. However, seeking care-as-usual was not a likely option for most of our intended target population; thus, we argue that the information app was the most comparable control. Another potential concern is the lack of face-to-face contact with a health care provider and that there was no professional assessment of the participants' ability to perform correct PFMT contractions. However, in a JAMA review from 2017, unsupervised PFMT is recommended as a first-line treatment after exclusion of serious underlying pathologies [9]. The treatment app included information on how to correctly perform contractions, and it recommended women who were uncertain of their contraction technique to contact their ordinary health care provider for advice. At follow-up, no treatment app users had sought help from usual care, and the majority stated that their ability to perform PFMT contractions had improved. Most of the women in the treatment group were satisfied with the treatment, and of those who were not satisfied, only a few intended to seek care elsewhere. However, since the qualitative feedback was optional, we do not know the reason behind their decision not to seek further care despite not being satisfied. Furthermore, this study focused on the presentation of quantitative data, and while this is a strength when investigating efficacy, a deeper analysis of the qualitative data collected might provide valuable information about the experiences of the participants in the trial. A common issue in research, and particularly in studies of eHealth interventions, is that the participants' education level is often higher than that of the general population. The average education level of people in Sweden is higher than that of people in many other countries; nonetheless, women with a university-level education were over-represented among our participants, which may potentially affect the generalizability of our results.

Comparison With Prior Work

Several systematic reviews report that antimuscarinic drugs and mirabegron yield a mean reduction corresponding to half a leakage per 24 hours at the 3-month follow-up when compared with the placebo [36-39]. The median reduction for the treatment group in our study was twice as large as this. Moreover, antimuscarinic medication is commonly associated with side-effects, such as dry mouth and constipation. Mirabegron is better tolerated but has side-effects, such as urinary tract infections, irregular heartrate, and palpitations [3,36,37]. In contrast, behavioral treatment and lifestyle advice carry no known side-effects. Women with urgency urinary incontinence are more likely to achieve improvement, cure, and satisfaction with behavioral therapy than with anticholinergics [40]. The cure and improvement rates in our treatment app group were in line with the findings from other studies. A systematic review

update from 2018 reported a 25%-30% cure rate of urgency urinary incontinence with neuromodulation, behavioral therapy, or combined anticholinergic and behavioral therapy. The OR for cure of urgency incontinence with behavioral therapy was 2.75 (95% CI 1.53-4.92) compared with the placebo, sham, or no treatment [40].

Conclusions and Outlook

UUI and MUI affect many women and can have a potentially large impact on their quality of life. Thus, it is important to offer effective treatment options that can reach many patients, and eHealth methods are a new potential means of supporting self-management. Providing treatment that does not require face-to-face contact with the health care service provider might facilitate increased care-seeking among these women. Additionally, eHealth tools (eg, smartphone apps) provide possibilities for adherence-promotion, such as reminder notifications, and can be tailored for the specific user, such as through the tailored advice provided in our presently tested treatment app. Concerns have recently been raised about a digital divide such that some groups might be less able to use these digital health aids [41-43]. Future research is needed to identify ways to improve the interventions, or the development process, to make eHealth treatment options more accessible or relevant for new user groups. Further investigations are also needed to evaluate the use of algorithms to select patients for self-management of these conditions, with regard to medical safety. For populations similar to the participants in our present study, our results indicate that the treatment app is already an effective treatment option. There remains a need to study the long-term effects, and to decide how to make this app available to patients—that is, whether the app can be offered as a stand-alone, first-line intervention for women with an uncomplicated medical history, or whether it should be regulated and prescribed only by health care professionals.

To our knowledge, this is the first study to demonstrate the potential to provide an effective, tailored, app-based treatment to women with urgency or mixed urinary incontinence suited for self-management. Our results show an efficacy that is comparable to other first-line treatments available. Therefore, we propose that this app could be added to the treatment options offered as part of usual care for women presenting with these conditions.

Acknowledgments

The authors would like to thank the following individuals: Susanne Johansson, for coordinating the study; Agneta Jervhed, for conducting telephone interviews and reminder phone calls to the participants; Johanna Ekdahl, for assisting the development of the psycho-educative component of the Tät II app and compiling the Incontinence Catastrophizing Scale; Karin Stenzelius, for participating in the planning for this study, developing the foundation of the bladder training program, creating some of the lifestyle advice contents available on the app, and participating in the translation and adaptation of some of the questionnaires used; the statisticians Lars Söderström and Hans Stenlund, for their involvement in developing the statistical plan, sample size calculations, and initial analyses; and Lars Malmberg, for participating in the early stages of planning for this study. We also want to thank all the women who participated in the study. Our study was funded by the Kamprad Family Foundation; the Swedish Research Council for Health, Working, and Welfare; and the Region Jämtland Härjedalen. These organizations provided grants but had no role in the study design, data collection, data analysis, data interpretation, or writing of the manuscript.

Authors' Contributions

ES conceptualized the study, was the principal investigator, and led the study at all stages. ES, EN, and TW designed the study with input from KF and EW. The statistical plan and protocol for the study were written by ES with support from TW, EN, and Lars Söderström (statistician). TW and EN performed the statistical analyses of the trial data, and ES and AL participated in the analyses. All authors had full access to all study data and participated in the interpretation of the data and statistical analyses. TW wrote the manuscript draft, with substantial input from ES and EN, and all authors critically reviewed successive drafts of the report and approved the final version of the manuscript before publication.

Conflicts of Interest

The logos Tät and Tät.nu are registered as trademark by The Swedish Patent and Registration office for E Samuelsson at Umeå University. None of the researchers have any financial interest in the programs.

Multimedia Appendix 1
Algorithm with structured questions regarding red-flag symptoms.
[PNG File , 101 KB - jmir_v23i4e19439_app1.png]

Multimedia Appendix 2
Information on the treatment program delivered via the Tät II treatment app.
[DOC File , 773 KB - jmir_v23i4e19439_app2.doc]

Multimedia Appendix 3

Criteria, tailored advice, and participant adherence data for treatment information delivered via the Tät II treatment app.
[DOC File , 46 KB - jmir_v23i4e19439_app3.doc]

Multimedia Appendix 4
Data collection timeline.
[PNG File , 33 KB - jmir_v23i4e19439_app4.png]

Multimedia Appendix 5
Continuous outcomes compared between the treatment group (n=60) and the information group (n=63) at follow-up, and within-group comparisons from baseline to follow-up.
[DOC File , 49 KB - jmir_v23i4e19439_app5.doc]

Multimedia Appendix 6
CONSORT-EHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1047 KB - jmir_v23i4e19439_app6.pdf]

References

1. Hannestad YS, Rortveit G, Sandvik H, Hunskaar S. A community-based epidemiological survey of female urinary incontinence: the Norwegian EPINCONT study. Epidemiology of Incontinence in the County of Nord-Trøndelag. J Clin Epidemiol 2000 Nov;53(11):1150-1157. [Medline: 11106889]
2. Hunskaar S, Burgio K, Diokno A, Herzog AR, Hjälmås K, Lapitan MC. Epidemiology and natural history of urinary incontinence in women. Urology 2003 Oct;62(4 Suppl 1):16-23. [Medline: 14550833]
3. Abrams P, Cardozo L, Wagg A, Wein A, editors. Incontinence 6th Edition. Bristol, UK: ICI-ICS. International Continence Society; 2017.
4. Milsom I, Coyne KS, Nicholson S, Kvasz M, Chen C, Wein AJ. Global prevalence and economic burden of urgency urinary incontinence: a systematic review. Eur Urol 2014 Jan;65(1):79-95. [doi: 10.1016/j.eururo.2013.08.031] [Medline: 24007713]
5. Haylen BT, de Ridder D, Freeman RM, Swift SE, Berghmans B, Lee J, et al. An International Urogynecological Association (IUGA)/International Continence Society (ICS) joint report on the terminology for female pelvic floor dysfunction. Neurourol Urodyn 2010;29(1):4-20. [doi: 10.1002/nau.20798] [Medline: 19941278]
6. Siddiqui NY, Levin PJ, Phadtare A, Pietrobon R, Ammarell N. Perceptions about female urinary incontinence: a systematic review. Int Urogynecol J 2014 Jul;25(7):863-871. [doi: 10.1007/s00192-013-2276-7] [Medline: 24310988]
7. Coyne KS, Sexton CC, Irwin DE, Kopp ZS, Kelleher CJ, Milsom I. The impact of overactive bladder, incontinence and other lower urinary tract symptoms on quality of life, work productivity, sexuality and emotional well-being in men and women: results from the EPIC study. BJU Int 2008 Jun;101(11):1388-1395 [FREE Full text] [doi: 10.1111/j.1464-410X.2008.07601.x] [Medline: 18454794]
8. Coyne KS, Zhou Z, Thompson C, Versi E. The impact on health-related quality of life of stress, urge and mixed urinary incontinence. BJU Int 2003 Nov;92(7):731-735 [FREE Full text] [Medline: 14616456]
9. Lukacz ES, Santiago-Lastra Y, Albo ME, Brubaker L. Urinary Incontinence in Women: A Review. JAMA 2017 Oct 24;318(16):1592-1604. [doi: 10.1001/jama.2017.12137] [Medline: 29067433]
10. National Institute for Health and Care Excellence (NICE). NICE Guidance - Urinary incontinence and pelvic organ prolapse in women: management. BJU Int 2019 Apr 22;123(5):777-803. [doi: 10.1111/bju.14763]
11. Bo K, Fernandes A, Duarte T, Brito L, Ferreira C. Is pelvic floor muscle training effective for symptoms of overactive bladder in women? A systematic review. Physiotherapy 2020 Mar;106:65-76. [doi: 10.1016/j.physio.2019.08.011]
12. Minassian VA, Yan X, Lichtenfeld MJ, Sun H, Stewart WF. The iceberg of health care utilization in women with urinary incontinence. Int Urogynecol J 2012 Aug;23(8):1087-1093 [FREE Full text] [doi: 10.1007/s00192-012-1743-x] [Medline: 22527544]
13. Hägglund D, Wadensten B. Fear of humiliation inhibits women's care-seeking behaviour for long-term urinary incontinence. Scand J Caring Sci 2007 Sep;21(3):305-312. [doi: 10.1111/j.1471-6712.2007.00481.x] [Medline: 17727542]
14. Bernard S, Boucher S, McLean L, Moffet H. Mobile technologies for the conservative self-management of urinary incontinence: a systematic scoping review. Int Urogynecol J 2019 Jul 02:1163-1174. [doi: 10.1007/s00192-019-04012-w] [Medline: 31267139]
15. Brown J, Michie S, Geraghty AWA, Yardley L, Gardner B, Shahab L, et al. Internet-based intervention for smoking cessation (StopAdvisor) in people with low and high socioeconomic status: a randomised controlled trial. Lancet Respir Med 2014 Dec;2(12):997-1006 [FREE Full text] [doi: 10.1016/S2213-2600(14)70195-X] [Medline: 25262458]
16. Massoudi B, Holvast F, Bockting CLH, Burger H, Blanker MH. The effectiveness and cost-effectiveness of e-health interventions for depression and anxiety in primary care: A systematic review and meta-analysis. J Affect Disord 2019 Feb 15;245:728-743. [doi: 10.1016/j.jad.2018.11.050] [Medline: 30447572]

17. Novara G, Checcucci E, Crestani A, Abrate A, Esperto F, Pavan N, et al. Telehealth in Urology: A Systematic Review of the Literature. How Much Can Telemedicine Be Useful During and After the COVID-19 Pandemic? European Urology 2020 Dec;78(6):786-811. [doi: 10.1016/j.eururo.2020.06.025]
18. Grimes CL, Balk EM, Crisp CC, Antosh DD, Murphy M, Halder GE, et al. A guide for urogynecologic patient care utilizing telemedicine during the COVID-19 pandemic: review of existing evidence. Int Urogynecol J 2020 Apr 27;31(6):1063-1089. [doi: 10.1007/s00192-020-04314-4]
19. Sudol NT, Adams-Piper E, Perry R, Lane F, Chen KT. In Search of Mobile Applications for Patients With Pelvic Floor Disorders. Female Pelvic Med Reconstr Surg 2019;25(3):252-256. [doi: 10.1097/SPV.0000000000000527] [Medline: 29219859]
20. Asklund I, Nyström E, Sjöström M, Umefjord G, Stenlund H, Samuelsson E. Mobile app for treatment of stress urinary incontinence: a randomized controlled trial. Neurourol Urodyn 2016 Sep 09:1369-1376. [doi: 10.1002/nau.23116] [Medline: 27611958]
21. Hoffman V, Söderström L, Samuelsson E. Self-management of stress urinary incontinence via a mobile app: two-year follow-up of a randomized controlled trial. Acta Obstet Gynecol Scand 2017 Oct;96(10):1180-1187 [FREE Full text] [doi: 10.1111/aogs.13192] [Medline: 28718223]
22. Nyström E, Sjöström M, Stenlund H, Samuelsson E. ICIQ symptom and quality of life instruments measure clinically relevant improvements in women with stress urinary incontinence. Neurourol Urodyn 2015 Nov;34(8):747-751 [FREE Full text] [doi: 10.1002/nau.22657] [Medline: 25154378]
23. Sjöström M, Lindholm L, Samuelsson E. Mobile App for Treatment of Stress Urinary Incontinence: A Cost-Effectiveness Analysis. J Med Internet Res 2017 May 08;19(5):e154 [FREE Full text] [doi: 10.2196/jmir.7383] [Medline: 28483745]
24. Sjöström M, Umefjord G, Stenlund H, Carlbring P, Andersson G, Samuelsson E. Internet-based treatment of stress urinary incontinence: 1- and 2-year results of a randomized controlled trial with a focus on pelvic floor muscle training. BJU Int 2015 Dec;116(6):955-964 [FREE Full text] [doi: 10.1111/bju.13091] [Medline: 25683075]
25. Huang AJ, Hess R, Arya LA, Richter HE, Subak LL, Bradley CS, et al. Pharmacologic treatment for urgency-predominant urinary incontinence in women diagnosed using a simplified algorithm: a randomized trial. Am J Obstet Gynecol 2012 May;206(5):444.e1-444.11 [FREE Full text] [doi: 10.1016/j.ajog.2012.03.002] [Medline: 22542122]
26. Avery K, Donovan J, Peters TJ, Shaw C, Gotoh M, Abrams P. ICIQ: a brief and robust measure for evaluating the symptoms and impact of urinary incontinence. Neurourol Urodyn 2004;23(4):322-330. [doi: 10.1002/nau.20041] [Medline: 15227649]
27. Klovning A, Avery K, Sandvik H, Hunskaar S. Comparison of two questionnaires for assessing the severity of urinary incontinence: The ICIQ-UI SF versus the incontinence severity index. Neurourol Urodyn 2009;28(5):411-415. [doi: 10.1002/nau.20674] [Medline: 19214996]
28. Abrams P, Avery K, Gardener N, Donovan J, ICIQ Advisory Board. The International Consultation on Incontinence Modular Questionnaire: www.iciq.net. J Urol 2006 Mar;175(3 Pt 1):1063-6; discussion 1066. [doi: 10.1016/S0022-5347(05)00348-4] [Medline: 16469618]
29. Sullivan MJL, Bishop SR, Pivik J. The Pain Catastrophizing Scale: Development and validation. Psychological Assessment 1995;7(4):524-532. [doi: 10.1037/1040-3590.7.4.524]
30. Yalcin I, Bump RC. Validation of two global impression questionnaires for incontinence. Am J Obstet Gynecol 2003 Jul;189(1):98-101. [Medline: 12861145]
31. Albers-Heitner PCP, Lagro-Janssen TALM, Joore MMA, Berghmans BLCM, Nieman FF, Venema PPL, et al. Effectiveness of involving a nurse specialist for patients with urinary incontinence in primary care: results of a pragmatic multicentre randomised controlled trial. Int J Clin Pract 2011 Jun;65(6):705-712. [doi: 10.1111/j.1742-1241.2011.02652.x] [Medline: 21564445]
32. Shamliyan T, Wyman JF, Ramakrishnan R, Sainfort F, Kane RL. Benefits and harms of pharmacologic treatment for urinary incontinence in women: a systematic review. Ann Intern Med 2012 Jun 19;156(12):861-74, W301. [doi: 10.7326/0003-4819-156-12-201206190-00436] [Medline: 22711079]
33. Lim R, Liong ML, Lim KK, Leong WS, Yuen KH. The Minimum Clinically Important Difference of the International Consultation on Incontinence Questionnaires (ICIQ-UI SF and ICIQ-LUTSqol). Urology 2019 Nov;133:91-95. [doi: 10.1016/j.urology.2019.08.004] [Medline: 31415780]
34. Sjöström M, Stenlund H, Johansson S, Umefjord G, Samuelsson E. Stress urinary incontinence and quality of life: a reliability study of a condition-specific instrument in paper and web-based versions. Neurourol Urodyn 2012 Nov;31(8):1242-1246. [doi: 10.1002/nau.22240] [Medline: 22517196]
35. Uren AD, Cotterill N, Parke SE, Abrams P. Psychometric equivalence of electronic and telephone completion of the ICIQ modules. Neurourol Urodyn 2017 Jun;36(5):1342-1349. [doi: 10.1002/nau.23103] [Medline: 27513627]
36. Nabi G, Cody JD, Ellis G, Herbison P, Hay-Smith J. Anticholinergic drugs versus placebo for overactive bladder syndrome in adults. Cochrane Database Syst Rev 2006 Oct 18(4):CD003781. [doi: 10.1002/14651858.CD003781.pub2] [Medline: 17054185]
37. Mirabegron for treating symptoms of overactive bladder - Technology appraisal guidance TA290. National Institute for HealthCare Excellence (NICE). 2013 Jun 26. URL: https://www.nice.org.uk/guidance/TA290 [accessed 2019-10-17]

38. Samuelsson E, Odeberg J, Stenzelius K, Molander U, Hammarström M, Franzen K, et al. Effect of pharmacological treatment for urinary incontinence in the elderly and frail elderly: A systematic review. Geriatr Gerontol Int 2015 May;15(5):521-534. [doi: 10.1111/ggi.12451] [Medline: 25656412]
39. Sexton CC, Notte SM, Maroulis C, Dmochowski RR, Cardozo L, Subramanian D, et al. Persistence and adherence in the treatment of overactive bladder syndrome with anticholinergic therapy: a systematic review of the literature. Int J Clin Pract 2011 May;65(5):567-585. [doi: 10.1111/j.1742-1241.2010.02626.x] [Medline: 21489081]
40. Balk E, Adam G, Kimmel H, Rofeberg V, Saeed I, Jeppson P, et al. Nonsurgical treatments for urinary incontinence in women: a systematic review update [Internet]. In: AHRQ Comparative Effectiveness Reviews. Report No. 212 (Prepared by the Brown Evidence-based Practice Center under Contract No. 290-2015-00002-I for AHRQ and PCORI). Rockville, MD: Agency for Healthcare Research and Quality; Aug 08, 2018.
41. Kontos E, Blake KD, Chou WS, Prestin A. Predictors of eHealth usage: insights on the digital divide from the Health Information National Trends Survey 2012. J Med Internet Res 2014 Jul 16;16(7):e172 [FREE Full text] [doi: 10.2196/jmir.3117] [Medline: 25048379]
42. Latulippe K, Hamel C, Giroux D. Social Health Inequalities and eHealth: A Literature Review With Qualitative Synthesis of Theoretical and Empirical Studies. J Med Internet Res 2017 Apr 27;19(4):e136 [FREE Full text] [doi: 10.2196/jmir.6731] [Medline: 28450271]
43. Neter E, Brainin E. eHealth literacy: extending the digital divide to the realm of health information. J Med Internet Res 2012 Jan 27;14(1):e19 [FREE Full text] [doi: 10.2196/jmir.1619] [Medline: 22357448]

Abbreviations

eHEALS: eHealth Literacy Scale
IC Scale: Incontinence Catastrophizing Scale
ICIQ-LUTSqol: International Consultation on Incontinence Questionnaire−Lower Urinary Tract Symptoms Quality of Life Module
ICIQ-OAB: International Consultation on Incontinence Questionnaire−Overactive Bladder Module
ICIQ-UI SF: International Consultation on Incontinence Questionnaire−Urinary Incontinence Short Form
IEF: incontinence episode frequency
MUI: mixed urinary incontinence
OR: odds ratio
PFMT: pelvic floor muscle training
PGI-I: Patient Global Impression of Improvement
SUI: stress urinary incontinence
UUI: urgency urinary incontinence

Edited by R Kukafka; submitted 22.05.20; peer-reviewed by J Hay-Smith, S Sankaran, S Farag; comments to author 16.07.20; revised version received 10.09.20; accepted 31.01.21; published 05.04.21.

Please cite as:
Wadensten T, Nyström E, Franzén K, Lindam A, Wasteson E, Samuelsson E
A Mobile App for Self-management of Urgency and Mixed Urinary Incontinence in Women: Randomized Controlled Trial
J Med Internet Res 2021;23(4):e19439
URL: https://www.jmir.org/2021/4/e19439
doi:10.2196/19439
PMID:33818395

Original Paper

Moderated Online Social Therapy for Young People With Active Suicidal Ideation: Qualitative Study

Eleanor Bailey[1,2,3], BA; Jo Robinson[1,3], BSc, MSc, PhD; Mario Alvarez-Jimenez[1,3], BSc, MResMeth, DClinPsy, PhD; Maja Nedeljkovic[2], PhD, MPsych(Clin), MPH; Lee Valentine[1,3], BA, MSocWk; Sarah Bendall[1,3], PhD, PGDipClinPsych; Simon D'Alfonso[1,4], BASc, PhD; Tamsyn Gilbertson[1,3], MPsych(Clin), MSc; Ben McKechnie[1], BA, MPsych(Clin); Simon Rice[1,3], MPsych(Clin), PhD

[1]Orygen, Parkville, Australia

[2]Swinburne University of Technology, Hawthorn, Australia

[3]Centre for Youth Mental Health, University of Melbourne, Parkville, Australia

[4]School of Computing and Information Systems, University of Melbourne, Parkville, Australia

Corresponding Author:
Eleanor Bailey, BA
Orygen
Locked Bag 10
35 Poplar Road
Parkville, 3052
Australia
Phone: 61 412483600
Email: eleanor.bailey@orygen.org.au

Related Article:

This is a corrected version. See correction statement: https://www.jmir.org/2021/6/e29645

Abstract

Background: Web-based interventions are a promising approach to support youth at risk of suicide, and those incorporating peer-to-peer social networking may have the added potential to target interpersonal states of perceived burdensomeness and thwarted belongingness. Owing to feasibility and safety concerns, including fear of contagion, this had not been tested until recently. In 2018, we conducted a pilot evaluation to test the feasibility, safety, and acceptability of a Moderated Online Social Therapy intervention, called Affinity, with a sample of young people with active suicidal ideation.

Objective: The aim of this study is to report qualitative data collected from study participants regarding their experience of the web-based social network and the consequent safety features.

Methods: Affinity is a closed website incorporating 3 key components: therapeutic content delivered via comics, peer-to-peer social networking, and moderation by peers and clinicians. Semistructured interviews were conducted with 17 young people who participated in the pilot study after 8 weeks of exposure to the intervention. Interview data from 2 young people who did not use Affinity were excluded from the analysis. The interviews were analyzed using thematic analysis, with the frequency of responses characterized using the consensual qualitative research method. The results are reported in accordance with the Consolidated Criteria for Reporting Qualitative Research checklist.

Results: A total of 4 overarching themes were identified: a safe and supportive environment, the importance of mutual experiences, difficulty engaging and connecting, and the pros and cons of banning discussions about suicide. Interestingly, although Affinity was perceived to be safe and free of judgment, concerns about negative evaluation and triggering others were significant barriers to posting on the social network. Participants generally supported the banning of conversations about suicide, although for some this was perceived to reinforce stigma or was associated with frustration and distress.

Conclusions: The results not only support the safety and potential therapeutic benefit of the social networking aspect of Affinity but also highlight several implementation challenges. There is a need to carefully balance the need for stringent safety and design features while ensuring that the potential for therapeutic benefit is maximized.

(J Med Internet Res 2021;23(4):e24260) doi:10.2196/24260

KEYWORDS

suicide; youth; social media; internet-based intervention

Introduction

Background

Globally, suicide is the second-leading cause of mortality in young people, defined as those aged between 15 and 24 years [1,2]. In Australia, it is the leading cause of death in youth, accounting for more than one-third of all deaths in this age group [3]. Suicidal ideation and behavior, including self-harm, are relatively common in young people, with approximately 10.6% and 1.2% of young people reporting past 12-month suicidal ideation and suicide attempts, respectively [4]. Suicidal ideation and behavior are risk factors for future suicidal behavior and suicide [5], along with mental disorders including, most significantly, a diagnosis of depression [6]. Moreover, interpersonal states, including perceived burdensomeness (feeling as though one is a burden on others) and thwarted belongingness (lacking a sense of belonging to any particular group) may also confer risk, as articulated by the Interpersonal Theory of Suicide [7,8]. Despite extensive investment and research into youth suicide prevention, the rates of suicide and suicide-related behavior in young people are rising both in Australia and internationally [3,9,10]. The reasons for this are complex, as such there is unlikely to be a single causal factor or solution. However, youth consultations suggest that there is a need to develop and evaluate new and innovative interventions that are theoretically based, target known risk factors for suicide, and are acceptable and relevant for young people [1,11].

The internet has increasingly received attention as a potential platform for the delivery of suicide prevention interventions for young people because of its reach, accessibility, acceptability, and cost-effectiveness [12-14]. Although several systematic reviews and meta-analyses have shown that digital interventions can reduce suicidal ideation and/or behavior in adults [15-18], far fewer studies have focused on developing and testing digital suicide prevention interventions for youth [12,19]. In addition, although the integration of digital interventions with existing mental health services may have several benefits, including improved treatment engagement and outcomes [20,21], to date there have been limited efforts to integrate web-based and face-to-face treatments for young people with suicidal thoughts or behaviors [22]. Finally, the internet has the capacity to facilitate mutual peer-to-peer support, which may in turn have the potential to target interpersonal risk factors for suicide (perceived burdensomeness and thwarted belongingness) [23]; however, no professionally developed intervention has included such a component. This may be in part because of concerns regarding the possible contagion of suicidal behavior, cyberbullying, and trolls [14,23].

Objectives

To address these gaps while also mitigating the potential risks, researchers at Orygen in Melbourne, Australia, developed and pilot tested *Affinity*, a Moderated Online Social Therapy (MOST) [24-26] intervention designed as an adjunct to face-to-face treatment for young people with active suicidal ideation. Quantitative data, reported elsewhere [27], revealed Affinity to be safe, feasible, and acceptable, with exploratory correlations indicating associations between clinical improvement and key aspects of Affinity usage. Given the novel nature of the Affinity intervention, particularly its inclusion of a web-based social network, quantitative data alone are insufficient to understand important aspects of the user experience of the platform. Thus, the aim of this study is to qualitatively explore the views and experiences of young people who participated in the Affinity pilot study. Qualitative outcomes focused on the social networking aspect of the intervention, including moderation and safety features.

Methods

Design

This study reports qualitative interview data collected from young people who took part in a single-arm, pre-post test pilot study of the Affinity intervention conducted in 2018.

Participants

Participants were current clients of a youth mental health clinic in Melbourne, Australia (the Youth Mood Clinic), that specializes in the treatment of young people with severe mood disorders [28]. The Youth Mood Clinic clients were referred to the study by their treating clinicians based on the clinician's perception of their suitability. They were eligible to participate in the trial if they had experienced suicidal ideation within the past 4 weeks (screened for by the lead author using the question, "Have you had any thoughts of suicide in the past four weeks?"), had regular and ongoing internet and telephone access, and were able to give informed consent and comply with study procedures. Additional inclusion criteria to ensure participants' safety and enable execution of safety protocols (if required) were as follows: well engaged with treatment and not approaching discharge in the next 4 weeks; familiar with, and willing to use, crisis supports; and willing and able to nominate 2 emergency contacts. There were no specific exclusion criteria related to the level of suicide risk, although clinicians were consulted on a case-by-case basis regarding participant suitability.

Of the 20 young people who participated in the pilot study, 17 (85%) completed a qualitative interview at follow-up. Of the 3 who did not complete a qualitative interview, 2 were not responsive to the research team's contact attempts at follow-up (but remained engaged in their treatment at the clinical service) and 1 moved overseas before the qualitative interview could be completed. A total of 2 participants did not use the social network during the intervention period, engaging only with the therapeutic comic component of Affinity; therefore, their interview transcripts were excluded from the analysis.

The 15 remaining participants had a mean age of 21.3 years (SD 2.7; range 17-24), with 9 identifying as female, 5 as male, and 1 as transgender. At baseline, 93% (14/15) participants had suicidal ideation scores above the clinical cutoff on the Adult Suicidal Ideation Questionnaire [29] and 73% (11/15) reported at least one previous suicide attempt. On the Patient Health Questionnaire-9 [30], 80% (12/15) participants were in the *severe* or *moderately severe* range for depressive symptoms.

Intervention

The Affinity intervention is described in detail elsewhere [23,27]. Affinity is a closed website that consists of 3 main components: (1) therapeutic content delivered in the form of illustrated comics; (2) peer-to-peer social networking; (3) moderation by clinical experts and young people with lived experience of mental ill health. Users create a profile using either their own first name or a pseudonym, with 67% (10/15) participants in this study choosing to use their real first name. Users can access Affinity 24 hours a day, as often or little as they like, and can use any or all of the components. Clinical and peer moderators promote engagement with the social network and therapeutic content and provide guidance, information, and emotional support to users. Although both clinical and peer moderators post publicly on Affinity and communicate with users via private message, only the clinical moderators are responsible for monitoring and managing clinical risk. Moderators checked Affinity for posts potentially indicative of clinical risk twice daily on weekdays and once daily on weekend days.

In addition to clinical moderation, the Affinity intervention and research design incorporated a number of safety features. First, posts containing keywords related to suicide risk (eg, suicide, suicidal, die) were automatically detected and blocked by the system, which would then trigger a notification to the lead author's business mobile phone. Participants were made aware that this was accessed only during business hours. If notified, the lead author would immediately review the post in question and, if concerned about risk, respond according to the approved study safety protocol (see the following paragraph). Affinity also includes a vent post function, which allows users to select "I'm just venting" before making a post. When selected, the post appears masked by a warning to other users about the potentially inflammatory nature of the post, and users can then elect to view the post's content. The vent post function is primarily designed to protect users from unwillingly viewing posts with swear words. Importantly, vent posts are not immune to the automatic detecting and blocking safety function; as such, posts expressing suicide risk could still be blocked even if "I'm just venting" is selected. Throughout the trial, 3 participants had at least one post that was automatically blocked. One of these posts communicated current suicidal ideation but not imminent risk, and the remainder of the blocked posts were not related to suicide risk but to other violent or aggressive behaviors that participants had experienced. Although the automatic blocking system used simple string matching and could not detect alternatives of keywords, such as sui*c*de, no participants in this study attempted to bypass the system.

A comprehensive safety protocol was in place outlining risk assessment and management procedures, including provisions for telephoning participants' emergency contacts and/or emergency services, if risk was assessed to be high. All information about clinical risk identified during participation in the trial was communicated to the participants' treating clinicians. Participants were also required to agree to a *terms of use* before being given access to Affinity, which included a request not to share personal or contact information with other users via the website to prevent conversations about suicide happening externally. These measures were in place to mitigate the risk of distress to participants or the contagion of suicidal behavior [23]. Accordingly, although participants were aware that the purpose of Affinity was to support young people who experience suicidal thoughts, they were actively discouraged from talking about suicide in the social network.

Procedure

Participants were referred to the study by their treating clinicians between April and August 2018 and were given access to Affinity from the point of entry into the study until the intervention was closed (October 31, 2018). Semistructured interviews were conducted with each participant approximately 8 weeks after they were given access to Affinity by the lead author (EB), an experienced research assistant and a PhD candidate. EB had an established relationship with the participants in that she had also conducted the baseline assessments but was not involved in the delivery of the intervention itself (ie, did not post on Affinity or use it to communicate with users). The interviews were conducted either in a private room at the mental health service or in participants' homes, with only the interviewer and interviewee present. The interview length ranged from 26 minutes to 75 minutes, with a mean duration of 48.3 (SD 16.4) minutes. Table 1 displays each participant's study ID, age, interview length, severity of depression symptoms at baseline, and suicidal ideation score at baseline. To protect the identity of the transgender participant, participants' genders are not provided. Participants were informed that the purpose of the interview was to explore positive, negative, and neutral feedback about Affinity, and were encouraged to be as open and honest as possible. As the purpose was to obtain feedback from as many participants as possible regarding their experience, data saturation was not assessed.

The interview schedule was designed to obtain participants' views and experiences about Affinity, with prompts included if answers were vague. For the purpose of this paper, only responses related to the social network, safety features, and peer and clinical moderation were analyzed. The full interview schedule is provided in Multimedia Appendix 1. All interviews were audio recorded and transcribed verbatim, and potentially identifiable information was removed. Brief field notes were recorded during interviews.

This study was approved by the Melbourne Health Human Research Ethics Committee (ID 2017.187). All participants provided written informed consent. Participants under the age of 18 years were required to provide consent from their parents or guardians.

Table 1. Participant ID, age, interview length, and clinical characteristics at baseline (N=15).

ID	Age (years)	Interview length (min)	Baseline depression category (Patient Health Questionnaire-9)	Baseline suicidal ideation score (ASIQ)[a]
2	20	48	Severe	130
3	23	68	Moderate	42
5	22	50	Moderately severe	61
6	24	70	Moderately severe	24
7	23	71	Severe	130
8	23	32	Moderately severe	120
10	21	48	Severe	79
11	18	53	Moderately severe	119
12	24	46	Moderate	78
13	17	26	Severe	132
15	17	26	Moderately severe	69
16	17	34	Severe	136
17	24	75	Severe	78
18	23	39	Moderate	74
19	23	39	Severe	79

[a]ASIQ: The Adult Suicidal Ideation Questionnaire. The scale has a possible score range of 0-150; scores of 31 or more are considered to be in the clinical range.

Data Reporting and Analysis

Data were reported in accordance with the Consolidated Criteria for Reporting Qualitative Research [31]. The checklist is provided in Multimedia Appendix 2. Only interview data pertaining to the social network, moderation, and safety features were analyzed for the purpose of this paper. Data were analyzed using inductive thematic analysis following 6 steps by Braun and Clarke [32]: (1) familiarizing with data, (2) generating an initial coding frame, (3) searching for themes, (4) reviewing themes, (5) defining and naming themes, and (6) reporting. The lead author (EB) read and reread the interview transcripts to immerse herself in the data, and based on this, an initial coding frame was generated. A coauthor (LV) checked the coding frame against a 10% subset of the data (ie, 2 interviews) and then discussed this with EB to refine the coding frame. The coding frame was then applied to the data provided by the lead author. Codes were then grouped into initial themes, which were reviewed and refined, with some combined and grouped into minor themes as necessary. SR, MN, and JR were regularly consulted regarding the codes and themes developed. The data were coded, and codes were grouped into themes by hand and then transferred to a web-based whiteboard platform (Miro) to produce hierarchical thematic maps, where they were organized, reviewed, and refined.

Data were reported according to the consensual qualitative research method [28], using the labels *few* (rare support; endorsed by 10%-20% of the respondents), *some* (variant support; endorsed by 21%-50% of the respondents), *many* (typical support; endorsed by 51%-90% of the respondents), or *most* (general support; endorsed by 91%-100% of the respondents).

Results

Overview

A total of 4 key themes were identified: (1) a safe and supportive environment, (2) the importance of mutual experiences, (3) difficulties in connecting and engaging, and (4) the pros and cons of banning discussions about suicide. These themes are outlined later and illustrated with exemplar quotes. Study ID and age are provided alongside quotes. The hierarchical thematic maps are available in Multimedia Appendix 3, and a table of themes, subthemes and/or codes, endorsement percentages, and exemplar quotes are available in Multimedia Appendix 4.

Theme 1: A Safe and Supportive Environment

The first theme describes participants experiencing Affinity as a positive environment, wherein they felt both safe and supported. Indeed, many participants (11/15, 73%) specifically discussed feeling safe on Affinity. This sense of safety was discussed both in terms of being shielded from negative content and feeling safe from ridicule or judgment. A few participants (2/15, 13%) also referred to feeling safe because they knew their privacy was protected.

Some participants (6/15, 40%) specifically referred to the presence of clinical moderators as contributing to the sense of safety, in that users knew "if something did happen that there were people there that could step in" (ID 08, aged 22 years). A few participants (3/15, 20%) also reported feeling safe because they knew that all users were well-intentioned, with one stating they knew other users "weren't there to harm me or affect me negatively in any way. They were just there to better themselves" (ID 02, aged 20 years). This belief in the good intentions of other users was partly attributed to their mutual

experiences, although this was identified as a distinct theme and is therefore elaborated separately later.

Some (6/15, 40%) specifically labeled Affinity as a *friendly* and/or *supportive* space. Many (8/15, 53%) participants said that they had provided support to other people on Affinity; 6 of these said this was a positive experience, with 1 participant noting "…it feels good to know you've helped someone" (ID 05, aged 22 years). Two of these participants noted negative impacts in addition to discussing positive outcomes of supporting others, with one stating "it can be frustrating when people don't want to listen to what you've got to say" (ID 07, aged 23 years), and another referring to their tendency to worry about other people at the expense of "fully looking after myself" (ID 17, aged 23 years).

Participants also expressed appreciation for the way the peer and clinical moderators connected with them and the social network. Many (8/15, 53%) valued when moderators reached out with personalized messages to check on them or suggest particular therapeutic components to try within Affinity:

> *[The moderators] sent us really big paragraphs which I really liked. I don't care how long they are, but the fact that they took time to write that and it was meaningful, it was really supportive.* [ID 16, aged 16 years]

Although many participants (9/15, 60%) also appreciated the effort the moderators put into facilitating group conversations and welcoming new users, some participants (4/15, 27%) felt they were overly enthusiastic in their attempts to promote connections within the network:

> *They were good. For one thing, it made it so that there was no post that was really ignored...But I will say that there were points where it felt a bit artificial to an extent.* [ID 11, aged 18 years]

Some participants (6/15, 40%) specifically drew comparisons between Affinity and other mainstream social networking sites, such as Facebook, stating that Affinity felt safer and more supportive:

> *People feel more secure being open in that area, as opposed to other social media. They could be attacked or feel triggered. But in Affinity, we're all here to support each other.* [ID 03, aged 22 years]

> *When I'm in a good mood Facebook was great. But when I'm in a bad mood or have anxiety, I'm not sure, something about it upsets me. I never got that from Affinity, I'm not sure if it's because I didn't know the people, or it's because the posts were positive, or because I knew they were going through the same thing I was going through. Whatever it was, I never felt upset.* [ID 12, aged 23 years]

For some (7/15, 47%), Affinity was therefore perceived to provide a way for safely and easily interacting, or connecting, with others, which was particularly helpful in periods of low mood or isolation when any form of interaction was difficult:

> *I mean, for me I feel very disconnected from the world, so I feel a little bit more connected when I'm on Affinity. Just a little bit more. It's not fixing anything per se but I do feel a little better using it.* [ID 06, aged 24 years]

> *Previous, what I would normally do is just cut contact off completely. This gave me one thing that I could stay on.* [ID 12, aged 23 years]

Theme 2: The Importance of Mutual Experiences

Many participants (13/15, 87%) specifically spoke about valuing participation in a network of other people with similar lived experiences (namely, depression and suicidal thoughts). Many (8/15, 53%) specifically spoke about how this helped them to feel less alone or *crazy* with regard to these experiences:

> *You read posts and it would be something, you'd be like oh I've had that thought or that's how I feel. Then in your head you're like oh, this person's feeling that way too. [...] You read that post and think oh okay I'm not crazy, I'm not the only person that thinks that.* [ID 08, aged 22 years]

> *Talking to other people that actually feel what I do is - I mean yeah, it's pretty good. I think the most terrible thing about depression is it makes me feel like I'm the only person in the world who feels that and it makes me feel lonely.* [ID 10, aged 21 years]

Some (7/15, 47%) also discussed how this led to them feeling particularly validated and understood by the other users, with one stating:

> *That was something I could talk about on Affinity, and be candid about on Affinity and not - I could tell my friends, but there was - it was - there was something about saying it in front of other people who understood about this.* [ID 11, aged 18 years]

Some participants (4/15, 27%) also reported that they were able to learn from other users, because they either provided relevant advice or a new perspective or posted about how they had dealt with their own problems. For example, one participant reflected that suggestions from other users were "more useful than people that haven't been through this kind of thing, purely because they themselves know more or less what helps and what doesn't" (ID 05, aged 22 years). Another user discussed how responses to a different user's post about a "problem that was similar to mine...helped me see things from the other person's side of view and look at their problem in a different way" (ID 02, aged 20 years). A few participants (3/15, 20%) said that hearing about the experiences of others was encouraging in terms of their own ability to "get through this." One participant said that knowing the peer moderators had lived experience of mental ill health and that they were in recovery "gives a little bit of hope" (ID 02, aged 20 years).

Finally, some participants (4/15, 27%) specifically stated that they experienced many of the benefits related to mutual experiences even though they did not actively engage in social networking; in other words, just reading posts was beneficial. One participant said that even though they did not post, "it helped to just read someone else's post about how they were feeling" (ID 17, aged 23 years).

Theme 3: Difficulties in Engaging and Connecting

Participants also reported that they encountered barriers to engaging with the affinity social network and/or connecting with other users. These barriers have been categorized into subthemes of internal barriers (related to feelings of anxiety or apprehension) and external barriers (attributes specific to Affinity).

Internal Barriers to Engaging and Connecting

Many participants (12/15, 80%) discussed feelings of anxiety or apprehension related to posting on the network and/or replying to posts made by others. A total of 2 subthemes related to the source of this anxiety were identified: fear of negative evaluation and fear of causing harm.

Fear of Negative Evaluation

Many (8/15, 53%) participants discussed feelings of anxiety related to posting or replying to posts on Affinity, often related to concerns that other users would not care about their post or that no one would respond to their disclosure or comment. For example, one participant (ID 18, aged 23 years) discussed not wanting to be "the first one to comment," for fear that other users "won't react to it and you'd just be sitting there and no-one will give a shit about what you said." Some participants (4/15, 27%) said this uncertainty manifested as "overthinking," with one participant stating, "I just kept second guessing whatever I would write whenever I wanted to make a post" (ID 02, aged 20 years).

One participant (ID 17, aged 23 years) likened this to anxiety associated with posting on other forms of social media, stating "social media has ruined people, and they're like, ah, is this going to get likes, or are people going to comment back to this?" Indeed, some (5/15, 33%) specifically stated that these thoughts and feelings were typical for them and extended to all social media and did not attribute them to Affinity specifically.

Some participants (4/15, 27%) expressed confusion about the fact that they felt anxious despite the supportive nature of the social network, with one saying, "It's weird because we know that it's a safe place, and yet we still get very anxious about those things" (ID 19, aged 23 years). Another said, "I don't know why I was being super scared, because I have all this proof that everybody on there is actually really supportive and nice and nothing that I'm posting is too bad or controversial" (ID 02, aged 20 years).

Fear of Causing Harm

Some participants (6/15, 40%) reported that they held back from posting or replying to posts because they were worried about causing distress or being unhelpful to others, with 2 users expressing uncertainty regarding permitted posts:

> *I guess I didn't even post things, like thoughts that I might be having, no one likes me or I'm just a complete failure or life's not worth living, I feel so hopeless, I feel like I don't have a future. I wouldn't say stuff like that, because I thought it would be too triggering for other people.* [ID 07, aged 23 years]

> *I felt even if I had something to say, I didn't feel comfortable saying it. I wasn't sure if I wrote something if it'd make it worse, or I'm not sure how to feel about giving other people advice. So, I kind of deliberately took a - even if I thought I had advice, I wouldn't give it, because I'm not sure how the reaction would be.* [ID 12, aged 23 years]

One participant suggested that this uncertainty could have been better managed if Affinity had contained a list of topics or phrases that were permitted and not permitted or if moderators had initiated group discussions around sensitive issues.

External Barriers to Engaging and Connecting

In total, 2 external barriers were identified: the small user base and the impersonal nature of the interactions with other users.

Small, Inactive User Base

Many participants (9/15, 60%) stated that there were too few users on Affinity and/or that the user base was too inactive. A few participants (2/15, 13%) specifically said that this reduced their motivation to log in altogether because "there isn't much happening" (ID 06, aged 24 years). A few (3/15, 20%) stated this hindered willingness to post because they anticipated a delayed response, and 2 participants reported that a lack of timely or prompt acknowledgment or validation from the network in relation to comments or disclosures made on Affinity had a negative impact:

> *You can post something and go unnoticed for two days. So, in that aspect, it [...] felt almost a little counterproductive [...]it just didn't necessarily help with the isolation.* [ID 05, aged 22 years]

> *Just say I type something, and no one responded within a day, I'm like, oh, they don't care about me. Oh, they don't want to listen. I made them feel shit, or something. Those thoughts keep going around in my head, and I'm like, oh, I shouldn't have said that.* [ID 16, aged 16 years]

Impersonal Interactions

Some participants (7/15, 47%) expressed dissatisfaction with their inability to interact more closely or on a long-term basis with other users. Specifically, the prohibition of sharing personal or contact information, the lack of a private user-to-user chat function, and the temporary nature of Affinity were perceived to have contributed to difficulties establishing meaningful connections with other users:

> *It probably helped a little bit and made me feel connected and heard, but not like - it didn't, you know, completely erase my feeling of isolation and loneliness because at the end of the day, it's a temporary thing, it's not permanent, I don't know who anyone is on there.* [ID 07, aged 23 years]

Theme 4: Pros and Cons of Banning Discussions About Suicide

The fourth general theme was related to the fact that both positive and negative feedback was received regarding the banning of suicide-related posts on Affinity. A total of 11 participants specifically discussed this policy; of these, 7 (64%) generally supported it, 2 (18%) did not, and 2 (18%) mentioned

both positive and negative impacts. Participants in support of prohibiting suicide-related posts stated that they felt conversations about suicide could be harmful or *triggering* for themselves or other users, with one stating:

> *For me, it would be very triggering if I were to see anything about suicidal thoughts. But that depends on the person I guess because I think I'm just very emotional, and very easily influenced.* [ID 19, aged 23 years]

Some participants (3/11, 27%) specifically stated that they valued the *vent post* function of Affinity.

One participant with mixed views about these terms of use stated:

> *It's definitely a little bit weird that [Affinity] is for people who are feeling suicidal but you can't really talk about it or - definitely seems like it had sort of a stigma as well. But I feel like it was good that you didn't let us talk about it like that because I feel like if I saw somebody talking about suicide, then it would get me thinking. Or I'd be super worried about them and I don't think either of those are helpful for me or anybody else on the website.* [ID 02, aged 20 years]

Another participant suggested that banning discussions about suicide could be challenging for people who *need to talk about* their suicidality. Indeed, some participants (3/11, 27%) spoke about needing to vent without necessarily needing a crisis intervention, for example, one said:

> *I know for myself sometimes when you want to - you just need to say something and to some people it's going to sound really bad but you're genuinely like I just need to get this out of my head.* [ID 08, aged 22 years]

One participant said it would have been helpful for them if they had been able to engage in discussions openly about suicide on Affinity:

> *So when I was alone and I was able to be on my phone and look through, it just would have helped [...] to have other people share their actual experiences.* [ID 17, aged 23 years]

A few participants (2/11, 18%) described the experience of having a post automatically blocked by the system and both found it a negative experience, mostly because of confusion about why it was blocked in the first place:

> *It was just like annoying. I was just like, really? Especially [because] there was no real obvious reason as to why it got blocked. I feel like for me if it was an obvious reason I'd be like okay, I can change that, but I'm looking at it going I can't actually change anything in this to make it any less - different to what it is.* [ID 08, aged 22 years]

> *I thought somebody took time out of their life to block my - or report me. I was like, are you serious? [...] Yeah, just seems like nothing now, but at the time I got really upset. Other people were using swearing and stuff. How can F-U-C-K be involved but not death? That's what I didn't understand, but yeah, I get it now.* [ID 16, aged 16 years]

Discussion

Principal Findings

This study used thematic analysis to examine the experiences of 15 young people who used the therapeutic web-based social networking platform, Affinity. We found that participants experienced Affinity as a safe and supportive environment where they felt less alone and understood by others, yet also experienced barriers to fully engaging and connecting. We also found that although participants generally supported banning discussions about suicide, some potential adverse effects were noted.

Positive Appraisals of the Affinity Social Network

The finding that Affinity was perceived to be supportive and safe and that participants valued being surrounded by others with similar experiences aligns with the quantitative results we have reported elsewhere supporting the acceptability and safety of Affinity [27]. The importance of mutual experiences has also been identified as a theme in other studies using the MOST platform [33,34] and the literature on online support groups for suicidal people [35,36]. The positive impact of providing support to others was also raised by some participants and has been identified in the broader online support group literature [36], although it is acknowledged that 2 participants in this study also reported negative aspects of supporting others, which are important and worthy of further exploration. For example, phenomena such as compassion fatigue or burnout may impact young people supporting other young people on Affinity; this has been hypothesized to occur in trained volunteers from at least one anonymous online forum for suicidal people [37].

Previously, we theorized that the peer-to-peer support facilitated by Affinity could mitigate the key risk factors described in the Interpersonal Theory of Suicide [7,8] of thwarted belongingness (via immersion in a network of similar others) and perceived burdensomeness (by enabling users to help others in need) [23]. The feeling of being less alone reported by participants suggests that they may have experienced an increased sense of belongingness; this aligns with our quantitative results showing a significant and large effect size (Cohen d=−0.96; P=.006) improvement in thwarted belongingness [27]. Regarding the possible influence of Affinity on perceived burdensomeness, as noted earlier, several participants reported that providing support to others was a positive experience. Although limited data preclude a deeper exploration of this, it is possible that helping others may effectively lessen the sense of burdensomeness [23]. Conversely, however, it may also be that fear of causing harm to others, either by triggering them or by providing advice perceived to be *bad* in response to a post, may also serve to increase burdensomeness. Therefore, further studies are warranted. Overall, these findings attest to the significance of the social networking component of Affinity and that it may have therapeutic benefits in and of itself. Moreover, some participants reported experiencing these benefits just by reading other users' posts, suggesting that the positive impact of mutual experiences may occur even without active participation. The

ability of participants to experience therapeutic benefit from passively participating in web-based peer support networks has also been found in previous studies using the MOST model [33,38] and in studies of online health-related support groups more broadly [39,40].

Barriers to Engaging and Connecting

Despite positive appraisals, participants also reflected on experiencing barriers to engaging and connecting in the trial. Perhaps unsurprisingly, given that most participants had depression symptoms in the severe or moderately severe range, fear of negative evaluation was a key barrier to engaging with the social network. Research has shown a link between symptoms of depression and a tendency to participate passively (or *lurk*) in web-based social networks [41-44]; indeed, participants in this study likened their apprehension about posting or commenting on Affinity to how they feel about social media more generally. Despite Affinity's vent post feature, which 3 participants specifically discussed in positive terms, a key barrier to engaging with the social network was fear of causing harm or triggering others. Participants also reported an inability to properly connect with other users, attributed to the finite period of access to Affinity as well as the lack of one-to-one chat and ban on sharing contact details. Building presence in a social network can be a significant investment of an individual's effort and time, and this was likely negated in this study, given the short intervention period and lack of opportunity to connect with users beyond the intervention. Although these safety and design choices were deliberately made because of the novel nature of the intervention and the high-risk nature of the sample, these will likely be carefully relaxed in future iterations of Affinity.

Balancing Risks and Benefits of Talking About Suicide

The finding that many participants supported the prohibition of suicide-related posts is somewhat surprising; we expected participants may have experienced this feature as authoritarian based on research suggesting that young people want to talk about suicide with one another and with adults [45]. This previous research, however, was conducted with young people who were not currently at risk. It is possible that young people with active suicidal ideation are more likely than those who have recovered to find open discussions about suicide distressing, particularly where suicidal cognitions are experienced as intrusive and involuntary. Indeed, concerns about becoming distressed or suicidal themselves were cited as key reasons for supporting the prohibition of discussions about suicide. Despite this, several adverse effects associated with banning discussions about suicide have also been raised by the participants. For example, one participant suggested that this may serve to perpetuate stigma. Given that openly talking about suicide to alleviate stigma and encourage help seeking is the rationale underpinning suicide prevention media campaigns [46], the potentially stigmatizing nature of prohibiting conversations about suicide on Affinity is worthy of consideration. Participants also suggested that talking about suicidal thoughts may be helpful for some people, and one participant said it would have been helpful to see how other people on Affinity deal with their suicidal thoughts; possibly, the benefits associated with mutual experiences would be heightened if these discussions were permitted. In addition, the participants who experienced having a post automatically blocked by the system reflected that this was a negative and confusing experience for them. This way, the automatic blocking system used in the Affinity pilot may actually be counterproductive and lead to increased distress, particularly for users who are already in a distressed state when posting. In the pilot, the list of risk words that would trigger a post to be blocked was not disclosed to participants nor were they informed which words had triggered the blocking after the fact. It is therefore possible that communicating this information to participants would have reduced their distress and frustration in response to having a post blocked. Given the mixed responses to the banning of discussions about suicide, more research is required to determine under what circumstances, by what mechanisms, and for whom, talking about suicide on platforms such as Affinity may be helpful or harmful. One possibility is that young people actively trying to suppress or avoid their suicidal thoughts as a coping strategy might experience anxiety or distress when encouraged to discuss them [47].

Implementation Challenges and Future Directions

Although the first 2 themes are indeed encouraging and provide further support for the acceptability, safety, and potential therapeutic benefit of Affinity, the barriers to engaging and connecting and mixed views on banning discussions about suicide pose a number of implementation challenges. First, there is a question of how to promote active participation in the social network to a population of users who may be particularly sensitive to judgment from others and who may perceive social media as being *unsafe*. Indeed, previous research has identified that adolescents tend to view social media as judgmental and threatening, particularly where they may be affected by mental illness, making them more susceptible to these cognitions [48,49]. Other challenges relate to how to allow users to connect more personally with each other and how to allow discussions about suicide to occur while maintaining adequate safety standards. There is clearly a careful balance to be struck so that the potential benefits of Affinity, particularly those related to mutual experiences, can be maximized while still protecting users who may be more susceptible to feeling distressed or triggered. This tension, associated with balancing safety features and ensuring potential benefits are maximized, is common to internet-based suicide prevention intervention research more broadly [50]. Given the quantitative [27] and qualitative findings of this study supporting the safety of Affinity, there is an opportunity to carefully relax some of the safety features used in the pilot study in the future. The inclusion criterion related to participants being sufficiently engaged with the clinical service, which in this study was implemented to ensure safety protocols could be appropriately executed, could potentially be relaxed in future iterations so that young people who are less well engaged in treatment could provide access to this potentially helpful intervention. Other options for consideration in future iterations of Affinity could include scheduled, moderator-facilitated discussions about suicidal thoughts focusing on helpful strategies and stories of hope and the promotion of guidelines about how to safely talk about suicide

[51], including specific advice about unacceptable content (eg, threatening suicide, inciting suicide in others). Importantly, users should be able to avoid discussions about suicide altogether on Affinity, without fear of judgment. In the absence of evidence-based guidelines for implementing digital interventions for this population, decisions regarding these features should be theoretically and empirically driven, consumer led, and carefully evaluated in an ongoing way regarding their acceptability and safety.

Limitations

Several limitations should be considered when interpreting the findings of this study. First, the interviews focused on breadth, rather than depth of data; as such, we were unable to explore in great detail the themes and subthemes that were identified. Future research is warranted to explore some ideas in detail; for example, understanding moderators of engagement and different engagement profiles would be important for the development of targeted consumer-informed web-based interventions for this population. Second, the sample size of 15, which is typical for qualitative research, prevents generalization of these findings beyond this study. Third, it is acknowledged that qualitative methods bring a degree of researcher subjectivity, which may have influenced the analysis and interpretation of results. To address this, regular discussions between members of the research team were held throughout the analysis process. Fourth, participants were not asked to review the transcripts or the study findings; however, as the transcripts were recorded and transcribed verbatim, the likelihood of error was minimal. Moreover, given the difficulty associated with contacting participants in this study, it was not appropriate or feasible to request that they provide feedback on the findings. Fifth, we did not analyze relationships between the themes, for example, whether participants who thought they benefited from passive participation were also those who experienced fear of negative evaluation in relation to posting; this should be a priority for future research. Finally, we did not include a measure of social anxiety, which would have shed more light on participants' experience of barriers to posting; this should be a focus of future evaluations of Affinity.

Conclusions

This study provides important preliminary data on the user experience of a web-based intervention for young people at risk of suicide, incorporating a social networking component and strict safety features. The findings suggest that the social networking component of Affinity is both safe and may possibly have therapeutic benefit, although participants experienced barriers to engaging and connecting related to their underlying feelings of anxiety as well as to design and the safety features of Affinity. Therefore, there is a need to carefully balance ensuring participants' safety with maximizing the potential for therapeutic benefit. That participants experienced benefit despite the ban on conversations about suicide and that passive participation was reported to be beneficial suggest that just knowing one part of a network of similar others may be in and of itself therapeutic for young people at risk of suicide. Future iterations of Affinity should consider carefully relaxing the safety features while continually monitoring and evaluating their acceptability and safety; these decisions should be based on the theoretical and empirical literature and made collaboratively with consumers.

Acknowledgments

EB is a PhD (Clinical Psychology) candidate at Swinburne University of Technology and is supported by an Australian Government Research Training Program Scholarship. She is supervised by JR, SR, and MN. JR (APP1142348) and SR (APP1158881) are supported by National Health and Medical Research Council Career Development Fellowships. MA was supported by an investigator grant (APP1177235) from the National Health and Medical Research Council. SR was also supported by the Mary Elizabeth Watson Early Career Fellowship in Allied Health from the Royal Melbourne Hospital. Future Generations Global provided additional financial support to this project.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Interview schedule.
[DOCX File , 15 KB - jmir_v23i4e24260_app1.docx]

Multimedia Appendix 2
Consolidated criteria for reporting qualitative research checklist.
[DOCX File , 16 KB - jmir_v23i4e24260_app2.docx]

Multimedia Appendix 3
Hierarchical thematic maps.
[DOCX File , 68 KB - jmir_v23i4e24260_app3.docx]

Multimedia Appendix 4

Summary table of themes, codes, percentage endorsed, and example quotes.
[DOCX File , 16 KB - jmir_v23i4e24260_app4.docx]

References

1. Suicide Data. World Health Organization. 2018. URL: http://www.who.int/mental_health/prevention/suicide/suicideprevent/en/ [accessed 2018-02-11]
2. Youth. United Nations. 2020. URL: https://www.un.org/en/sections/issues-depth/youth-0/ [accessed 2020-10-18]
3. 3303.0 - Causes of death, Australia, 2018. Australian Bureau of Statistics. 2018. URL: https://www.abs.gov.au/AUSSTATS/abs@.nsf/Explanatory%20Notes/3303.0 [accessed 2021-02-18]
4. Mortier P, Cuijpers P, Kiekens G, Auerbach RP, Demyttenaere K, Green JG, et al. The prevalence of suicidal thoughts and behaviours among college students: a meta-analysis. Psychol Med 2018 Mar;48(4):554-565. [doi: 10.1017/S0033291717002215] [Medline: 28805169]
5. Castellví P, Lucas-Romero E, Miranda-Mendizábal A, Parés-Badell O, Almenara J, Alonso I, et al. Longitudinal association between self-injurious thoughts and behaviors and suicidal behavior in adolescents and young adults: A systematic review with meta-analysis. J Affect Disord 2017 Jun;215:37-48. [doi: 10.1016/j.jad.2017.03.035] [Medline: 28315579]
6. Gili M, Castellví P, Vives M, de la Torre-Luque A, Almenara J, Blasco MJ, et al. Mental disorders as risk factors for suicidal behavior in young people: a meta-analysis and systematic review of longitudinal studies. J Affect Disord 2019 Feb 15;245:152-162. [doi: 10.1016/j.jad.2018.10.115] [Medline: 30390504]
7. Joiner TE. Why people die by suicide. Cambridge, USA: Harvard University Press; 2007.
8. Van Orden KA, Witte TK, Cukrowicz KC, Braithwaite SR, Selby EA, Joiner TE. The interpersonal theory of suicide. Psychol Rev 2010 Apr;117(2):575-600 [FREE Full text] [doi: 10.1037/a0018697] [Medline: 20438238]
9. Twenge JM, Cooper AB, Joiner TE, Duffy ME, Binau SG. Age, period, and cohort trends in mood disorder indicators and suicide-related outcomes in a nationally representative dataset, 2005-2017. J Abnorm Psychol 2019 Apr;128(3):185-199. [doi: 10.1037/abn0000410] [Medline: 30869927]
10. Griffin E, McMahon E, McNicholas F, Corcoran P, Perry IJ, Arensman E. Increasing rates of self-harm among children, adolescents and young adults: a 10-year national registry study 2007-2016. Soc Psychiatry Psychiatr Epidemiol 2018 Jul;53(7):663-671. [doi: 10.1007/s00127-018-1522-1] [Medline: 29721594]
11. Robinson J, Bailey E, Browne V, Cox G, Hooper C. Raising the bar for youth suicide prevention. The National Centre of Excellence in Youth Mental Health. 2016. URL: https://www.orygen.org.au/Policy/Policy-Reports/Raising-the-bar-for-youth-suicide-prevention/orygen-Suicide-Prevention-Policy-Report?ext [accessed 2021-02-18]
12. Perry Y, Werner-Seidler A, Calear A, Christensen H. Web-based and mobile suicide prevention interventions for young people: a systematic review. J Can Acad Child Adolesc Psychiatry 2016;25(2):73-79 [FREE Full text] [Medline: 27274742]
13. Rice S, Robinson J, Bendall S, Hetrick S, Cox G, Bailey E, et al. Online and social media suicide prevention interventions for young people: a focus on implementation and moderation. J Can Acad Child Adolesc Psychiatry 2016;25(2):80-86 [FREE Full text] [Medline: 27274743]
14. Marchant A, Hawton K, Stewart A, Montgomery P, Singaravelu V, Lloyd K, et al. A systematic review of the relationship between internet use, self-harm and suicidal behaviour in young people: the good, the bad and the unknown. PLoS ONE 2017 Aug 16;12(8):-. [doi: 10.1371/journal.pone.0181722]
15. Torok M, Han J, Baker S, Werner-Seidler A, Wong I, Larsen ME, et al. Suicide prevention using self-guided digital interventions: a systematic review and meta-analysis of randomised controlled trials. The Lancet Digital Health 2020 Jan;2(1):e25-e36. [doi: 10.1016/s2589-7500(19)30199-2]
16. Witt K, Spittal MJ, Carter G, Pirkis J, Hetrick S, Currier D, et al. Effectiveness of online and mobile telephone applications ('apps') for the self-management of suicidal ideation and self-harm: a systematic review and meta-analysis. BMC Psychiatry 2017 Aug 15;17(1):297 [FREE Full text] [doi: 10.1186/s12888-017-1458-0] [Medline: 28810841]
17. Arshad U, Farhat-Ul-Ain, Gauntlett J, Husain N, Chaudhry N, Taylor PJ. A Systematic Review of the Evidence Supporting Mobile- and Internet-Based Psychological Interventions For Self-Harm. Suicide Life Threat Behav 2020 Feb;50(1):151-179 [FREE Full text] [doi: 10.1111/sltb.12583] [Medline: 31448847]
18. Melia R, Francis K, Hickey E, Bogue J, Duggan J, O'Sullivan M, et al. Mobile Health Technology Interventions for Suicide Prevention: Systematic Review. JMIR Mhealth Uhealth 2020 Jan 15;8(1):e12516 [FREE Full text] [doi: 10.2196/12516] [Medline: 31939744]
19. Robinson J, Bailey E, Witt K, Stefanac N, Milner A, Currier D, et al. What works in youth suicide prevention? A systematic review and meta-analysis. EClinicalMedicine 2018;4-5:52-91 [FREE Full text] [doi: 10.1016/j.eclinm.2018.10.004] [Medline: 31193651]
20. Erbe D, Eichert H, Riper H, Ebert DD. Blending face-to-face and internet-based interventions for the treatment of mental disorders in adults: systematic review. J Med Internet Res 2017 Sep 15;19(9):e306 [FREE Full text] [doi: 10.2196/jmir.6588] [Medline: 28916506]
21. Fairburn CG, Patel V. The impact of digital technology on psychological treatments and their dissemination. Behav Res Ther 2017 Jan;88:19-25 [FREE Full text] [doi: 10.1016/j.brat.2016.08.012] [Medline: 28110672]

22. Mohr DC, Riper H, Schueller SM. A solution-focused research approach to achieve an implementable revolution in digital mental health. JAMA Psychiatry 2018 Feb 01;75(2):113-114. [doi: 10.1001/jamapsychiatry.2017.3838] [Medline: 29238805]
23. Bailey E, Rice S, Robinson J, Nedeljkovic M, Alvarez-Jimenez M. Theoretical and empirical foundations of a novel online social networking intervention for youth suicide prevention: a conceptual review. J Affect Disord 2018 Oct 01;238:499-505 [FREE Full text] [doi: 10.1016/j.jad.2018.06.028] [Medline: 29936387]
24. Lederman R, Wadley G, Gleeson J, Bendall S, Álvarez-Jiménez M. Moderated online social therapy: designing and evaluating technology for mental health. ACM Trans Comput Hum Interact 2014 Feb 01;21(1):1-26. [doi: 10.1145/2513179]
25. Alvarez-Jimenez M, Rice S, D'Alfonso S, Leicester S, Bendall S, Pryor I, et al. A novel multimodal digital service (moderated online social therapy+) for help-seeking young people experiencing mental ill-health: pilot evaluation within a national youth e-mental health service. J Med Internet Res 2020 Aug 13;22(8):- [FREE Full text] [doi: 10.2196/17155] [Medline: 32788151]
26. Rice S, Gleeson J, Davey C, Hetrick S, Parker A, Lederman R, et al. Moderated online social therapy for depression relapse prevention in young people: pilot study of a 'next generation' online intervention. Early Interv Psychiatry 2018 Aug;12(4):613-625. [doi: 10.1111/eip.12354] [Medline: 27311581]
27. Bailey E, Alvarez-Jimenez M, Robinson J, D'Alfonso S, Nedeljkovic M, Davey CG, et al. An enhanced social networking intervention for young people with active suicidal ideation: safety, feasibility and acceptability outcomes. Int J Environ Res Public Health 2020 Apr 03;17(7):- [FREE Full text] [doi: 10.3390/ijerph17072435] [Medline: 32260111]
28. Rice SM, Halperin S, Cahill S, Cranston I, Phelan M, Hetrick SE, et al. The Youth Mood Clinic: an innovative service for the treatment of severe and complex depression. Australas Psychiatry 2017 Apr;25(2):112-116. [doi: 10.1177/1039856216689002] [Medline: 28135816]
29. Reynolds WM. Adult suicidal ideation questionnaire: professional manual. Psychological Assessment Resources. Florida: Odessa; 1991. URL: http://www.str-tn.org/suicidal_ideation_questionnaire_professional_manual.pdf [accessed 2021-02-18]
30. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
31. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
32. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
33. Valentine L, McEnery C, O'Sullivan S, Gleeson J, Bendall S, Alvarez-Jimenez M. Young people's experience of a long-term social media-based intervention for first-episode psychosis: qualitative analysis. J Med Internet Res 2020 Jun 26;22(6):- [FREE Full text] [doi: 10.2196/17570] [Medline: 32384056]
34. Santesteban-Echarri O, Rice S, Wadley G, Lederman R, D'Alfonso S, Russon P, et al. A next-generation social media-based relapse prevention intervention for youth depression: qualitative data on user experience outcomes for social networking, safety, and clinical benefit. Internet Interv 2017 Sep;9:65-73 [FREE Full text] [doi: 10.1016/j.invent.2017.06.002] [Medline: 30135839]
35. Mokkenstorm JK, Mérelle SYM, Smit JH, Beekman ATF, Kerkhof AJFM, Huisman A, et al. Exploration of benefits and potential harmful effects of an online forum for visitors to the suicide prevention platform in the netherlands. Crisis 2020 May;41(3):205-213. [doi: 10.1027/0227-5910/a000627] [Medline: 31657643]
36. Robinson J, Cox G, Bailey E, Hetrick S, Rodrigues M, Fisher S, et al. Social media and suicide prevention: a systematic review. Early Interv Psychiatry 2016 Apr;10(2):103-121. [doi: 10.1111/eip.12229] [Medline: 25702826]
37. Barak A. Emotional support and suicide prevention through the internet: a field project report. Comp Hum Beh 2007 Mar;23(2):971-984. [doi: 10.1016/j.chb.2005.08.001]
38. Rice S, O'Bree B, Wilson M, McEnery C, Lim M, Hamilton M, et al. Leveraging the social network for treatment of social anxiety: pilot study of a youth-specific digital intervention with a focus on engagement of young men. Internet Interv 2020 Apr;20:- [FREE Full text] [doi: 10.1016/j.invent.2020.100323] [Medline: 32435600]
39. Han JY, Hou J, Kim E, Gustafson DH. Lurking as an active participation process: a longitudinal investigation of engagement with an online cancer support group. Health Commun 2014;29(9):911-923 [FREE Full text] [doi: 10.1080/10410236.2013.816911] [Medline: 24345206]
40. Steadman J, Pretorius C. The impact of an online Facebook support group for people with multiple sclerosis on non-active users. Afr J Disabil 2014;3(1):132 [FREE Full text] [doi: 10.4102/ajod.v3i1.132] [Medline: 28730005]
41. Seabrook EM, Kern ML, Rickard NS. Social networking sites, depression, and anxiety: a systematic review. JMIR Ment Health 2016 Nov 23;3(4):e50 [FREE Full text] [doi: 10.2196/mental.5842] [Medline: 27881357]
42. Shaw AM, Timpano KR, Tran TB, Joormann J. Correlates of Facebook usage patterns: the relationship between passive Facebook use, social anxiety symptoms, and brooding. Comp Hum Beh 2015 Jul;48:575-580 [FREE Full text] [doi: 10.1016/j.chb.2015.02.003]
43. Escobar-Viera CG, Shensa A, Bowman ND, Sidani JE, Knight J, James AE, et al. Passive and active social media use and depressive symptoms among united states adults. Cyberpsychol Behav Soc Netw 2018 Jul;21(7):437-443. [doi: 10.1089/cyber.2017.0668] [Medline: 29995530]

44. Thorisdottir IE, Sigurvinsdottir R, Asgeirsdottir BB, Allegrante JP, Sigfusdottir ID. Active and passive social media use and symptoms of anxiety and depressed mood among icelandic adolescents. Cyberpsychol Behav Soc Netw 2019 Aug;22(8):535-542. [doi: 10.1089/cyber.2019.0079] [Medline: 31361508]
45. Gibson K, Wilson J, Grice JL, Seymour F. Resisting the silence: the impact of digital communication on young people's talk about suicide. Youth & Society 2017 Jul 17;51(8):1011-1030. [doi: 10.1177/0044118x17720986]
46. Pirkis J, Rossetto A, Nicholas A, Ftanou M. Advancing knowledge about suicide prevention media campaigns. Crisis 2016 Sep;37(5):319-322. [doi: 10.1027/0227-5910/a000441] [Medline: 27868447]
47. Pettit JW, Temple SR, Norton PJ, Yaroslavsky I, Grover KE, Morgan ST, et al. Thought suppression and suicidal ideation: preliminary evidence in support of a robust association. Depress Anxiety 2009;26(8):758-763. [doi: 10.1002/da.20512] [Medline: 18781662]
48. Singleton A, Abeles P, Smith I. Online social networking and psychological experiences: the perceptions of young people with mental health difficulties. Comp Hum Beh 2016 Aug;61:394-403 [FREE Full text] [doi: 10.1016/j.chb.2016.03.011]
49. Rideout V, Robb MB. Social media, social life: teens reveal their experiences. 2018. URL: https://www.commonsensemedia.org/sites/default/files/uploads/research/2018_cs_socialmediasociallife_executivesummary-final-release_3_lowres.pdf [accessed 2021-02-18]
50. Bailey E, Mühlmann C, Rice S, Nedeljkovic M, Alvarez-Jimenez M, Sander L, et al. Ethical issues and practical barriers in internet-based suicide prevention research: a review and investigator survey. BMC Med Ethics 2020 May 13;21(1):37 [FREE Full text] [doi: 10.1186/s12910-020-00479-1] [Medline: 32404098]
51. Robinson J, Hill N, Thorn P, Teh Z, Battersby R, Reavley N. A young person's guide for communicating safely online about suicide. In: #chatsafe. Australia: Orygen; 2018:1-34.

Abbreviations

MOST: Moderated Online Social Therapy

Edited by R Kukafka; submitted 11.09.20; peer-reviewed by D Reidenberg, J Han, B Cliffe; comments to author 09.11.20; revised version received 26.11.20; accepted 08.02.21; published 05.04.21.

Please cite as:
Bailey E, Robinson J, Alvarez-Jimenez M, Nedeljkovic M, Valentine L, Bendall S, D'Alfonso S, Gilbertson T, McKechnie B, Rice S
Moderated Online Social Therapy for Young People With Active Suicidal Ideation: Qualitative Study
J Med Internet Res 2021;23(4):e24260
URL: https://www.jmir.org/2021/4/e24260
doi:10.2196/24260
PMID:33818392

Original Paper

Internet-Based Cognitive Behavioral Therapy for Informal Caregivers: Randomized Controlled Pilot Trial

Ieva Biliunaite[1], MSc; Evaldas Kazlauskas[2], Prof Dr; Robbert Sanderman[3,4], Prof Dr; Inga Truskauskaite-Kuneviciene[2], PhD; Austeja Dumarkaite[2], MSc; Gerhard Andersson[1,5,6], Prof Dr

[1]Department of Behavioural Sciences and Learning, Linköping University, Linköping, Sweden

[2]Center for Psychotraumatology, Institute of Psychology, Vilnius University, Vilnius, Lithuania

[3]Department of Health Psychology, University Medical Center Groningen, University of Groningen, Groningen, Netherlands

[4]Department of Psychology, Health & Technology, University of Twente, Enschede, Netherlands

[5]Department of Biomedical and Clinical Sciences, Linköping University, Linköping, Sweden

[6]Department of Clinical Neuroscience, Karolinska Institutet, Stockholm, Sweden

Corresponding Author:
Ieva Biliunaite, MSc
Department of Behavioural Sciences and Learning
Linköping University
Campus Valla
Linköping, 581 83
Sweden
Phone: 46 13 28 69 10
Email: ieva.biliunaite@liu.se

Abstract

Background: Caregiving for a family member can result in reduced well-being for the caregiver. Internet-delivered cognitive behavioral therapy (ICBT) may be one way to support this population. This is especially the case for caregivers in countries with limited resources, but high demand for psychological services.

Objective: In this study we evaluated the effects of a therapist-guided 8-week-long ICBT intervention for informal caregivers.

Methods: In total, 63 participants were recruited online and randomized either to the intervention or to the wait-list control group. The main study outcome was the Caregiver Burden Inventory (CBI). Secondary outcomes included measures of caregiver depression, anxiety, stress, and quality of life.

Results: Moderate between-group effect sizes were observed for the CBI measure, in favor of the intervention group, with a Cohen *d*=–0.70 for the intention-to-treat analysis. Analyses of the subscales of the CBI showed significant reductions on the subscales of Development and Physical Health. Moderate reductions were found for depression and anxiety scores as indicated by the Patient Health Questionnaire-9 (PHQ-9) and Generalized Anxiety Disorder-7 (GAD-7) scores. Large between-group effects were observed for reduction in stress and increase in quality of life as indicated by the Perceived Stress Scale-14 (PSS-14), The Brunnsviken Brief Quality of Life Scale (BBQ), and The World Health Organization-Five Well-Being Index (WHO-5). In addition, participants experienced little to no difficulty in using the program and were mostly satisfied with the intervention's platform and the choice of content.

Conclusions: This is the first internet intervention study for informal caregivers in Lithuania. The results suggest that therapist-guided ICBT can be effective in reducing caregiver burden, anxiety, depression, stress, and improving quality of life.

Trial Registration: ClinicalTrials.gov NCT04052724; https://clinicaltrials.gov/ct2/show/NCT04052724

(J Med Internet Res 2021;23(4):e21466) doi:10.2196/21466

KEYWORDS

caregiver burden; informal caregivers; internet intervention; cognitive behavioral therapy; eHealth; mHealth

Introduction

Because of an increase in longevity and a decrease in fertility, it has been suggested that in the future more individuals will be involved in taking care of the elderly than raising children [1]. Approximately one-third of the population could be described as informal caregivers as they look after someone in their close environment: individual(s) who are either physically or mentally ill, or experience difficulties due to old age [2]. Over the course of the history, women were more often expected to take up caregiving tasks within the family context [3]. Today, caregiving is still more prevalent among women [2]. Across Europe, informal caregiving is also more common among those aged between 50 and 59 as well as among non-employed or religious individuals [2]. Informal caregivers are becoming increasingly relied on for managing health care and societal costs for the elderly and chronically ill [4]. Yet, caregivers themselves are often faced with many challenges which puts them at risk for their own well-being.

Caregiving tasks and intensity vary greatly based on the condition of care recipients, available resources, and a combination of many other factors. Accordingly, caregiving might affect individuals differently too. For example, it has been shown that informal caregiving can bring satisfaction and fulfilment [5], especially if appraised positively [6]. However, this is only one side of the coin. It has been previously found that caregivers, in comparison to noncarers, experience reduced mental well-being [7], chronic stress, and increase in depressive symptoms [8]. Such a combination of physical, psychological, financial, and social demands of caregiving can be overall referred to as caregiver burden [3]. According to this definition, caregiver burden is viewed as a multidimensional experience that affects several aspects of caregiver's life. Hence, despite the positive consequences, informal caregiving can also have a negative effect on the informal caregiver's physical and emotional well-being.

Intercultural differences exist in the type of support that informal caregivers receive across Europe [9]. For example, caregivers in the Northern countries, such as Sweden or Denmark, tend to receive tax-funded professional help, and therefore spend on average less time for caregiving duties (approximately 2-3 hours per week) [10]. Contrary to this, caregivers in Southern European countries might be required to provide 24-hour support [10]. These are rather drastic cross-country differences that might in turn have a different effect on caregiver's overall well-being and experience of burden. Apart from availability of formal support, countries differ in other factors such as traditions and societal expectations toward care provision [3]. It is therefore important to consider specific cultural context when developing support services, and in the present context psychological services delivered via the internet.

Because of the increasing numbers of informal caregivers and evidence of their experienced burden, recent years have seen a rise in research studies investigating ways to improve their well-being. Different types of support interventions could be distinguished: interventions providing education and information, interventions providing practical support via respite services for the caregiver, and interventions providing psychological support [3]. Interventions in the latter group are of particular interest because they target informal caregiver well-being. Specifically, these interventions aim at supporting caregivers in their emotion management as well as problem-solving skills, and hence improving their overall quality of life.

Apart from the traditional, face-to-face format, psychological support interventions are being increasingly offered online. One of the benefits of internet interventions is that it can reach a wide range of individuals, even in remote places [11]. This is a very important benefit as opposed to the face-to-face format, especially in the context of the current COVID-19 pandemic, which is having an impact not only on individual mobility, but also on accessibility of various support services. Regarding internet intervention programs for informal caregivers, differences can be observed in both the mode of delivery and the content of programs [12]. To give an example, some programs provide video materials, whereas others mainly rely on text. In addition, some programs offer human support via professional feedback, whereas others offer opportunities for peer support. Despite differing formats of such interventions, based on the current findings there is some evidence that internet interventions for informal caregivers can be effective in improving their well-being [13]. At the same time, despite the emergence of internet intervention studies for informal caregivers, further high-quality research is encouraged [14].

Internet-delivered cognitive behavioral therapy (ICBT) is one of the treatment formats that now has been tested in various populations with success rate similar to that of face-to-face therapy [15]. Few attempts have been previously made to implement ICBT for informal caregivers (eg, [16]). However, in most cases treatments were targeted to a specific group of caregivers. For example, caregivers of individuals with dementia, cancer, or other disorders. Hence, there is little existing knowledge about the effectiveness of transdiagnostic ICBT for informal caregivers. Transdiagnostic treatments target common mechanisms observed in various psychiatric disorders, instead of focusing on one specific disorder. Such treatments are especially useful in addressing comorbidity. To give an example, the same intervention program could be used by individuals with stress and anxiety as well as depression symptoms. Therefore, if effective, transdiagnostic ICBT could be applied for a wide range of informal caregivers.

Considering the need for further research studies, the growing number of caregivers, and the need for accessible interventions, the aim of this study was to develop and evaluate the effectiveness of an internet-based, therapist-guided ICBT intervention in reducing caregiver burden. Caregiver burden was chosen as a focus for the intervention because current research findings regarding the effectiveness of internet intervention in reducing caregiver burden are inconclusive [6]. The intervention was targeted at informal caregivers in Lithuania and was a set up as a pilot randomized controlled trial. A decrease in the caregiver burden was defined as the primary goal with reduction in depression, anxiety, and stress as well as increase in quality of life as secondary goals. Results of this study will provide information about the acceptability and

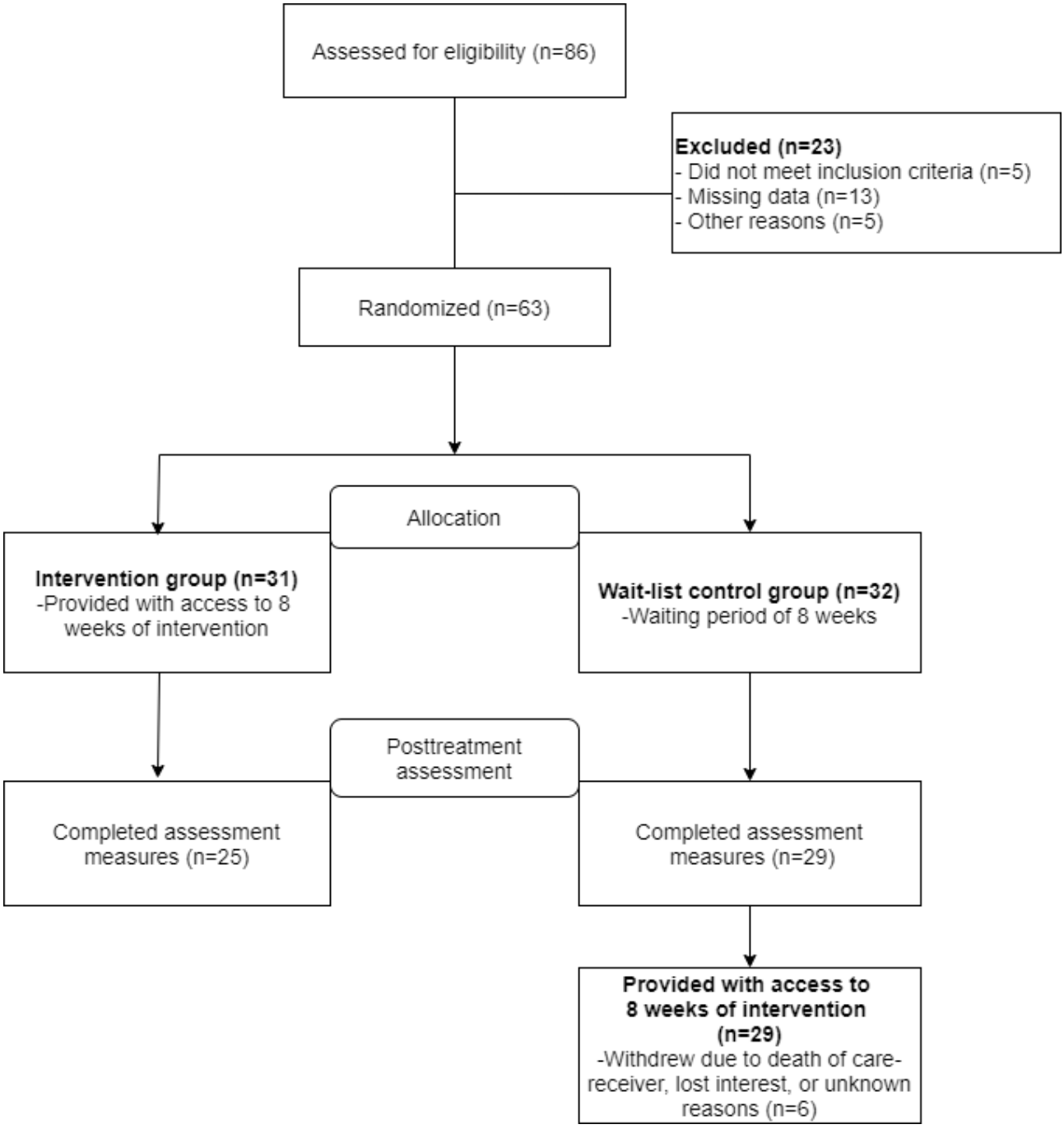
effectiveness of such interventions in a unique population where the demand for such services is high [17].

Methods

Design

A 2-armed randomized controlled trial was conducted online in Lithuania with participants recruited from the general population. Participants were randomly allocated to either an 8-week internet treatment or a wait-list control condition. Measures were collected online and administered at 2 points in time: at the start of the treatment (pretreatment) and at the end of the treatment (posttreatment). In addition, 4 weeks after the start of the intervention, participants were contacted for a short telephone interview. The main purpose of this conversation was to provide participants with an opportunity to ask questions about the use of the treatment as well as to raise any related concerns. Control group participants received the same treatment once the initial treatment phase had finished (posttreatment). A flowchart representation of the study is presented in Figure 1. Ethical approval for this study was received from the Vilnius University Psychology Research Ethics Committee (08-07-2019 No. 26). The trial was registered in ClinicalTrials.gov (registration number NCT04052724) and reported in accordance with the CONSORT-EHEALTH checklist (Multimedia Appendix 1).

Figure 1. Flowchart of the participant recruitment to the randomized controlled pilot trial.



Participants

Eligible participants were able to register from September 2019 to October 2019. Recruitment was conducted in various forms. The study was announced by the Vilnius University press release and posted on the Vilnius University website, Facebook, and other social media. Additionally, invitation to participate in the study was disseminated among the health care institutions, social services, and patient organizations targeted to reach informal caregivers. The study publicity campaign reached national media, and the study was covered by national TV and radio programs. In addition, information about the study was posted in national and regional press. After registration, individuals who filled in all the required measures online were contacted for a structured interview over telephone, after which eligibility to participate in the study was finalized.

To participate in the study, individuals had to have access to the internet via a computer or any other compatible device (eg, smartphone or tablet) as well as the ability to use it throughout the duration of the intervention. Eligible participants had to have a score of 24 or more on the Caregiver Burden Inventory (CBI) [18], be 18 years old or over, and be fluent in comprehending, writing, and reading Lithuanian language. The intervention was not tailored to any specific group of caregivers (eg, those caring for patients with dementia). Individuals were not eligible to participate in the study if they were experiencing severe physical or mental health problems, alcohol addiction, other severe traumatic events, suicide risk, severe interpersonal violence, or their care receiver had a life expectancy below or approximately around 6 months.

Overall, 86 individuals expressed interest to participate and provided informed written consent. Of these, 13 did not fill in all the required questionnaires. One participant was excluded as we were not able to reach the person for screening. Eligibility of the remaining 72 participants was assessed via phone interviews. Mini International Neuropsychiatric Interview Version 5 [19] was translated into Lithuanian language by the research group and was used in phone interviews for additional assessment of inclusion criteria. Following interviews, 9 participants were excluded due to either not meeting inclusion criteria (n=5) or losing interest in participation (n=3). One remaining application was removed, because the participant had registered twice. We have made an exception regarding cut-off scores for CBI: 1 participant with a score of 21 was included in the study because the individual expressed feeling burdened as well as wished for support during the phone interview. Hence, 63 participants were randomized to either the intervention (n=31) or the wait-list control group (n=32). Randomization was conducted by an independent researcher not involved in the trial and performed according to a 1:1 ratio. The website for generation of random numbers [20] was used. Randomization procedure was also used for randomly allocating therapists to participants.

All the included participants were requested about the condition of their care receiver. As there would be several small groups among our sample regarding different medical conditions of the care receivers, we chose not to further categorize these data. Moreover, many caregivers indicated that their care receiver experiences several comorbidities, which would make it difficult to categorize the data. Nevertheless, some of the more common health conditions of the care receivers were dementia and frailty due to old age.

Measures

Primary Outcome Measure

Caregiver Burden Inventory

The CBI [18] was chosen as it views caregiver burden as a multidimensional experience. The CBI comprises 24 items that are distributed within 5 areas: development, physical health, time dependency, emotional health, and social relationships. Response options are presented on a 5-item Likert scale ranging from 0 (Never) to 4 (Nearly always). All areas have 5 questions dedicated to them with the exception of physical health, which has 4. The total score of the CBI is summed up and ranges from 0 to 96, with higher scores indicating higher levels of burden. Reliability coefficients for each of the subscales of the CBI were previously shown to be high: development (α=.87), physical health (α=.86), time dependency (α=.85), emotional health (α=.81), and social relationships (α=.69) [21]. In this sample, Cronbach α for each of the subscales was high: development (α=.85), physical health (α=.80), time dependency (α=.89), emotional health (α=.83), and social relationships (α=.81). Cronbach α for CBI altogether in this sample was also high (α=.87). The CBI has been translated and previously applied in research studies in several different countries (eg, [22,23]) as well as for specific caregiver groups. In our study, the CBI was translated into Lithuanian language by the research group.

Secondary Outcome Measures

For these measures existing official Lithuanian language translations were used.

Patient Health Questionnaire-9

The Patient Health Questionnaire-9 (PHQ-9) [24] comprises 9 questions aimed at evaluating depressive symptoms. Response options are presented on a 4-item Likert scale ranging from 0 (Not at all) to 3 (Nearly every day). Higher scores indicate higher symptom severity. This questionnaire is widely used in research studies, due to easy administration and very good psychometric properties (α=.89) [24]. In this sample, Cronbach α for PHQ-9 was high (α=.82).

Generalized Anxiety Disorder-7

The Generalized Anxiety Disorder-7 (GAD-7) questionnaire [25] is used for evaluating symptoms of generalized anxiety disorder. This questionnaire consists of 7 questions. As with PHQ-9, answers for this questionnaire are distributed on a 4-item Likert scale ranging from 0 (Not at all) to 3 (Nearly every day). Higher score indicates higher levels of anxiety. This measure displays very good psychometric properties (α=.92) [25]. In this sample, Cronbach α for GAD-7 was high (α=.87).

Perceived Stress Scale-14

The Perceived Stress Scale-14 (PSS-14) [26] is a 14-item questionnaire that contains questions measuring stress on a 5-point Likert scale ranging from 0 (‘Never’) to 4 (‘Very Often’) [26]. Items on the PSS-14 touch upon aspects such as feeling

of nervousness and ability to cope as experienced over the last 4 weeks. Higher scores on PSS-14 indicate higher levels of stress. This measure has previously shown good psychometric properties with Cronbach α ranging from .75 to .89 [27]. In this sample, Cronbach α was found to be high (α=.85).

Brunnsviken Brief Quality of Life Scale (BBQ)

The Brunnsviken Brief Quality of Life Scale (BBQ) [28] is a questionnaire developed for evaluating quality of life for both clinical and nonclinical samples. It contains 12 statements that cover topics such as creativity, leisure time, and view on oneself. Responses are distributed on a 5-point Likert scale ranging from 0 (Strongly disagree) to 4 (Strongly agree). Higher scores on this measure indicate higher quality of life. Psychometric properties of BBQ are good (α=.76) [28]. The BBQ in our sample showed good internal consistency (α=.87).

The World Health Organization-Five Well-Being Index

The World Health Organization-Five Well-Being Index (WHO-5) [29] is a widely used short questionnaire that was developed by the World Health Organization [30] and has shown to be a valid and reliable measure of overall well-being (α=.88) [29]. This questionnaire contains 5 statements regarding an individual's well-being over the last 2 weeks. Each of the statements is evaluated using a 6-item Likert scale ranging from 0 (At no time) to 5 (All the time). Higher scores indicate higher well-being. The internal consistency for WHO-5 in our sample was acceptable (α=.76).

Baseline Measures

Life Events Checklist

The Life Events Checklist (LEC) [31] was developed at the National Center for Post-Traumatic Stress Disorder to screen for potentially traumatic events in a respondent's lifetime. In this study, we used 19-item version that has been adapted to the Lithuanian cultural context [32]. This questionnaire has acceptable psychometric properties (κ=0.61) [33]. In our sample, the internal consistency for LEC was acceptable (α=.78).

The Alcohol Use Disorders Identification Test (AUDIT)

The Alcohol Use Disorders Identification Test (AUDIT) [34] is a 10-item instrument for assessing alcohol consumption and indicating its consequences on the well-being of an individual. Each statement is evaluated by choosing an answer from either 5 or 3 different response options. This instrument has been a subject of extensive evaluation and has consistently displayed good psychometric properties with a median reliability value of 0.82 for non-English versions [35]. The AUDIT in our sample displayed acceptable internal consistency (α=.73).

Intervention

The intervention was an ICBT program consisting of 8 modules, each dedicated to 1 theme. The themes were (chronologically): introduction, thoughts, stress, and relaxation, problem solving, communication, anxiety, behavioral activation, and maintenance. The content was mainly presented in the form of written text with the exception of relaxation instructions, which participants could choose to either read or listen to. See Figure 2 for a screenshot of the 8 modules in the program provided to participants after logging in. In Lithuanian language, the name of the program was *Slaugau artimą*, which if translated literally means, "I take care of my close one." The intervention content was partly adapted from materials previously used in other ICBT studies (eg, [36-38]). The choice of content, examples, and exercises of the intervention were carefully selected and adopted for the target population grounded on the practices and guidelines in the field of informal caregiving (eg, [3]). A short description of the 8 modules as well as a list of exercises for each module is presented in Table 1.

Figure 2. Screenshot of the eight intervention modules as presented to participants after logging in.

Participants were provided access to 1 new module every week over the 8 weeks of intervention, starting from module 1. Once accessible, modules remained available throughout the duration of the treatment. The basic structure of the modules is very similar throughout the intervention: each module starts with psychoeducation, followed by one or few case examples which are then followed by several exercises. At the end of each module, participants were provided with suggestions about which exercises they should practice further that week. In this way participants were encouraged to learn and apply gained knowledge in managing maladaptive thoughts or behaviors in their own situation. In cases of lack of clarity, participants were able to contact their assigned therapist via a secure internal messaging function on the platform. Therapists monitored participants' questions and provided feedback on completed exercises daily via the intervention's platform.

Table 1. Description of intervention content.

Module (length)	Goals of the module	Exercises
Introduction (1089 words)	• Instruct about how to use the intervention • Introduce main principles of cognitive behavioral therapy • Inform about what participants can expect throughout the treatment • Encourage to formulate individual goals for the duration of the program	• Writing exercise I—Describing emotionally difficult caregiving-related situation and further reflecting on thoughts, behavior, and short- and long-term consequences. • Writing exercise II—Setting goals for the treatment
Thoughts (1682 words)	• Discuss how thoughts impact on one's well-being • Introduce the concept of automatic thoughts • Instruct about how to practice recognizing automatic thoughts • Introduce the most common cognitive distortions	• Thinking of own thought patterns and cognitive distortions • Thought Change Record—participants use a predefined structure to evaluate their situation, automatic thoughts, emotions, rational response, and outcome
Stress and relaxation (2426 words)	• Define stress • Discuss what effect stress can have on well-being • Discuss how one can reduce stress in daily life • Familiarize with relaxation methods	• Writing exercise—Thinking and describing which caregiving-related situation(s) causes most stress • Thinking of ways to increase inner resources and lower mental or physical load • Planning when to engage in pleasurable activities • Listening and practicing short meditations
Problem solving (1524 words)	• Encourage to reflect on own problem-solving strategies • Familiarize with different coping strategies • Emphasize helpful and harmful coping strategies • Learn about factors that interfere with problem solving	• Writing exercise—Thinking and describing strategies that an individual uses to deal with problems • Based on provided theory and examples, reflecting once more on problem-solving strategies • Solving own problem(s) a using step-by-step approach
Communication (2635 words)	• Discuss possible changes in communication after one becomes a caregiver • Discuss possible ways of maintaining or improving communication with people in a close environment	• Reflecting on changes in communication after becoming a caregiver • Step-by-step structure from previous module for solving communication problems • Perspective taking exercise—Evaluating difficulties in communication from a different point of view • Writing exercise—Using writing to express own emotions
Anxiety (2090 words)	• Discuss what anxiety is and what effect it can have on well-being • Discuss ways in which anxiety could be reduced	• Reflecting on sources of worry in own situation • Exercise "Balcony"—Evaluating a worrisome situation by observing it from the distance • Listening and practicing short meditations
Behavioral activation (1338 words)	• Emphasize the importance of personal time and small rewards • Encourage to spend time for own needs	• Reflecting on behavior and activities that are not useful • Thinking of how one can find more time and opportunities for pleasant activities
Maintenance (906 words)	• Summarize most important points • Provide with tips for maintaining gained knowledge • Encourage to reflect on previously formed goals and goals for the future • Farewell	• Reflecting on what was the most useful throughout the program and which knowledge or exercises one can also apply in the future • Reflecting on initial goal and progress throughout the program

Therapists

Three Master's in Clinical Psychology program students trained to deliver this intervention under supervision and 3 clinical psychologists were involved as therapists in the study. The main role of the therapists was to answer participant questions as well as to provide feedback on completed modules via a secure messaging system in the intervention's platform [39]. Weekly scheduled and on-demand supervision meetings for therapists were conducted and lead by one of the experienced clinical psychologists (AD) in the study. These supervision meetings varied in length between 1 and 2 hours.

Procedure

Upon opening the study website, participants were provided with information regarding inclusion and exclusion criteria as

well as the registration process, management, and research team. After participants provided informed consent on a secure study website, they were asked to fill in the screening questionnaires. Participants were then invited for a phone interview. Following phone interviews, the final decision about participation in the study was jointly discussed and agreed upon by the 3 coauthors (IB, IT-K, and AD). Decision about exclusion or inclusion in the study was communicated to the participants in a span of a few days. After randomization, participants in the intervention and wait-list control group were provided with information about the start of the intervention. Participants in the waiting-list control group were also explained that they will be able to receive access to the same treatment after the intervention group is finished.

A secure online platform (*Iterapi* [40]) was used for communication between therapists and participants, distribution of program materials, and collection of assessments. Participants' personal information was anonymized by assigning each participant a code, which participants then used for logging into the program. Included participants were also able to extend their data security by receiving a code (one-time password) to their phone number, which had to be entered upon every logging in, after entering their self-generated password.

At the start of the intervention, all participants received an email containing their username as well as personalized link for creating their own password. Throughout the duration of the program, every Thursday, participants received an email stating availability of the new material. Participants who did not view a particular week's materials or have not conducted exercises were sent 1 weekly reminder. Reminder contained a predefined short encouraging message and was sent on Mondays, by therapists to participants in own groups.

Statistical Analysis

Analysis was performed using IBM SPSS Statistics (version 25). The significance level was set at .05. Independent samples *t* test and Fisher exact test were used for investigating possible differences among participants in the 2 groups at baseline. Data collected at posttreatment were treated according to the intention-to-treat (ITT) principle. Following ITT all participants were analyzed according to their treatment assignment included in the analysis [41]. This method was chosen as it preserves the integrity of the randomization and minimizes the risks of bias that could occur due to the differences in groups following attrition or nonadherence [42]. For this purpose, a multiple imputation procedure was chosen considering that data were missing at random (MAR) [43]. According to MAR, missingness depends on the observed data, where incomplete values are replaced by values that are based on the complete data [44]. In our study MAR was considered, as no clear patterns in the missing data regarding pretreatment scores or demographic variables were identified. During the multiple imputation procedure, 20 simulations were imputed in a sequence. Because it is practically difficult to be fully confident that data are missing at random, complete case analyses were also performed.

Descriptive statistics were used for evaluating attrition and overall satisfaction with the program. Regarding main outcomes, effect sizes and confidence intervals were calculated for within and between groups. Analysis of covariance (ANCOVA) was conducted to investigate treatment effects at posttreatment for primary and all secondary measures with baseline scores entered as covariates [45].

The Jacobson and Truax [46] method was used for calculating Reliable Change Index (RCI) and investigating clinical significance of change in primary outcome at posttreatment. Because population norms were not available, criteria *a* was calculated, resulting in a cut-off point of 28.8. According to this, individuals who scored below 28.8 points at posttreatment should fall outside of the dysfunctional population and be considered recovered [46]. As for reliability of the scores, RCI values were calculated. In cases where a positive reliable change was achieved, participants were deemed improved. In turn, negative reliable change scores were used to determine deterioration. In the current sample, participants had to obtain a reduction of 8.8 scores on CBI at postassessment for a positive reliable change to be achieved.

Results

Participants

Most of the recruited participants were female (57/63, 90%) caring for either their elderly mother or father (44/63, 70%). Most participants had provided care for either 1-4 years (28/63, 44%) or more than 4 years (25/63, 40%). Most caregivers were spending 5-7 days per week (54/63, 86%), and either 3-7 (24/63, 38%) or 12 or more hours (23/63, 37%) per day providing care. In this respect, the sample could be described as consisting of high-intensity, long-term carers. Demographic characteristics of the sample are presented in Table 2. No pretreatment differences regarding demographic or outcome measures between participants randomized to treatment and wait-list control group were detected.

Table 2. Sociodemographic characteristics of the sample at baseline.

Participant characteristics	Overall (n=63)	Intervention group (n=31)	Wait-list control group (n=32)
Age caregiver (year), mean (SD)	52 (8.4)	54 (7.9)	50 (8.57)
Age recipient (year), mean (SD)	71 (21.1)	70 (23.13)	72 (19.28)
Gender (female) caregiver, n (%)	57 (90)	28 (90)	29 (91)
Gender (female) recipient, n (%)	44 (70)	20 (65)	24 (75)
Relation receiver, n (%)			
Husband/wife/partner	8 (13)	5 (16)	3 (9)
Father/mother	44 (70)	20 (65)	24 (75)
Other	11 (17)	6 (19)	5 (16)
Time caring (months), n (%)			
<12	10 (16)	6 (19)	4 (13)
12-48	28 (44)	13 (42)	15 (47)
>48	25 (40)	12 (39)	13 (41)
Time week (days), n (%)			
1-2	3 (5)	1 (3)	2 (6)
3-4	6 (10)	3 (10)	3 (9)
5-7	54 (86)	27 (87)	27 (84)
Time day (hours), n (%)			
3<	9 (14)	5 (16)	4 (13)
3-7	24 (38)	14 (45)	10 (31)
8-11	7 (11)	5 (16)	2 (6)
>12	23 (37)	7 (23)	16 (50)
Residing with care receiver (yes), n (%)	49 (78)	24 (77)	25 (78)
Individual is the only caregiver (yes), n (%)	31 (49)	18 (58)	13 (41)
Highest education level, n (%)			
High school or lower	4 (6)	3 (10)	1 (3)
Professional/vocational training	17 (27)	9 (29)	8 (25)
College or applied science education	6 (10)	2 (6)	4 (13)
University degree	36 (57)	17 (55)	19 (59)
Marital status, n (%)			
Single	12 (19)	7 (23)	5 (16)
Married/partner	39 (62)	18 (58)	21 (66)
Divorced/widowed or other	12 (19)	6 (19)	6 (19)

Attrition

Participants were regarded as dropouts if the posttreatment measures were missing. Independent samples *t* test and Fisher exact test were performed for investigating differences between dropouts and completers. No differences were detected between the groups regrading demographic characteristics or scores on outcome measures.

At posttreatment, out of 31 participants in the intervention group, 25 have filled in the measures (81%), yielding a dropout rate of approximately 19%. At the same time, 29 out of 32 participants have filled in the measures in the wait-list control group (91%). Overall, posttreatment measures were collected from 54 participants (86%), which indicated a dropout rate of 14%. Following the intervention, out of 32 participants in the wait-list control group, 23 had filled in the measures (72%), showing a dropout rate of approximately 28%. In total, the overall dropout rate (ie, both groups combined) after the intervention period ended was 24% (15/63).

We were not able to reach all the participants who did not fill in the posttreatment measures. From participants that we did manage to reach, there were 2 main reasons for ceased participation: losing interest in participation (n=3) or death of care receiver (n=3).

Main Outcomes

Within- and between-group effect sizes were calculated for both completers and ITT sample. We found moderate within- and between-group effect sizes for the CBI measure (Table 3). Besides, moderate to large between- and small to large within-group effect sizes for secondary outcome measures were observed. Effect sizes for the completers were higher than those for ITT, indicating that the ITT result should be the one to consider. Further, only the ITT results will be reported. Completer results are provided in Multimedia Appendix 2.

The ANCOVAs were performed to determine differences between the control and intervention group on the posttreatment scores when controlling for the pretreatment scores. For primary and all secondary measures, significant effects of group on posttreatment scores were found in the ITT sample—CBI: $F_{1,60}$=5.39, P=.02; PHQ-9: $F_{1,60}$=6.12, P=.01; GAD-7: $F_{1,60}$=8.24, P=.004; PSS-14: $F_{1,60}$=13.56, P<.001; BBQ: $F_{1,60}$=10.88, P=.001; and WHO-5: $F_{1,60}$=10.7, P=.001. The ANCOVAs were also conducted to investigate changes over 5 subscales of the CBI separately at posttreatment. In the ITT sample significant changes have been detected for the reduction in postassessment scores in the subscales of Development and Physical Health, $F_{1,60}$=6.99, P=.008 and $F_{1,60}$=5.5, P=.02, respectively. No significant changes were observed for the remaining 3 subscales—Time Dependency: $F_{1,60}$=0.25, P=.62; Emotional Health: $F_{1,60}$=2.05, P=.15; and Social Relationships: $F_{1,60}$=2.67, P=.10.

Table 3. Means, standard deviations, and effect sizes (Cohen *d*) with confidence intervals.

Measures and condition	Intention to treat, mean (SD)		Completers, mean (SD)		Effect size (95% CI)	
	Pretreatment (n=63)	Posttreatment (n=63)	Pretreatment (n=63)	Posttreatment (n=54)	Within-group prepost	Between-group post-treatment
CBI[a]						
Intervention	51.94 (12.79)	46.39 (13.78)	51.94 (12.79)	46.28 (13.92)	0.41 (–0.09 to 0.92)	–0.7 (–1.2 to –0.19)
Wait-list	55.84 (12.44)	56.99 (16.22)	55.84 (12.44)	57.34 (16.1)	0.13 (–0.36 to 0.62)	
PHQ-9[b]						
Intervention	9.32 (4.42)	8.23 (4.79)	9.32 (4.42)	7.84 (4.78)	0.24 (–0.27 to 0.73)	–0.69 (–1.19 to –0.17)
Wait-list	10.41 (5.36)	12.05 (6.2)	10.41 (5,36)	12.14 (6.28)	–0.28 (–0.77 to 0.21)	
GAD-7[c]						
Intervention	8.19 (4.34)	6.83 (4.28)	8.19 (4.34)	6.48 (4.12)	0.32 (–0.19 to 0.81)	–0.74 (–1.24 to –0.22)
Wait-list	9 (4.95)	10.77 (6.14)	9 (4.95)	10.9 (6.25)	–0.32 (–0.81 to 0.18)	
PSS-14[d]						
Intervention	27.03 (5.97)	21.83 (6.84)	27.03 (5.97)	21.08 (6.21)	0.81 (0.28 to 1.32)	–1.06 (–1.57 to –0.52)
Wait-list	29.28 (7.89)	30.07 (8.58)	29.28 (7.89)	30.69 (8.3)	–0.10 (–0.58 to 0.40)	
BBQ[e]						
Intervention	45.26 (22.32)	56.69 (25.04)	45.26 (22.32)	58.44 (25.18)	–0.48 (–0.98 to 0.03)	0.8 (0.28 to 1.30)
Wait-list	41.75 (23.19)	37.77 (22.18)	41.75 (23.19)	37.14 (21.78)	0.18 (–0.32 to 0.66)	
WHO-5[f]						
Intervention	34.97 (12.04)	48.66 (19.72)	34.97 (12.04)	49.92 (19.8)	–0.8 (–1.3 to –0.3)	0.85 (0.32 to 1.35)
Wait-list	32.75 (16.52)	32.72 (17.84)	32.75 (16.52)	32.55 (17.75)	0 (–0.50 to 0.49)	

[a]CBI: Caregiver Burden Inventory.

[b]PHQ-9: Patient Health Questionnaire-9.

[c]GAD-7: Generalized Anxiety Disorder-7.

[d]PSS-14: Perceived Stress Scale-14.

[e]BBQ: The Brunnsviken Brief Quality of Life Scale.

[f]WHO-5: The World Health Organization-Five Well-Being Index.

Clinically Significant Change and RCI of Caregiver Burden

Results regarding RCI for the ITT sample are presented in Table 4. McNemar test was performed to test whether the difference between participants in the 2 groups regarding positive reliable change is significant. An exact McNemar test result indicated that significantly more participants in the intervention group as opposed to the wait-list control group have achieved a positive reliable change for CBI scores at postassessment (P<.001). Regarding clinical significance, 2 (6%) participants in the intervention group scored below the cut-off score of 28.8. These

participants were deemed as recovered because they achieved a clinically significant change. In the control group, 1 (3%) participant was found to achieve a clinically significant change.

Table 4. RCI of participants in the ITT sample.

Group	RCI[a] for CBI[b] scores		
	Positive RCI	No change	Negative RCI
Intervention (n=31), n (%)	14 (45)	13 (42)	4 (13)
Wait-list (control; n=32), n (%)	3 (9)	23 (72)	6 (19)

[a]RCI: Reliable Change Index.

[b]CBI: Caregiver Burden Inventory.

Participant Evaluation of the Intervention

Overall, 56% (14/25) of participants who filled in posttreatment measures in the intervention group indicated that throughout the program their well-being had improved. Most participants also indicated that using the program was either easy (13/25, 52%) or very easy (6/25, 24%) and was either useful (12/25, 48%) or very useful (7/25, 28%). Participants were also positive about the opportunity to contact and receive feedback from a therapist, with 44% (11/25) indicating that this opportunity was useful, while 28% (7/25) of participants rated it as very useful. Lastly, information and tasks in the program were mostly rated as useful (10/25, 40%) or very useful (9/25, 36%).

Almost half of the participants (12/25, 48%) indicated that all the modules in the program were useful and none (0/25, 0%) noted that modules were not useful. The highest ranking in terms of usefulness was module 7 (“behavioral activation”; 16/25, 64%). The second most useful was module 6 (“anxiety”; 14/25, 56%). The modules “stress and relaxation” and “problem solving” were also rated as very useful by more than half of the participants (13/25, 52%). Module 1 ("introduction") was rated as the least useful (5/25, 20%).

Discussion

Principal Findings

The aim of this study was to evaluate the effectiveness of an internet-based therapist-guided program for reducing caregiver burden. Moderate between- and within-group effect sizes were found in reducing caregiver burden in the intervention group. This is a promising finding keeping in mind that previous face-to-face studies have achieved small to moderate between-group effect sizes (d=0.09-0.23) [47]. Our results also revealed that out of 5 caregiver-burden components, significant posttreatment reductions were observed for the subscales of Physical Health and Development. The latter subscale consists of items regarding missing out on life, emotional tiredness, and limited social life. In turn, the Physical Health subscale questions sleep disturbances, worsened health, and tiredness. It could be argued that behavioral and cognitive components of the intervention, such as time scheduling, thought diary, stress management exercises, and relaxation methods, helped participants to address these areas the most. To our knowledge, there is only 1 previous study that has reported improvement in these subscales of caregiver burden [48]. This is a study by Spatuzzi and colleagues [48], in which caregivers (of patients with cancer) with low spiritual well-being were compared with those with high spiritual well-being, with the latter found to be experiencing significantly less burden in the Development and Physical Health subscales of CBI. In their study, spiritual well-being was described as one of the core domains in quality of life that also helps in dealing with caregiving tasks. Improvements in quality of life ratings in our sample come in hand with this explanation. However, further investigation is warranted before any conclusions can be drawn.

Moderate to high between-group effect sizes were found on measures of anxiety, depression, and stress as well as increase in quality of life scores with only slight differences between ITT and completer analysis. These findings are in line with existing evidence of internet intervention potential in improving caregiver well-being [14]. We consider this finding of great importance because participants in our sample were found to be highly invested in caregiving: caring on average 5-7 days per week, between 3 and 7 or more than 12 hours per day. The latter could also explain the finding that only 2 participants had clinically recovered and slightly less than half had achieved a positive reliable change in relation to their burden. Caregiver burden, as described by Novak and Guest [18], is a complex experience that compromises emotional, physical, developmental, social, and time-dependency factors. Although our intervention provided participants with tools to deal with negative automatic thoughts, manage stress and anxiety, improve communication, and be attentive to their own needs, it could not change caregiving tasks or the amount of time required for supporting the care receiver. Nevertheless, moderate to high effect sizes found for the primary and secondary measures indicate that internet-delivered CBT (ICBT) can be effective in improving caregiver well-being in Lithuania. This argument is further supported by participant evaluations: more than half of the participants indicated that their well-being improved throughout the duration of the program and almost half noted that all modules were useful and the program was easy to use.

Limitations and Strengths

One of the main limitations in this study is that we did not control for external help that caregivers were receiving. Although all participants in the study identified themselves as primary caregivers, they could have been receiving various levels of support from other family members or professional workers such as nurses. A second limitation concerns the implementation of the CBI measure and The Mini International Neuropsychiatric Interview (MINI) that were used for the

screening purposes. This was the first time both instruments were translated and used in Lithuanian language which could bias the validity and reliability of these measures. Importantly, forward and backward translations were performed independently by the 2 bilingual members (IB and AD) of the research team. Independent versions were then compared and finalized following discussion among the research group members. In addition, translations were presented to a small convenience sample for evaluation of the comprehension of the items. By following such a procedure, we wanted to make sure that translated versions are as accurate as possible while culturally appropriate. We have chosen such an approach due to the pilot nature of the study. We aim to further investigate psychometric properties of these instruments before their implementation in subsequent trials.

Few other limitations must be mentioned. One is that all participants in the study were self-referred which limits the generalizability of our findings and presents the risk of the volunteer bias. Another important limitation is lack of information regarding the long-term effects of the intervention. In addition to the postintervention assessment, follow-up with the participants is highly desirable for investigating if the effects of the intervention were maintained long term. Consequently, it is important to mention that we included 1 participant with a score few points below the initial cut-off point for the CBI. One approach could have been to exclude this individual, but taking into account the pilot nature of the study, we have decided to include the participant and to further reflect on this decision when reviewing inclusion criteria for a prospective larger trial in the future. Lastly, even though positive changes in caregiver well-being were identified, at this point it is not possible to determine which components of the intervention were responsible for this improvement. We aim to further investigate this in the future.

Informal caregivers supporting individuals with a range of support needs were included in this study. For example, there were individuals who were caring for a sibling with a mental disorder as well as their own underage children or elderly parents diagnosed with dementia. For this reason, it could be argued that the intervention did not equally suit everyone's need. However, our initial idea was to create an intervention that would be suitable to a wide group of caregivers. In that way, independent of the care receiver's condition, everyone would find something that is applicable to one's own situation. The fact that we managed to detect improvement in caregiver well-being indicates that such transdiagnostic interventions can be effective. As another strength, we would like to outline the relatively low drop-out rates in our study. We consider the therapist support, weekly reminders, and feedback that participants received from the therapist as possibly the most effective factors in keeping participants engaged with the intervention. Lastly, there were no technical problems with the intervention that could have impacted participant's overall experience. A full-time IT technician was responsible for managing the technical aspects of the intervention's platform and thus possible problems would have been handled rapidly.

Clinical Implications

As briefly outlined previously, informal caregivers represent an increasing part of the society. Their task, even though rewarding, is still demanding. As our results indicate, ICBT can be effective in reducing caregiver burden as well as other negative states such as stress. In turn, it can also be effective in improving the overall quality of life. Importantly, it can reach a wide range of individuals, even the ones in remote locations. Internet interventions as opposed to face-to-face interventions are most likely even of a greater importance currently, considering the COVID-19 pandemic. In addition, for many individuals such a solution could be much more affordable not only time wise, but also financially. As for our target group, informal caregivers in Lithuania, such solutions are of even higher importance considering that psychological support services in general are very limited and not always accessible for all [17]. At this point, we must stress the pilot nature of the study and that our results should be interpreted with caution. Yet, we encourage additional research to further build on this idea and extend our findings in larger randomized trials. Because the numbers of informal caregivers is increasing worldwide and is only considered to further grow in the future, researchers are encouraged to further pursue these ideas in diverse cultural backgrounds, possibly with additional emphasis on the countries in which strong familial norms and little formal support are available to support informal caregivers.

Conclusion

This study is the first internet intervention study aimed at informal caregivers in Lithuania. The internet-based therapist-guided intervention based on CBT principles was found to be moderately effective in reducing caregiver burden, anxiety, and depressive symptoms. It was also found to be highly effective in reducing stress and improving quality of life for informal caregivers supporting individuals with various care needs. Such an intervention could be of special importance for caregivers who due to time-constraints, geographical, medical, or other reasons are not able to attend face-to-face therapy. Because of close cultural and historical aspects, this experience could further extend to neighboring Baltic countries in the region such as Latvia, where the need for psychological services for the informal caregivers is also high.

Acknowledgments

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under the Marie Skłodowska-Curie grant agreement number 814072 and is part of The European Training Network on Informal Care (ENTWINE). This work was highly supported by George Vlaescu. We also thank Vaiva Milasiunaite, Julija Mingaudaite, and Irma Sabonyte for their assistance.

Authors' Contributions

IB contributed to the study conception, design, coordination, data collection, data analysis, and drafting of the manuscript. EK, IT-K, and AD and contributed to the study design, coordination, data collection, and data analysis. RS contributed to the design of the study and data analysis. GA contributed to the study conception, design, coordination, data collection, and data analysis. All authors were involved in revising the manuscript. Final approval was obtained from all the authors for the publication of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT-EHEALTH checklist.
[PDF File (Adobe PDF File), 1640 KB - jmir_v23i4e21466_app1.pdf]

Multimedia Appendix 2
Results based on completer sample.
[DOCX File , 17 KB - jmir_v23i4e21466_app2.docx]

References

1. Clancy RL, Fisher GG, Daigle KL, Henle CA, McCarthy J, Fruhauf CA. Eldercare and Work Among Informal Caregivers: A Multidisciplinary Review and Recommendations for Future Research. J Bus Psychol 2019 Jan 9;35(1):9-27. [doi: 10.1007/s10869-018-9612-3]
2. Verbakel E, Tamlagsrønning S, Winstone L, Fjær EL, Eikemo T. Informal care in Europe: findings from the European Social Survey (2014) special module on the social determinants of health. Eur J Public Health 2017 Feb 01;27(suppl_1):90-95. [doi: 10.1093/eurpub/ckw229] [Medline: 28355645]
3. Revenson T, Griva K, Luszczynska A. Caregiving in the illness context. Berlin, Germany: Springer; 2016.
4. Riedel M, Kraus M. Differences and similarities in monetary benefits for informal care in old and new EU member states. International Journal of Social Welfare 2015 Jan 28;25(1):7-17. [doi: 10.1111/ijsw.12157]
5. Cohen CA, Colantonio A, Vernich L. Positive aspects of caregiving: rounding out the caregiver experience. Int J Geriatr Psychiatry 2002 Feb;17(2):184-188. [doi: 10.1002/gps.561] [Medline: 11813283]
6. Maguire R, Hanly P, Maguire P. Beyond care burden: associations between positive psychological appraisals and well-being among informal caregivers in Europe. Qual Life Res 2019 Aug 4;28(8):2135-2146. [doi: 10.1007/s11136-019-02122-y] [Medline: 30715685]
7. Pinquart M, Sörensen S. Differences between caregivers and noncaregivers in psychological health and physical health: a meta-analysis. Psychol Aging 2003 Jun;18(2):250-267. [doi: 10.1037/0882-7974.18.2.250] [Medline: 12825775]
8. Schulz R, Sherwood PR. PHYSICAL AND MENTAL HEALTH EFFECTS OF FAMILY CAREGIVING. Journal of Social Work Education 2017 Jan 12;44(sup3):105-113. [doi: 10.5175/jswe.2008.773247702]
9. Lamura G, Mnich E, Nolan M, Wojszel B, Krevers B, Mestheneos L, EUROFAMCARE Group. Family carers' experiences using support services in Europe: empirical evidence from the EUROFAMCARE study. Gerontologist 2008 Dec;48(6):752-771. [doi: 10.1093/geront/48.6.752] [Medline: 19139249]
10. Brandt M. INTERGENERATIONAL HELP AND PUBLIC ASSISTANCE IN EUROPE. European Societies 2013 Feb;15(1):26-56. [doi: 10.1080/14616696.2012.726733]
11. Andersson G, Titov N, Dear BF, Rozental A, Carlbring P. Internet-delivered psychological treatments: from innovation to implementation. World Psychiatry 2019 Feb 02;18(1):20-28 [FREE Full text] [doi: 10.1002/wps.20610] [Medline: 30600624]
12. Guay C, Auger C, Demers L, Mortenson WB, Miller WC, Gélinas-Bronsard D, et al. Components and Outcomes of Internet-Based Interventions for Caregivers of Older Adults: Systematic Review. J Med Internet Res 2017 Sep 19;19(9):e313 [FREE Full text] [doi: 10.2196/jmir.7896] [Medline: 28928109]
13. Spencer L, Potterton R, Allen K, Musiat P, Schmidt U. Internet-Based Interventions for Carers of Individuals With Psychiatric Disorders, Neurological Disorders, or Brain Injuries: Systematic Review. J Med Internet Res 2019 Jul 09;21(7):e10876 [FREE Full text] [doi: 10.2196/10876] [Medline: 31290399]
14. Sherifali D, Ali MU, Ploeg J, Markle-Reid M, Valaitis R, Bartholomew A, et al. Impact of Internet-Based Interventions on Caregiver Mental Health: Systematic Review and Meta-Analysis. J Med Internet Res 2018 Jul 03;20(7):e10668 [FREE Full text] [doi: 10.2196/10668] [Medline: 29970358]
15. Carlbring P, Andersson G, Cuijpers P, Riper H, Hedman-Lagerlöf E. Internet-based vs. face-to-face cognitive behavior therapy for psychiatric and somatic disorders: an updated systematic review and meta-analysis. Cogn Behav Ther 2018 Jan 07;47(1):1-18. [doi: 10.1080/16506073.2017.1401115] [Medline: 29215315]

16. Meichsner F, Theurer C, Wilz G. Acceptance and treatment effects of an internet-delivered cognitive-behavioral intervention for family caregivers of people with dementia: A randomized-controlled trial. J Clin Psychol 2019 Apr 31;75(4):594-613. [doi: 10.1002/jclp.22739] [Medline: 30597537]
17. Žalimienė L, Blažienė I, Junevičienė J. What type of familialism is relevant for Lithuania? The case of elderly care. Journal of Baltic Studies 2020 Mar 29;51(2):159-178. [doi: 10.1080/01629778.2020.1746368]
18. Novak M, Guest C. Application of a multidimensional caregiver burden inventory. Gerontologist 1989 Dec 01;29(6):798-803. [doi: 10.1093/geront/29.6.798] [Medline: 2516000]
19. Lecrubier Y, Sheehan D, Weiller E, Amorim P, Bonora I, Sheehan KH, et al. The Mini International Neuropsychiatric Interview (MINI). A short diagnostic structured interview: reliability and validity according to the CIDI. Eur. psychiatr 2020 Apr 16;12(5):224-231. [doi: 10.1016/s0924-9338(97)83296-8]
20. RANDOM.ORG. URL: www.random.org [accessed 2021-03-24]
21. Caserta MS, Lund DA, Wright SD. Exploring the Caregiver Burden Inventory (CBI): Further Evidence for a Multidimensional View of Burden. Int J Aging Hum Dev 1996 Jan;43(1):21-34. [doi: 10.2190/2dkf-292p-a53w-w0a8]
22. Farajzadeh A, Akbarfahimi M, Maroufizadeh S, Rostami HR, Kohan AH. Psychometric properties of Persian version of the Caregiver Burden Scale in Iranian caregivers of patients with spinal cord injury. Disabil Rehabil 2018 Feb 21;40(3):367-372. [doi: 10.1080/09638288.2016.1258738] [Medline: 28637138]
23. Valer DB, Aires M, Fengler FL, Paskulin LMG. Adaptation and validation of the Caregiver Burden Inventory for use with caregivers of elderly individuals. Rev Lat Am Enfermagem 2015 Feb;23(1):130-138 [FREE Full text] [doi: 10.1590/0104-1169.3357.2534] [Medline: 25806641]
24. Kroenke K, Spitzer RL, Williams JBW. The PHQ-9. J Gen Intern Med 2001 Sep;16(9):606-613. [doi: 10.1046/j.1525-1497.2001.016009606.x]
25. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
26. Cohen S, Kamarck T, Mermelstein R. A Global Measure of Perceived Stress. Journal of Health and Social Behavior 1983 Dec;24(4):385. [doi: 10.2307/2136404]
27. Lee E. Review of the psychometric evidence of the perceived stress scale. Asian Nurs Res (Korean Soc Nurs Sci) 2012 Dec;6(4):121-127 [FREE Full text] [doi: 10.1016/j.anr.2012.08.004] [Medline: 25031113]
28. Lindner P, Frykheden O, Forsström D, Andersson E, Ljótsson B, Hedman E, et al. The Brunnsviken Brief Quality of Life Scale (BBQ): Development and Psychometric Evaluation. Cogn Behav Ther 2016 Apr 17;45(3):182-195 [FREE Full text] [doi: 10.1080/16506073.2016.1143526] [Medline: 26886248]
29. Topp CW, Østergaard SD, Søndergaard S, Bech P. The WHO-5 Well-Being Index: a systematic review of the literature. Psychother Psychosom 2015 Mar 28;84(3):167-176 [FREE Full text] [doi: 10.1159/000376585] [Medline: 25831962]
30. WHO Regional Office for Europe. Wellbeing Measures in Primary Health Care/The Depcare Project. Report on a WHO Meeting, Stockholm, Sweden, 12-13 February. Copenhagen, Denmark: WHO Regional Office for Europe; 1998. URL: https://www.euro.who.int/__data/assets/pdf_file/0016/130750/E60246.pdf [accessed 2021-03-30]
31. Weathers F, Blake D, Schnurr P, Kaloupek D, Marx B, Keane T. The life events checklist for DSM-5 (LEC-5) - Standard. National Center for PTSD. 2013. URL: https://www.ptsd.va.gov/ [accessed 2026-03-21]
32. Mažulytė E, Skerytė-Kazlauskienė M, Eimontas J, Gailienė D, Grigutytė N, Kazlauskas E. Trauma experience, psychological resilience and dispositional optimism: Three adult generations in Lithuania. Psichologija 2014;49:21-33 [FREE Full text]
33. Gray MJ, Litz BT, Hsu JL, Lombardo TW. Psychometric properties of the life events checklist. Assessment 2004 Dec 26;11(4):330-341. [doi: 10.1177/1073191104269954] [Medline: 15486169]
34. Saunders JB, Aasland OG, Babor TF, de la Fuente JR, Grant M. Development of the Alcohol Use Disorders Identification Test (AUDIT): WHO Collaborative Project on Early Detection of Persons with Harmful Alcohol Consumption--II. Addiction 1993 Jun;88(6):791-804. [doi: 10.1111/j.1360-0443.1993.tb02093.x] [Medline: 8329970]
35. Reinert DF, Allen JP. The alcohol use disorders identification test: an update of research findings. Alcohol Clin Exp Res 2007 Feb;31(2):185-199. [doi: 10.1111/j.1530-0277.2006.00295.x] [Medline: 17250609]
36. Andersson G, Carlbring P, Holmström A, Sparthan E, Furmark T, Nilsson-Ihrfelt E, et al. Internet-based self-help with therapist feedback and in vivo group exposure for social phobia: A randomized controlled trial. Journal of Consulting and Clinical Psychology 2006;74(4):677-686. [doi: 10.1037/0022-006x.74.4.677]
37. Andersson G, Paxling B, Roch-Norlund P, Östman G, Norgren A, Almlöv J, et al. Internet-based psychodynamic versus cognitive behavioral guided self-help for generalized anxiety disorder: a randomized controlled trial. Psychother Psychosom 2012 Sep 6;81(6):344-355 [FREE Full text] [doi: 10.1159/000339371] [Medline: 22964540]
38. Andersson G. Clinician-Supported Internet-Delivered Psychological Treatment of Tinnitus. Am J Audiol 2015 Sep;24(3):299-301. [doi: 10.1044/2015_aja-14-0080]
39. Andersson G. Chapter 6: Guided Internet-based CBT (ICBT) programs: introduction. In: The internet and CBT : a clinical guide. Boca Raton, Florida: CRC Press, 2015; 2015:53-54.
40. Vlaescu G, Alasjö A, Miloff A, Carlbring P, Andersson G. Features and functionality of the Iterapi platform for internet-based psychological treatment. Internet Interv 2016 Nov;6:107-114 [FREE Full text] [doi: 10.1016/j.invent.2016.09.006] [Medline: 30135819]

41. Gupta S. Intention-to-treat concept: A review. Perspect Clin Res 2011 Jul;2(3):109-112 [FREE Full text] [doi: 10.4103/2229-3485.83221] [Medline: 21897887]
42. McCoy E. Understanding the Intention-to-treat Principle in Randomized Controlled Trials. West J Emerg Med 2017 Oct 18;18(6):1075-1078 [FREE Full text] [doi: 10.5811/westjem.2017.8.35985] [Medline: 29085540]
43. Jakobsen JC, Gluud C, Wetterslev J, Winkel P. When and how should multiple imputation be used for handling missing data in randomised clinical trials - a practical guide with flowcharts. BMC Med Res Methodol 2017 Dec 06;17(1):162 [FREE Full text] [doi: 10.1186/s12874-017-0442-1] [Medline: 29207961]
44. Little R, Rubin D. Introduction. In: Statistical analysis with missing data. Hoboken, New Jersey: John Wiley & Sons, Inc., 2002; 2002:11-13.
45. Van Breukelen GJ. ANCOVA versus change from baseline: more power in randomized studies, more bias in nonrandomized studies [corrected]. J Clin Epidemiol 2006 Sep;59(9):920-925. [doi: 10.1016/j.jclinepi.2006.02.007] [Medline: 16895814]
46. Jacobson N, Truax P. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. J Consult Clin Psychol 1991 Feb;59(1):12-19. [doi: 10.1037//0022-006x.59.1.12] [Medline: 2002127]
47. Adelman RD, Tmanova LL, Delgado D, Dion S, Lachs MS. Caregiver burden: a clinical review. JAMA 2014 Mar 12;311(10):1052-1060. [doi: 10.1001/jama.2014.304] [Medline: 24618967]
48. Spatuzzi R, Giulietti MV, Ricciuti M, Merico F, Fabbietti P, Raucci L, et al. Exploring the associations between spiritual well-being, burden, and quality of life in family caregivers of cancer patients. Pall Supp Care 2018 May 11;17(03):294-299. [doi: 10.1017/s1478951518000160]

Abbreviations

AUDIT: The Alcohol Use Disorders Identification Test
BBQ: The Brunnsviken Brief Quality of Life Scale
CBI: Caregiver Burden Inventory
CBT: cognitive behavioral therapy
GAD-7: Generalized Anxiety Disorder-7
ICBT: Internet-delivered cognitive behavioral therapy
ITT: Intention-to-treat principle
LEC: Life Events Checklist
MAR: missing at random
PHQ-9: Patient Health Questionnaire-9
PSS-14: Perceived Stress Scale-14
RCI: Reliable Change Index
WHO-5: The World Health Organization-Five Well-Being Index

Edited by G Eysenbach, R Kukafka; submitted 16.06.20; peer-reviewed by S Zabihi, M Lozano-Lozano, A Dominguez-Rodriguez; comments to author 29.09.20; revised version received 21.10.20; accepted 12.03.21; published 07.04.21.

Please cite as:
Biliunaite I, Kazlauskas E, Sanderman R, Truskauskaite-Kuneviciene I, Dumarkaite A, Andersson G
Internet-Based Cognitive Behavioral Therapy for Informal Caregivers: Randomized Controlled Pilot Trial
J Med Internet Res 2021;23(4):e21466
URL: https://www.jmir.org/2021/4/e21466
doi:10.2196/21466
PMID:33825687

Original Paper

A Digital Patient-Provider Communication Intervention (InvolveMe): Qualitative Study on the Implementation Preparation Based on Identified Facilitators and Barriers

Berit Seljelid[1,2,3], RN, MSc; Cecilie Varsi[1,4], RN, PhD; Lise Solberg Nes[1,2,5], LP, PhD; Kristin Astrid Øystese[6,7], MD, PhD; Elin Børøsund[1], RN, PhD

[1]Department of Digital Health Research, Division of Medicine, Oslo University Hospital, Oslo, Norway

[2]Institute of Clinical Medicine, Faculty of Medicine, University of Oslo, Oslo, Norway

[3]Department of Cooperation, Patient Education and Equivalent Health Services, CEO's Staff, Oslo University Hospital, Oslo, Norway

[4]Faculty of Health and Social Sciences, University of South-Eastern Norway, Drammen, Norway

[5]Department of Psychiatry & Psychology, College of Medicine & Science, Mayo Clinic, Rochester, MN, United States

[6]Section of Specialized Endocrinology, Department of Endocrinology, Morbid Obesity and Preventive Medicine, Oslo University Hospital, Oslo, Norway

[7]Department of Medical Biochemistry, Institute of Clinical Medicine, Faculty of Medicine, University of Oslo, Oslo, Norway

Corresponding Author:
Elin Børøsund, RN, PhD
Department of Digital Health Research
Division of Medicine
Oslo University Hospital
Box 4950 Nydalen
Oslo, N-0424
Norway
Phone: 47 22894357
Email: elin.borosund@rr-research.no

Abstract

Background: Chronic health conditions are affecting an increasing number of individuals, who experience various symptoms that decrease their quality of life. Digital communication interventions that enable patients to report their symptoms have been shown to positively impact chronic disease management by improving access to care, patient-provider communication, clinical outcomes, and health-related quality of life. These interventions have the potential to prepare patients and health care providers (HCPs) before visits and improve patient-provider communication. Despite the recent rapid development and increasing number of digital communication interventions that have shown positive research results, barriers to realizing the benefits offered through these types of interventions still exist.

Objective: The aim of this study is to prepare for the implementation of a digital patient-provider communication intervention in the daily workflow at 2 outpatient clinics by identifying potential determinants of implementation using the Consolidated Framework for Implementation Research (CFIR) to tailor the use of digital communication intervention to the intended context and identify key aspects for an implementation plan.

Methods: A combination of focus groups, workshops, and project steering committee meetings was conducted with HCPs (n=14) and patients (n=2) from 2 outpatient clinics at a university hospital. The CFIR was used to guide data collection and analysis. Transcripts, written minutes, and notes were analyzed and coded into 5 CFIR domains using thematic analysis.

Results: Data were examined and analyzed into 18 CFIR constructs relevant to the study purpose. On the basis of the identified determinants, important intervention tailoring includes adjustments to the digital features and adjustments to fit the clinical workflow and a decision to conduct a future pilot study. Furthermore, it was decided to provide the intervention to patients as early as possible in their disease trajectory, with tailored information about its use. Key aspects for the implementation plan encompassed maintaining the identified engagement and positive attitude, involving key stakeholders in the implementation process, and providing the needed support and training.

Conclusions: This study offers insight into the involvement of stakeholders in the tailoring and implementation planning of a digital communication intervention in clinical practice. Stakeholder involvement in the identification of implementation facilitators and barriers can contribute to the tailoring of digital communication interventions and how they are used and can also inform systematic and targeted implementation planning.

(J Med Internet Res 2021;23(4):e22399) doi:10.2196/22399

KEYWORDS

eHealth; digital communication; secure messages; digital symptom assessment; implementation; tailoring; Consolidated Framework for Implementation Research; CFIR; facilitators; barriers; stakeholders

Introduction

Background

Living with a chronic health condition causes symptoms that negatively affect health-related quality of life (HRQoL) [1-4]. Symptom recognition may be challenging for patients and, by extension, for health care providers (HCPs) [5,6]. This can relate to patients' poor understanding of disease mechanisms and progression, lack of knowledge, and low levels of health literacy [5,7] and to practical barriers to inform HCPs about symptoms [6]. Experiences of difficulties in communication and interaction between patients and HCPs are common, including poor timing of information and challenging symptom recognition, which are factors that may interfere with symptom management and help seeking [5,7]. eHealth communication interventions may offer the potential to alleviate such difficulties.

Studies of eHealth communication interventions have reported benefits in terms of patient-provider communication [8-14], patient-provider relationship [15], patient self-management [16], symptom management [11,17,18], preparation before hospital visits [9,10,18,19], and HRQoL [8,17,20] in chronic health care settings. Despite these benefits, barriers to benefit realization still exist [21], including staff familiarity with technology [22,23], level of patient education [20,23], and issues with user-friendliness [22,24].

Although the positive effects of eHealth interventions on patient-provider communication and patient outcomes are known, HCPs report concerns regarding the integration of eHealth interventions into daily workflow [25,26] and concerns about increased workload [27-29]. In addition, the use of eHealth interventions can challenge HCPs' competence [25,26]. Such challenges may act as barriers to implementation and actual use, which could be another barrier for use. Successful implementation of such eHealth interventions requires attention to the development and evaluation of strategies to implement the interventions [30-32]. An important factor for the acceptance and success of eHealth implementation is the tailoring of the interventions to suit the local context [33]. Stakeholders representing the target group and the actual context can provide important input to reduce system complexity, increase system acceptability, and make systems as user-friendly as possible [33,34]. To increase the likelihood of implementation success, there is also a need to examine intervention characteristics from the end-user perspective in order to inform the tailoring of the intervention to suit contextual needs.

Implementation refers to the systematic uptake of research into HCP practice to improve the quality of health care services [35]. Implementation strategies can be explained as methods or techniques used to enhance the adoption, implementation, and sustainability of HCP clinical practice [36]. However, there is limited guidance regarding the types of implementation strategies that may be effective when implementing eHealth interventions to practice [30]. Nevertheless, it has been suggested that implementation strategies should be selected and tailored to address the unique contextual needs based on an identification of determinants (ie, factors that act as facilitators or barriers) that may influence the implementation process [32]. The identification of determinants can be used to address barriers and leverage facilitators [32,37].

Implementation frameworks can guide the identification of determinants that might influence the implementation, its effectiveness, and the implementation process [32,35]. The Consolidated Framework for Implementation Research (CFIR) is a widely used framework to identify facilitators and barriers [38-40]. The CFIR was developed from a synthesis of 20 existing theories and frameworks and consists of 5 overarching domains, including 39 specific constructs within these 5 domains [38]. The first domain of the CFIR is the *Intervention Characteristics* and includes constructs such as the adaptability of the intervention, the perceived relative advantage, and the complexity and cost of the intervention [38]. The *Outer Setting* domain includes constructs such as the patient's needs and resources related to the intervention, whereas the *Inner Setting* domain includes constructs such as implementation climate and readiness for implementation, the organization's culture, and leadership engagement. The fourth domain is the *Characteristics of Individuals* involved in the intervention or implementation process; it relates to personal attributes, including personal traits such as motivation, values, and competence. The last domain relates to the *Process* and includes planning, execution, and evaluation of the implementation process [38].

Objectives

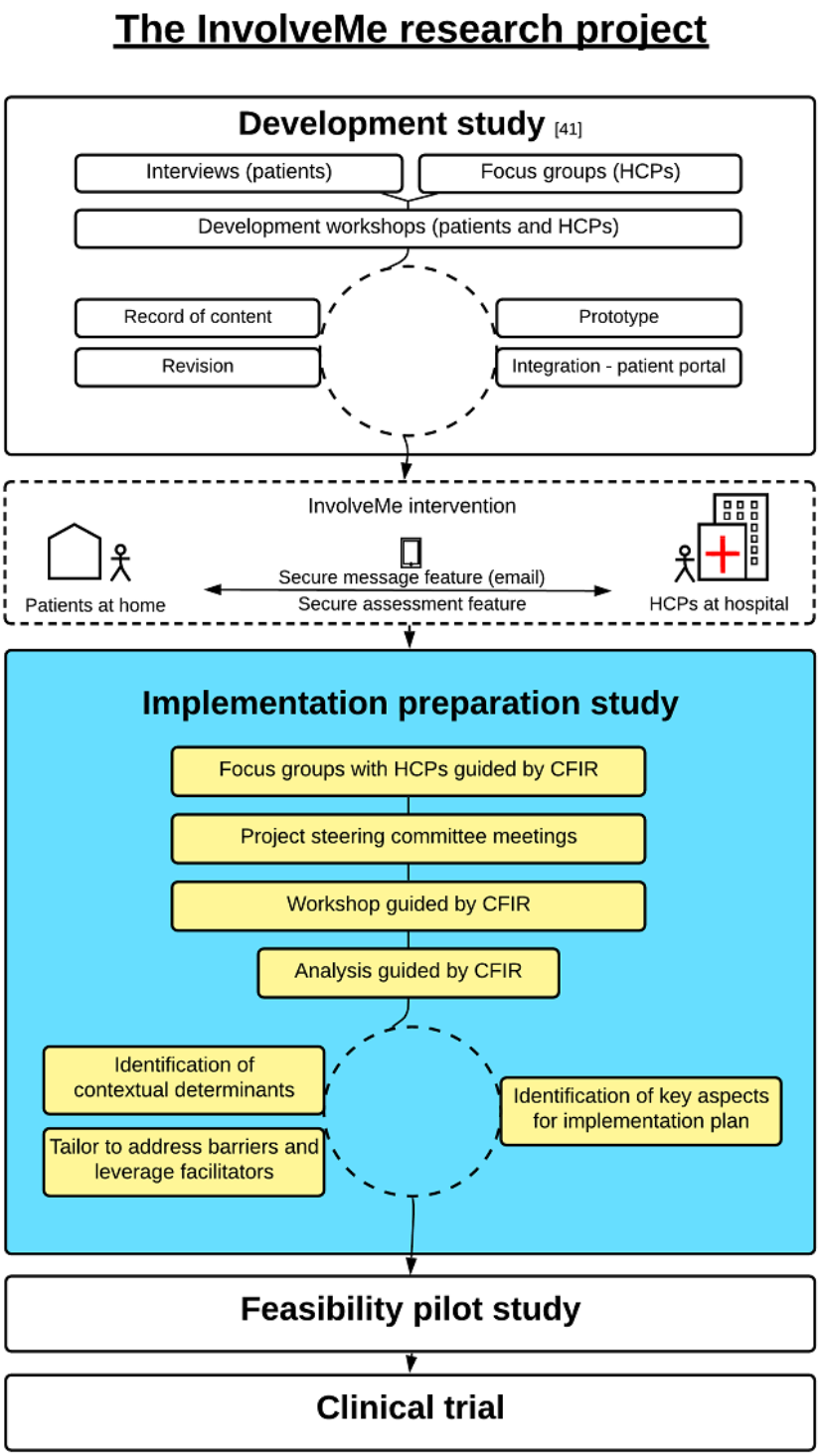
The aim of this study is to prepare the implementation of a digital patient-provider communication intervention, *InvolveMe*, into the daily workflow at 2 outpatient clinics where patients with chronic health conditions are treated by identifying potential facilitators and barriers to implementation using CFIR as the conceptual framework to (1) tailor the *InvolveMe* intervention to the intended context and (2) identify key aspects for an implementation plan.

Methods

Overview

This study is part of the *InvolveMe* research project, which includes the development, implementation, and evaluation of a digital intervention (Figure 1). The *InvolveMe* research project is a collaboration between 2 outpatient clinics and 1 research department at a large university hospital in Norway. The *InvolveMe* intervention will be implemented in 2 outpatient clinics and tested in a future clinical trial (Figure 1).

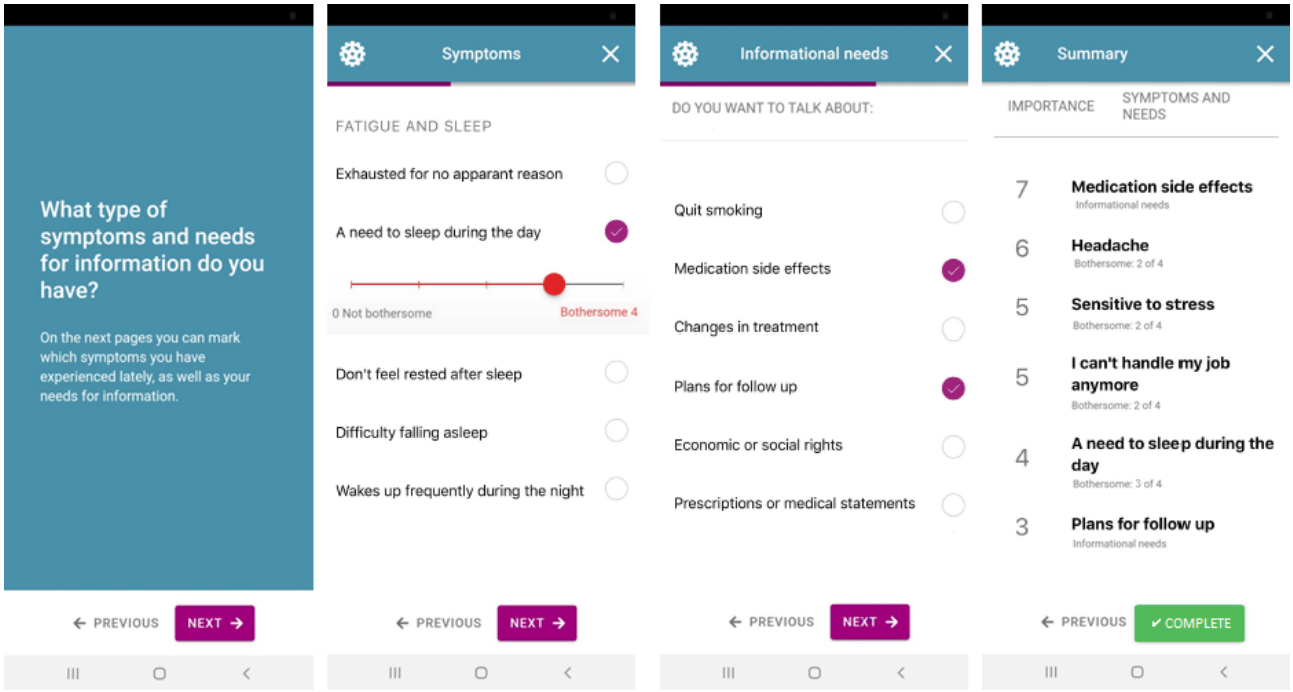
Figure 1. Overview of the InvolveMe research project. CFIR: Consolidated Framework for Implementation Research; HCP: health care providers.



Description of the InvolveMe Intervention

InvolveMe was developed at the initiative of HCPs aiming to improve follow-up for two specific categories of patients: renal transplant recipients and patients with nonfunctioning pituitary adenomas [41]. *InvolveMe* was internally developed at the hospital in close cooperation with registered nurses, physicians, health support personnel, patients, researchers, and system developers (Figure 1). A detailed description of the content and system development of *InvolveMe* is provided elsewhere [41]. The *InvolveMe* intervention contains two features: (1) a secure message feature and (2) a secure assessment feature (ie, predefined list) where patients can prioritize their need for symptom management, information, and preferences for care from home and on a scale from 0 to 10 (Figure 2).

Figure 2. Screenshots of InvolveMe.



Patients and HCPs can use the secure message feature to interact with each other between or after outpatient visits [41]. Completion of the secure assessment feature generates a summary that is sent to the patients' HCPs for use in upcoming consultations. The secure message(s) and the assessment(s) are integrated into an existing patient portal that allows patients to read their electronic patient record (EPR). However, the opposite is not possible (ie, the EPR cannot receive data from the patient portal) owing to information safety regulations.

Design

This study used a participatory and iterative design approach [42] using qualitative methods for data collection. The data collection period was November 2017 to December 2018 and proceeded through focus groups, project steering committee meetings, and a workshop (Figure 1). Data collection from focus groups and workshops was guided by CFIR [38] (Multimedia Appendices 1 and 2). The variety of data collection activities and diverse data collection approaches allowed for mutual stakeholder learning and comprehension of all stakeholder perspectives involved [43]. As part of the main focus of this study is to identify local determinants to tailor the intervention into the intended context (ie, HCPs' practice) and identify key aspects for an implementation plan, most study participants represented the HCP perspective. However, to ensure that the patient perspective was also continuously involved, 2 patients with experience from the *InvolveMe* development study [41] were included in the steering committees to ensure knowledge, inclusion, and prioritization of the patient perspective(s).

The determinants identified from data collection and analysis informed the tailoring of the *InvolveMe* intervention to suit workflow at the 2 participating outpatient clinics and aided in the identification of key aspects to include in an implementation plan. Tailoring, as described in this study, refers to addressing intervention barriers and leveraging facilitators as key aspects of the implementation planning process.

Data Collection Guided by CFIR

CFIR allows researchers to select constructs that they perceive as most relevant and use them to guide the assessment of determinants in the implementation context [38]. The operationalization of CFIR domains in this study was based on discussion and consensus in the research team, where all 5 domains of CFIR were explored for the development of focus group and workshop guides (Multimedia Appendices 1 and 2). Significant themes were to be discussed for determinants important for tailoring the intervention and providing input for the implementation plan (ie, key aspects). By asking about themes to discuss within each domain, rather than questions for each construct, several CFIR constructs would most likely not be covered by the focus group and workshop guides.

Focus Groups

The focus group guide centered around HCPs' experiences from previous successful implementation projects, their perception

of possible advantages and challenges of using a digital patient-provider communication intervention, and how to successfully implement the *InvolveMe* intervention (Multimedia Appendix 1).

Workshop

The workshop guide consisted of the 5 operationalized domains of CFIR (Multimedia Appendix 2): (1) the *InvolveMe* intervention (*Intervention Characteristics*); (2) the patients who will be offered the *InvolveMe* intervention (*Outer Setting*); (3) the 2 outpatient clinics where the patients are being treated (*Inner Setting*); (4) the HCPs (*Characteristics of Individuals*); and (5) the preparation for implementation of the *InvolveMe* intervention (*Process*).

Settings, Participants, and Recruitment

Participants were HCPs and patients. HCPs were purposely selected and recruited from an outpatient nephrology clinic and an outpatient endocrine clinic at a large university hospital in Norway. They were registered nurses, physicians, and health support personnel responsible for the treatment and care of patients with renal transplants or nonfunctioning pituitary adenomas. HCPs were provided with written information about the study and those willing to participate were included. The patients were participants in the development study [41] (Figure 1). They were asked to participate in the project steering committees, representing each of the 2 categories of patients. This was based on detailed knowledge about the *InvolveMe* research project [41], in addition to their own experience of being a patient. The patients were contacted by HCPs at the clinics and asked to participate before being contacted by the first author (BS), who described study participation in detail before the final study participation agreement was received.

HCPs had 2 to 38 years of clinical experience from specialist health care, with a median of 49.5 years (range 28-63 years), and most were female (10/14, 71%). Some HCPs participated in all data collection activities, whereas others participated in 1 or 2 activities (Table 1). The patient participants were female.

Table 1. Overview of participants in the focus groups, project steering committees, and workshop.

Stakeholders[a]	Focus groups (n=11)	Project steering committee (n=6)	Workshop (n=7)
Endocrine outpatient clinic			
Head of clinic (physician)		✓[b]	✓
Registered nurse	✓	✓	✓
Registered nurse	✓		✓
Registered nurse	✓		✓
Physician	✓		
Physician	✓		
Nephrology outpatient clinic			
Head of clinic (physician)		✓	✓
Registered nurse	✓	✓	✓
Registered nurse	✓		
Registered nurse	✓		
Physician	✓		
Physician	✓		
Health support personnel	✓		
Health support personnel			✓
Other stakeholders			
Patient participant		✓	
Patient participant		✓	

[a]All stakeholders participated in the development study [41], except for one head of the clinic and one health support personnel.

[b]Participated in data collection.

Data Collection

Focus Groups

HCPs were invited to participate in focus groups, a method suitable for exploring attitudes and experiences, and to encourage group discussion [44]. The HCPs from the nephrology (n=6) and endocrine (n=5) clinics participated in separate groups to explore context-related determinants and key aspects of an implementation plan (Table 1). The focus groups were facilitated by the first (BS) and last (EB) authors. Both focus group sessions lasted approximately 50 minutes and were recorded with a digital voice recorder and transcribed verbatim.

Project Steering Committee Meetings

A total of 2 project steering committees were established, one for each participating clinic, to promote leadership and stakeholder engagement in the intervention and to ensure input on the process of tailoring the intervention to fit contextual needs. Participants in each of the 2 project steering committees represented either nephrology (n=3) or endocrine (n=3) outpatient clinics (Table 1). The first (BS) and last (EB) authors facilitated the committee meetings. Each group met twice in the preimplementation phase of the study, which lasted for 1 year. All committee meetings lasted approximately 60 minutes and had a set agenda with topics to discuss (eg, workshop preparation and integration of the intervention into practice). Data collection from committee meetings was based on written minutes made by the last author (EB) during meetings. Each participant received and approved the minutes before the analysis.

Workshop

On the basis of the project steering committee meeting discussions and decisions, a joint workshop (n=7) for both participating clinics was considered expedient to share insights and experiences and to identify determinants for tailoring and key aspects for implementation (Table 1). A workshop with HCPs from both clinics was therefore conducted to gain further insight into participants' reflections and expectations about the *InvolveMe* intervention and elaborate on how to implement the intervention in the 2 clinics [45]. Workshop participants were invited to share their reflections through an exercise in which they were presented with the 5 operationalized domains of CFIR to facilitate narration (Multimedia Appendix 2). The presentation of each domain was followed by group discussions on potential facilitators and barriers. Thereafter, Post-it notes were used to present group reflections and encourage discussions between participants from the 2 clinics. The workshop was facilitated by the first (BS) and last (EB) authors and lasted 180 minutes, including a 15-minute break. Data collection from the workshop resulted in a report based on written notes made by the last author (EB) and pictures of the written Post-it notes made and shared by workshop participants.

Analysis

Transcripts, meeting minutes, and notes from the 3 data collection activities were deductively analyzed as one data set, based on thematic analysis by Braun and Clarke [46,47], and into the 5 domains of CFIR (ie, themes). The first author (BS) led the analysis process, which involved 2 coauthors (EB and CV). The first step was to read through and become familiar with the transcripts, meeting minutes, and notes. Early impressions were captured during the writing process. Next, the data were coded (ie, coding by CFIR constructs) using an Excel spreadsheet. Quotes from the focus group transcripts were copied and pasted into the spreadsheet along with text sections from meeting minutes and notes. Colors were used to mark data based on sources. The codes were then resorted and re-evaluated based on the CFIR domains and constructs. Through regular coauthor meetings (BS, EB, and CV), codes were discussed and revised to reach a consensus. Codes that did not appear to fit any of the CFIR constructs were also re-evaluated. The analysis was then refined, and the results were written and reviewed. In the final step, quotes were chosen for representation.

Ethics

The study was performed in accordance with the Helsinki Declaration and approved by the Department for Data Protection and Information Security (equivalent to an institutional review board) at Oslo University Hospital (20178/9223). Written informed consent was obtained from all participants (ie, HCPs and patient representatives). To guarantee confidentiality, the transcripts were coded with project ID numbers and stored on a secure server for sensitive research data, as required by the Department for Data Protection and Information Security at Oslo University Hospital. Only the first author and project administrator (BS) and last author and principal investigator (EB) had access to the code connecting the project ID numbers and the actual participant's name. Owing to the design and implementation emphasis of the study, the need for study approval was waived by the Regional Committee for Medical and Health Research Ethics for South East Norway, which is in line with the Norwegian legislation [48].

Results

Overview

Data were examined and analyzed into 18 CFIR constructs relevant to the study purpose. The constructs are presented by the domains of CFIR, which include description and considerations regarding tailoring of the *InvolveMe* intervention to the intended context (Tables 2-5) and identification of key aspects for the implementation plan (Table 6). A brief description of the relevant CFIR construct is provided to support the interpretation of the results.

Table 2. Intervention Characteristics: determinants, tailoring, and identification of key aspects for implementation planning.

Construct and study results	Considerations regarding determinants	Tailoring and key aspects
Intervention Source: intervention considered as internally developed	Facilitator: the involvement in the development study [41] and this study may promote ownership to the intervention, which may support intervention implementation.	Key aspect: ownership
Relative Advantage: the intervention as an advantage to current practice	Facilitator: HCPs[a] pointed to aspects perceived to be advantages of the intervention. This may be considered as a positive attitude to what is being implemented.	Key aspect: a positive attitude to implement the intervention
Adaptability: use of existing system	Facilitator: participants perceived the previous integration of *InvolveMe* in a patient portal to be beneficial. This may support acceptance and adoption	Key aspect: system acceptance and adoption
Adaptability: a new work task; the secure assessment feature	Barrier: The assessment was a new work task for HCPs, which caused a concern for increased workload and potentially increased time pressure on consultations.	Tailoring: the assessment feature was condensed to a brief list and refinement was made to the summary
Trialability: a need to test before the clinical trial	Barrier: HCPs highlighted that a pilot study would be important to test the intervention and the implementation strategies. A test of the intervention would also inform HCPs that were not formerly involved in the research project and potentially address concerns in advance.	Tailoring: decision, agreed upon by all parties involved, to conduct a pilot study
Complexity: lack of integration between EPR[b] and patient portal	Barrier: it was recognized as important to improve accessibility and avoid paper printouts of the assessment summary.	Tailoring: the summary was created in a format that could be copied and pasted from the patient portal and into the EPR
Complexity: messages sent directly to the physicians	Barrier: it was considered important to tailor the intervention to suit the physician's clinical workflow to succeed with intervention implementation.	Tailoring: a shared email inbox with a dedicated triage moderator was established

[a]HCP: health care provider.

[b]EPR: electronic patient record.

Table 3. Outer Setting: determinants, tailoring, and identification of key aspects for implementation planning.

Construct and study results	Considerations regarding determinants	Tailoring and key aspects
Patient Needs and Resources: *InvolveMe* could potentially contribute to less anxiety	Facilitator: HCPs[a] described patients being worried and anxious early in the disease trajectory. *InvolveMe* could be beneficial to patients in terms of increased information and thereby potentially help patients avoid or experience less anxiety.	Tailoring: to provide the intervention as early as possible to patients
Patient Needs and Resources: patient's motivation to use *InvolveMe*	Facilitator: it can be difficult to reach HCPs on the telephone. *InvolveMe* may have the potential to represent a place where patients can get in contact with HCPs.	Key aspect: motivated patients could contribute to HCPs implementing *InvolveMe*
Patient Needs and Resources: patient acceptance—use of a patient portal	Facilitator: the potential for intervention integration into a patient portal seemed acceptable. This supports findings from the development study [41].	Key aspect: intervention acceptance and adoption
Patient Needs and Resources: patient acceptance—use of a digital health service	Barrier: HCPs described being concerned that *InvolveMe* might be technically demanding for some patients. Therefore, *InvolveMe* was designed to be a voluntary supplement to standard care, not a replacement. The assessment in *InvolveMe* can act as preparation before consultations. To make this clear to all patients, relevant information should be provided.	Tailoring: provide patients with tailored information about the intervention

[a]HCP: health care provider.

Table 4. Inner Setting: determinants, tailoring, and identification of key aspects for implementation planning.

Construct and study results	Considerations regarding determinants	Tailoring and key aspects
Structural Characteristics: 2 outpatient clinics organized differently from each other	Facilitator: knowledge about the different organization and staffing may be of importance for tailoring the intervention to fit each outpatient clinics (ie, the moderator functioning).	Tailoring: one clinic designated a nurse to be the moderator, and the other clinic designated health support personnel
Network and Communication: weekly meetings for activity planning	Facilitator: existing meetings were considered appropriate and feasible to discuss and evaluate the implementation process in the research project.	Key aspect: use of existing weekly meetings to monitor implementation process
Culture: interest in innovations	Facilitator: an interest in innovations may provide opportunities to interact with end users (here HCPs[a]) regarding the intervention. This has the potential to support a collaborative relationship between researchers and HCPs.	Key aspect: a collaborative relationship
Tension for Change: improve patient follow-up	Facilitator: HCPs perceived the current situation as demanding, which could contribute to strengthened motivation to change practice (ie, intervention implementation).	Key aspect: monitoring the number of phone calls and the measurement of HRQoL[b] before and after intervention to visualize change
Leadership Engagement: the heads of the clinics were engaged and active	Facilitator: by their participation in the research project, the heads of the clinics display their commitment and accountability, which may contribute to staff engagement and support a culture for change.	Key aspect: providing anchoring and acceptance for the intervention and a change of practice

[a]HCP: health care provider.

[b]HRQoL: health-related quality of life.

Table 5. Characteristics of Individuals: determinants, tailoring, and identification of key aspects for implementation planning.

Construct and study results	Considerations regarding determinants	Tailoring and key aspects
Knowledge and Beliefs: a positive attitude about using a digital intervention such as *InvolveMe*	Facilitator: a positive attitude may act as a facilitator for the implementation process. Reflection on how to maintain a positive attitude throughout the implementation process should be done to establish a close researcher-clinician relationship. The provision of positive feedback along the implementation process might contribute to maintenance of use and collaboration.	Key aspect: maintaining the positive attitude

Table 6. Process: determinants and identification of key aspects for implementation planning.

Construct and study results	Considerations regarding determinants	Key aspects
Planning: lack of information and assignment of responsibility may reduce the motivation of HCPs[a]	Barrier: providing information and intervention guidance to staff involved in intervention implementation could include providing project information at meetings and brief updates via email or other information channels to all staff members. Meetings and updates could also allow for information exchange on implementation strategies. Easy access to researchers and technical support in case of questions may be of importance.	Providing: • Timely information • Someone to call on a specific number • Technical training and support • Joint project steering committee meetings for mutual exchange of experiences
Engagement: attracting and involving HCPs	Facilitator: some participants initiated writing abstracts to present study details at local and national conferences.	• Maintaining engagement
Opinion Leaders	Facilitator: physicians (and head of clinics) were described by some of the nurses as filling an Opinion Leader role.	• Involving Opinion Leaders in implementation
Implementation Leaders	Facilitator: participants of the project steering committee were suggested as filling the positions as Implementation Leaders.	• Involving members of steering committees in implementation
Champions	Facilitator: registered nurses with a responsibility for the project were seen as potential Champions and drivers of the implementation, inspiring, motivating, and helping other staff members.	• Involving registered nurses in implementation
External Change Agents: provide support to clinics	Facilitator: the clinics wanted a designated external facilitator from the research team to provide support for staff members in implementation.	• First author (BS) designated as External Change Agent in this study

[a]HCP: health care provider.

Intervention Characteristics

Intervention Source is defined as the key stakeholders' perception of whether the intervention is externally or internally developed [38]. Most participants were involved in activities to prepare the content and development of the system underlying *InvolveMe*, as described elsewhere [41]. The *InvolveMe* intervention was collaboratively developed within the hospital, and the participating HCPs expressed a perception of *InvolveMe* ownership.

Relative Advantage refers to the stakeholder's perception of the advantage of implementing the intervention rather than an alternative solution [38]. Although symptom assessments were a part of routine consultations, they were performed based on the preference and prioritization of the HCP and the history of the patients. Most participants perceived that the *InvolveMe* intervention could be an advantage compared with current practice where there is no digital communication between patients and HCPs. The participants reported that such a digital intervention could increase patient safety; raise awareness about the patient's perspective (ie, symptoms and informational needs); and improve patient-provider communication, patient satisfaction, and HRQoL. An intervention that could document contact between patient and provider and reduce the number of phone calls from patients was seen as warranted. One participant stated:

> *An email is much less disruptive than a phone call. An email I open when I have some spare time, while the phone call I have to answer while in the middle of something, while doing something else.* [HCP 10, focus group]

Adaptability refers to the degree to which an intervention can be adapted, tailored, and refined to meet local needs [38]. The participants shared their opinions on how they thought *InvolveMe* could fit into existing workflows in the clinics. The use of an already existing system was perceived as positive. Participants perceived that there were "already too many digital clinical systems" and that it was beneficial for *InvolveMe* to be integrated into an excising system (ie, patient portal) [41]. Some HCPs expressed concern that the intervention would introduce additional work tasks in an already hectic work environment. These concerns were raised surrounding worries that patients might complete extensive assessments, expecting everything to be addressed in the consultation. One participant stated:

> *If it becomes one more thing I have to deal with when meeting a patient for half an hour, we'll have to start considering extending the consultation time.* [HCP 6, focus group]

Trialability relates to the ability to test the intervention on a small scale in the organization and be able to reverse course if warranted [38]. Participants were positive for participating in

a clinical trial to test *InvolveMe*. However, HCPs raised concerns about carrying out the planned trial without them being able to test the intervention in advance.

Complexity is defined as the perceived difficulty of implementation, reflected by duration, scope, radicalness, disruptiveness, centrality, intricacy, and the number of steps required for implementation [38]. The participants stated that the digital communication tool must be intuitive and easy for them to use. As expressed by one participant:

> *It has to be something that is intuitive and easy to answer and...something you don't spend a lot of extra time on.* [HCP 9, focus group]

Although *InvolveMe* could be integrated into a patient portal, the participants expressed concern that the intervention, because of data protection and privacy regulations, likely would not be allowed to communicate directly with the hospital EPR. If current regulations would require that paper printouts from *InvolveMe* had to be manually scanned into the EPR, rather than received directly from the patient portal, participants were concerned that this would add to their workload. The participating physicians also raised concerns about receiving secure messages directly to an individual mailbox, without some form of triage. There were also some concerns that the message functionality might become more like a chat, with messages going back and forth between patients and HCPs, potentially increasing the time HCPs spend communicating with each other for management of the patients' many questions.

Outer Setting

The construct of *Patient Needs and Resources* concerns the extent to which patients' needs and facilitators and barriers to meeting these needs are accurately known and prioritized by the organization [38]. The participating HCPs explained that they, based on their own experience, perceived patients as the most worried and anxious early in the disease trajectory. They described a structured follow-up for the 2 categories of patients, with room for improvement. As expressed by one participant:

> *That's also what we, me too, have been thinking about for many years when it comes to our patients, that they come to the 3-month check-up and they have questions that we could have answered for them [before the time of check-up], but they've had no place to pose their questions before consultation.* [HCP 2, focus group]

HCPs described that they thought most patients would be motivated to use *InvolveMe* and that the intervention could improve patient-provider communication related to symptoms, needs, and preferences, but also serve as a secure digital channel where patients knew that they could get in touch with their HCPs between consultations. This aspect was also discussed in the workshop. One participant described the following:

> *Satisfied patients, they will feel more seen and heard.* [Post-it note, workshop]

The use of the existing system was considered positive for patient use. However, the HCPs were concerned about various aspects of patient acceptance. They expressed thoughts that some patients might be afraid of losing in-person contact with their HCPs and that digital communication might not suit all patients. This issue was particularly raised as a digital intervention could potentially require a level of digital competence that some patients might not have. One participant stated:

> *Some patients might be afraid to use technology.* [Post-it note, workshop]

Adding to the HCP input, the patient participants in this study supplemented patient input from the development study [41] and strengthened the patient's voice by providing direct input on the *InvolveMe* intervention. They were very positive toward the use of *InvolveMe* and expressed their view that digital patient-provider communication would strengthen patient follow-up.

Inner Setting

The construct of *Structural Characteristics* is explained as the social architecture, age, maturity, and size of an organization [38]. The 2 included outpatient clinics were organized differently from each other, although both clinics described staff stability. One clinic was larger than the other and included 2 registered nurses and several physicians. This clinic also had several health support personnel who organized much of the patient-administrative work for registered nurses and physicians. The other clinic included registered nurses and physicians in a relatively small HCP group, which was perceived as an advantage by the HCPs in question in terms of implementation. One participant stated:

> *It is probably an advantage that we are a small group, and not thousands of people.* [HCP 5, focus group]

The construct of *Network and Communication* involves the nature and quality of social networks and the nature and quality of formal and informal communication in an organization [38]. Both clinics had weekly meetings for activity planning, where research projects, including this study, were discussed.

Culture, as a construct, includes the norms, values, and basic assumptions of a given organization [38]. The HCPs reported that they were generally interested in innovations. This interest was also displayed in attendance and discussions at presentations and meetings about *InvolveMe*. Most of the participating HCPs also pointed to the potential for improved symptom management through interventions such as *InvolveMe*. One participant stated:

> *It's the issue of identifying the patient's problem...that we sometimes struggle to capture.* [HCP 4, focus group]

Tension for Change is the degree to which HCPs perceive the current situation as intolerable or needing change [38]. With regard to digital patient-provider communication, there was a general tension for change among all groups of HCPs in this study. All participants described receiving many phone calls from patients, and that they needed and wanted an easier method for patient follow-up than what current practice allowed, suggesting that digital communication could be one way to improve this issue. One participant stated:

> *There will be less "noise" if we have one of those electronic communication channels...then we would have the opportunity to convey something, and at the same time reduce the patient's level of anxiety.* [HCP 4, focus group]

Leadership Engagement refers to the commitment, involvement, and accountability of leaders regarding implementation [38]. The heads of the participating clinics were positive and engaged members of the project steering committee. They were also supportive and involved in the research project, facilitating and participating in research activities and allocating clinic personnel to participate in research project activities and meetings.

Characteristics of Individuals

The construct of *Knowledge and Beliefs* about the intervention involves individuals' attitudes and the value placed on the intervention and familiarity with facts, truths, and principles related to the intervention [38]. The participants expressed, for the most part, a positive attitude toward using *InvolveMe* and stated that they believed the use of such an intervention could improve clinical practice through highlighting the importance of good patient-provider communication related to symptom management. One participant stated:

> *Being able to clarify some expectations makes it easier to...the patient is better prepared for consultation, they understand what they are struggling with and why they come in for consultation...it will potentially make it easier to talk to them when some things are clarified in advance...* [HCP 4, focus group]

Participating HCPs expressed that digital interventions, such as *InvolveMe*, should be a part of modern practice. The positive attitude toward an intervention such as *InvolveMe* was also expressed in other ways, for example, written on a Post-it note:

> *Will provide structure to the workday.* [Post-it note, workshop]

HCPs also stated that they believed that such an intervention could make patients feel safe and cared for. One participant said:

> *Possibly an increased level of security [for the patient] provided by a communication channel that is not filtered through a switchboard...* [HCP 5, focus group]

Process

The construct of *Planning* is explained as the degree to which a scheme or method of behavior, and tasks for implementing an intervention in advance, corresponds with the consideration of the quality of those schemes or methods [38]. The participants in this study stated that a lack of information and assignment of responsibility could potentially reduce HCP motivation. The importance of providing information and guidance for use to everyone involved at the clinics was highlighted:

> *When switching to new systems, it is always important to have an easily accessible support person who can help solve issues right away.* [HCP 5, focus group]

The heads of the participating clinics suggested joint project steering committee meetings to exchange information on implementation strategies in the implementation process. In addition, availability from someone from the research team, including the possibility to call if the HCPs had any questions, was suggested. The need for technical support and training was also suggested:

> *Some training in the use of the software maybe...*
>
> *Yes, but I often think we get too much of that...*
>
> *Agreed, but not too long in advance then, as it is so easy to forget. But you could get help with specific things that you wonder about, and then you learn and acquire knowledge, while if you're sitting in a classroom, learning about a lot of things that you can't really easily relate to...* [Discussion between HCP 6 and 9, focus group]

The construct of *Engagement* involves attracting and involving appropriate individuals in the implementation and use of the intervention through a combined strategy of social marketing, education, role modeling, training, or similar activities [38]. There was definite engagement in the planned intervention in this study. For example, some participants initiated writing abstracts to present study details at local and national conferences. This initiation was discussed in project steering committees, and the research team allocated responsibility for the writing process. Abstracts written for the part of the process also received two Best Poster Awards and a Meritorious Abstract Award [49] and contributed to maintaining engagement.

Opinion Leaders are individuals in an organization who have formal or informal influence on the attitudes and beliefs of their colleagues regarding the implementation of the intervention [38]. Physicians (and head of clinics) were described by some of the nurses as filling an *Opinion Leader* role. Participants of the project steering committee were suggested to fill the positions as *Implementation Leaders*. Registered nurses who were responsible for the project were seen as potential *Champions* and drivers of the implementation.

External Change Agents are individuals who are affiliated with an outside entity and who formally influence and facilitate interventions in a desirable direction [38]. By facilitating the project steering committee meetings, the first (BS) and last (EB) authors potentially influenced and facilitated the intervention as *External Change Agents*. In addition, the participants suggested a facilitator from the research team to be available to the clinic staff members for support during the implementation process.

Discussion

Principal Findings

This study identifies the determinants using the CFIR framework [38] to inform tailoring of the *InvolveMe* intervention and to identify key aspects for implementation planning based on context.

The identification of determinants in this study supports findings from existing literature [23,33,50]. However, the influence of

context on implementation outcomes must be considered to understand the need to tailor interventions [51]. To the best of the authors' knowledge, descriptions of how identified determinants can be used to tailor interventions to context are largely lacking. The HCPs participating in this study were mostly positive toward implementing the digital communication intervention *InvolveMe* and perceived the intervention as having the potential to improve patient-provider communication. This is in line with existing research showing that improvements in communication can act as facilitators in eHealth implementation [24,50].

In the development study [41] preceding this study, the participants voiced a concern about lack of integration with existing systems, which corresponds with findings from other studies where lack of accessibility and fit into organizational structures have been identified as barriers to implementation of eHealth interventions [23,33,50]. Therefore, the *InvolveMe* intervention was integrated into a patient portal already in use by patients and HCPs. What was initially perceived as a barrier in the development study [41] was hence turned into something, perceived by participants, beneficial in this study. This confirmed the decision made in the development study [41], acting as a potential facilitator for intervention implementation and potentially improving acceptance and adoption [33].

The HCPs in this study voiced some concerns regarding the assessment feature in terms of being a new work task that could potentially increase the workload. The assessment feature was therefore condensed to a brief list and refinements were made to the assessment summary by conducting several user tests in close collaboration with HCPs, patient participants, and the research team. This strategy is supported by the literature, showing user-friendliness and integration into care as known facilitators for the implementation of eHealth interventions [22,50,52]. In addition, this strategy can prevent the need for intervention redesign, which is likely to delay use and increase costs, which are known barriers to eHealth interventions [33].

The involvement of relevant stakeholders is a known facilitator of implementation [50,53]. Although stakeholders were involved in the development [41] of the *InvolveMe* intervention, the participants in this study provided valuable additional information for the tailoring of the intervention, including the need to test the intervention before any upcoming clinical trial. Participants in this study raised concerns about potential implications for clinical workflow and workload when using a digital communication intervention. Concerns about increased workload are a well-known barrier to the implementation of eHealth interventions [28,50]. A growing point of importance is also to tailor eHealth interventions to existing clinical workflows, minimizing potential burdens [30]. A moderator function for triaging secure messages should therefore be organized in a flexible way, depending on the clinic organization and available HCPs. Such tailoring to the local context may facilitate intervention integration into clinical workflow, a known facilitator for implementation of eHealth interventions in practice [22,50].

In this study, HCPs also expressed concerns about patient acceptance of a digital communication intervention, described as a lack of digital competence among some patients. This concern is supported by patient education literature [20,24,54]. However, there are some indications that it is feasible to deliver eHealth interventions to improve eHealth and health literacy skills among patients with chronic health conditions [22,54]. To be able to offer interventions, such as *InvolveMe* to patients, regardless of digital competence, studies have suggested employing blended care models, involving a mixture of in-person, technology, or telephone contact as a way to help facilitate use [23,55,56]. Furthermore, alternating health care delivery between digital communication and in-person meetings has been described as a way to avoid losing in-person contact with patients [23].

To ensure successful implementation of an eHealth intervention in a certain context, the need to develop and follow an implementation plan is widely recognized [36] and the lack of such a plan is considered a barrier to implementation [33]. In this study, results from the CFIR *Planning* construct provided insight into the participants' thoughts on how to involve key stakeholders, secure leadership support, and how to provide information training and coaching. These factors have previously been described as important for incorporation into an implementation plan [32,37]. In addition, several facilitators (eg, ownership, positive attitude, and system acceptance) identified in the other CFIR domains were considered important to build on (ie, leverage facilitators) when planning for implementation. HCPs struggling with the use of technology are a known barrier in the implementation of eHealth interventions [23]. Training of HCPs is therefore a preferred and widely used implementation strategy [53], and a combination of software training and training in how to incorporate the intervention into daily clinical workflow may be required [30,31]. In this study, it was hence considered an important aspect to include training and follow-up of HCPs in the implementation plan. Training of designated clinicians who would subsequently train others to use *InvolveMe* was planned [32]. External support (ie, provided by a member of the research team) was also identified as a key aspect of the implementation plan.

The heads of both clinics participating in this study were involved in the development study [41] and in the tailoring and implementation planning. Several studies have shown that implementation strategies that encourage leadership support and engagement are crucial to implementation success [30]. Leaders are often seen as providers of new knowledge and as key influencers related to implementation initiatives, including facilitating effective teamwork and cultivating a culture of learning [57]. In addition, leaders can assign dedicated staff to perform the required change, which may ease workload and the concern for increased work [57]. Therefore, strategies targeting leaders, such as continuing the project steering committee meetings, should be considered key aspects to include in the implementation plan of interventions, such as *InvolveMe*, into outpatient clinics. In addition, carrying out implementation preparation workshops (Multimedia Appendix 2), as in this study, might also capture local knowledge of what works and not, knowledge that can be shared between implementation sites [32].

In this study, the project steering committee meetings were also intended to build a coalition between patient participants, HCPs, and the research team to cultivate a good relationship in the implementation effort, another described implementation strategy [32,37,53], and thus a key aspect to include in an implementation plan. Collaborative relationships are crucial for implementing plans and, through a social exchange communicating the potential impact of innovations, for the implementation of interventions, such as *InvolveMe*, into clinical practice may be facilitated [57].

Strengths and Limitations

This study has some limitations that need to be addressed. *First*, the study was conducted at a single university hospital. This might limit transferability to other settings. However, the inclusion of 2 outpatient clinics, following up 2 different patient categories, might increase transferability to other settings. *Second*, the study had a relatively small sample size, a factor potentially limiting transferability. A small sample size may limit the ability of data to describe the entire local context at the clinics involved. However, the study was performed in outpatient clinics where the implementation of the intervention is planned to take place, thus enabling identification of local determinants and key aspects that may be crucial for an implementation plan. *Third*, there was an imbalance in the number of HCPs compared with patients in this study. Traditionally, implementation involves the improvement of HCP practices [35]. CFIR does not differ from this tradition, as CFIR places patients under the *Outer Setting* domain, where only one single construct is intended to capture patients' needs and resources [38]. As such, patients are considered to have a peripheral role in their implementation. However, this study examined all perspectives and interplay of all stakeholders involved, including patient participants, which helped to illuminate stakeholder aspects and may help increase the likelihood of successful implementation to practice [33]. In addition, even with a limited number of patient participants, patient participants' experience from the development study [41] and subsequent direct input on various topics in this study may have ensured relevance and reliability from the patient perspective. *Fourth*, the perceived facilitators and barriers of participants in this study might not necessarily correspond to facilitators and barriers experienced in clinical practice in general. However, recent evidence indicates that the limited use of tailoring to context could explain the limited implementation success [30]. Knowledge generated during preparation for implementation can contribute to intervention tailoring and context-specific individualization, which implies that stakeholders' needs are more likely met, and hence intervention design and implementation preparation are improved [58].

This study has some strengths. Applying a structured and comprehensive framework such as CFIR within the field of implementation is considered to be a strength guiding data collection and analysis [59]. Identifying and describing determinants that affect implementation, as well as identifying key aspects for implementation planning, are also strengths [59]. Furthermore, the use of CFIR in this study provided a common language through the use of constructs and definitions for the analysis of data and thus may provide comparable results that may make it easier to assess why and how certain elements work. However, it should be noted that the strength of CFIR as a comprehensive framework may also be a weakness. CFIR does not distinguish between the relative importance of all of its constructs, which may imply that the details necessary for implementation success could be lost if trying to capture as many constructs as possible. In this study, a number of CFIR constructs were not covered by the data collection, as only discrete but significant themes and questions were targeted (Multimedia Appendices 1 and 2). Another challenge using CFIR is that some constructs are broad and difficult to capture and some may overlap with other constructs. Further descriptions and explanations related to constructs could enhance the CFIR and thus make the framework more intuitive to use when planning for implementation.

Future Directions

The ongoing COVID-19 pandemic has triggered a significant need for a wide range of digital communication services between patients and HCPs and has led to increased demand for digital intervention from within the health care services themselves. As such, and with the tremendous challenges posed by this significant health challenge, the pandemic might turn out to be a powerful facilitator for the implementation of digital interventions in health care services.

This study revealed some specific aspects that need to be investigated in future research. In particular, the results show that a pilot study may contribute to identify gaps and inform further necessary tailoring of the intervention and an implementation plan (ie, strategies) before clinical trials. This emphasizes that, regardless of stakeholder involvement in intervention development, a pilot test should always be considered.

Future studies should also aim to better understand how the CFIR framework can inform an implementation planning process in terms of tailoring interventions before implementation and the selection of implementation strategies based on identified determinants. In addition, refinements of the CFIR to strengthen the patient-related constructs and make the framework easier to apply would be beneficial for researchers and for HCPs conducting implementation in clinical practice.

Conclusions

This study contributes to the field of implementation science by using identified determinants to inform the tailoring of a digital communication intervention (ie, *InvolveMe*) and to identify key aspects of an implementation plan to context. Important intervention tailoring aspects identified were adjustments to the digital features and adjustments to fit the clinical workflow as well as recommendations to conduct a future pilot study before testing in larger clinical trials. Future research into the implementation of digital communication interventions should focus on the early identification of determinants and attention to tailoring to address barriers and leverage facilitators. In addition, key aspects of implementation planning should be identified, raising the probability of implementation success.

Acknowledgments

This study was funded by the Norwegian Nurses Organization grant 847382 (principal investigator: EB). The authors would like to thank the patient participants and HCPs who participated in this study. The authors would also like to thank all collaborating project team partners, including the editorial group and software team in the Department of Digital Health Research at Oslo University Hospital.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Focus group guide with Consolidated Framework for Implementation Research domains.
[PDF File (Adobe PDF File), 98 KB - jmir_v23i4e22399_app1.pdf]

Multimedia Appendix 2
Workshop theme guide with Consolidated Framework for Implementation Research domains.
[PDF File (Adobe PDF File), 144 KB - jmir_v23i4e22399_app2.pdf]

References

1. Burström K, Johannesson M, Diderichsen F. Health-related quality of life by disease and socio-economic group in the general population in Sweden. Health Policy 2001 Jan;55(1):51-69. [doi: 10.1016/s0168-8510(00)00111-1] [Medline: 11137188]
2. Saarni SI, Härkänen T, Sintonen H, Suvisaari J, Koskinen S, Aromaa A, et al. The impact of 29 chronic conditions on health-related quality of life: a general population survey in Finland using 15D and EQ-5D. Qual Life Res 2006 Oct;15(8):1403-1414. [doi: 10.1007/s11136-006-0020-1] [Medline: 16960751]
3. Heyworth ITM, Hazell ML, Linehan MF, Frank TL. How do common chronic conditions affect health-related quality of life? Br J Gen Pract 2009 Nov;59(568):353-358 [FREE Full text] [doi: 10.3399/bjgp09X453990] [Medline: 19656444]
4. Wu M, Brazier JE, Kearns B, Relton C, Smith C, Cooper CL. Examining the impact of 11 long-standing health conditions on health-related quality of life using the EQ-5D in a general population sample. Eur J Health Econ 2015 Mar;16(2):141-151 [FREE Full text] [doi: 10.1007/s10198-013-0559-z] [Medline: 24408476]
5. May CR, Cummings A, Myall M, Harvey J, Pope C, Griffiths P, et al. Experiences of long-term life-limiting conditions among patients and carers: what can we learn from a meta-review of systematic reviews of qualitative studies of chronic heart failure, chronic obstructive pulmonary disease and chronic kidney disease? BMJ Open 2016 Oct 05;6(10) [FREE Full text] [doi: 10.1136/bmjopen-2016-011694] [Medline: 27707824]
6. Lubberding S, van Uden-Kraan CF, Te Velde EA, Cuijpers P, Leemans CR, Leeuw IMV. Improving access to supportive cancer care through an eHealth application: a qualitative needs assessment among cancer survivors. J Clin Nurs 2015 May;24(9-10):1367-1379. [doi: 10.1111/jocn.12753] [Medline: 25677218]
7. Friis K, Lasgaard M, Osborne RH, Maindal HT. Gaps in understanding health and engagement with healthcare providers across common long-term conditions: a population survey of health literacy in 29,473 Danish citizens. BMJ Open 2016 Jan 14;6(1) [FREE Full text] [doi: 10.1136/bmjopen-2015-009627] [Medline: 26769783]
8. Goldzweig CL, Towfigh AA, Paige NM, Orshansky G, Haggstrom DA, Beroes JM, et al. Systematic review: secure messaging between providers and patients, and patients' access to their own medical record: evidence on health outcomes, satisfaction, efficiency and attitudes. VA Evidence-based Synthesis Program Reports 2012. [Medline: 22973584]
9. Heyn L, Ruland CM, Finset A. Effects of an interactive tailored patient assessment tool on eliciting and responding to cancer patients' cues and concerns in clinical consultations with physicians and nurses. Patient Educ Couns 2012 Feb;86(2):158-165. [doi: 10.1016/j.pec.2011.04.024] [Medline: 21592719]
10. Heyn L, Finset A, Eide H, Ruland CM. Effects of an interactive tailored patient assessment on patient-clinician communication in cancer care. Psychooncology 2013 Jan;22(1):89-96. [doi: 10.1002/pon.2064] [Medline: 21919122]
11. Johansen MA, Berntsen GKR, Schuster T, Henriksen E, Horsch A. Electronic symptom reporting between patient and provider for improved health care service quality: a systematic review of randomized controlled trials. part 2: methodological quality and effects. J Med Internet Res 2012 Oct 03;14(5):e126 [FREE Full text] [doi: 10.2196/jmir.2216] [Medline: 23032363]
12. Whitehead L, Seaton P. The effectiveness of self-management mobile phone and tablet apps in long-term condition management: a systematic review. J Med Internet Res 2016 May 16;18(5):e97 [FREE Full text] [doi: 10.2196/jmir.4883] [Medline: 27185295]

13. de Jong CC, Ros WJ, Schrijvers G. The effects on health behavior and health outcomes of Internet-based asynchronous communication between health providers and patients with a chronic condition: a systematic review. J Med Internet Res 2014 Jan 16;16(1):e19 [FREE Full text] [doi: 10.2196/jmir.3000] [Medline: 24434570]
14. Børøsund E, Cvancarova M, Moore SM, Ekstedt M, Ruland CM. Comparing effects in regular practice of e-communication and web-based self-management support among breast cancer patients: preliminary results from a randomized controlled trial. J Med Internet Res 2014 Dec 18;16(12):e295 [FREE Full text] [doi: 10.2196/jmir.3348] [Medline: 25525672]
15. Qudah B, Luetsch K. The influence of mobile health applications on patient - healthcare provider relationships: a systematic, narrative review. Patient Educ Couns 2019 Jun;102(6):1080-1089. [doi: 10.1016/j.pec.2019.01.021] [Medline: 30745178]
16. Lee JA, Choi M, Lee SA, Jiang N. Effective behavioral intervention strategies using mobile health applications for chronic disease management: a systematic review. BMC Med Inform Decis Mak 2018 Feb 20;18(1):12 [FREE Full text] [doi: 10.1186/s12911-018-0591-0] [Medline: 29458358]
17. Chen J, Ou L, Hollis SJ. A systematic review of the impact of routine collection of patient reported outcome measures on patients, providers and health organisations in an oncologic setting. BMC Health Serv Res 2013 Jun 11;13:211 [FREE Full text] [doi: 10.1186/1472-6963-13-211] [Medline: 23758898]
18. Ruland CM, Holte HH, Røislien J, Heaven C, Hamilton GA, Kristiansen J, et al. Effects of a computer-supported interactive tailored patient assessment tool on patient care, symptom distress, and patients' need for symptom management support: a randomized clinical trial. J Am Med Inform Assoc 2010;17(4):403-410 [FREE Full text] [doi: 10.1136/jamia.2010.005660] [Medline: 20595307]
19. Ruland CM, White T, Stevens M, Fanciullo G, Khilani SM. Effects of a computerized system to support shared decision making in symptom management of cancer patients: preliminary results. J Am Med Inform Assoc 2003;10(6):573-579 [FREE Full text] [doi: 10.1197/jamia.M1365] [Medline: 12925545]
20. Jacobs RJ, Lou JQ, Ownby RL, Caballero J. A systematic review of eHealth interventions to improve health literacy. Health Informatics J 2016 Jun;22(2):81-98 [FREE Full text] [doi: 10.1177/1460458214534092] [Medline: 24916567]
21. eHealth Action Plan 2012-2020: innovative healthcare for the 21st century. 2012. URL: https://ec.europa.eu/digital-agenda/en/news/ehealth-action-plan-2012-2020-innovative-healthcare-21st-century [accessed 2021-03-22]
22. Keyworth C, Hart J, Armitage CJ, Tully MP. What maximizes the effectiveness and implementation of technology-based interventions to support healthcare professional practice? A systematic literature review. BMC Med Inform Decis Mak 2018 Nov 07;18(1):93 [FREE Full text] [doi: 10.1186/s12911-018-0661-3] [Medline: 30404638]
23. Kruse CS, Karem P, Shifflett K, Vegi L, Ravi K, Brooks M. Evaluating barriers to adopting telemedicine worldwide: a systematic review. J Telemed Telecare 2018 Jan;24(1):4-12 [FREE Full text] [doi: 10.1177/1357633X16674087] [Medline: 29320966]
24. Kruse CS, Argueta DA, Lopez L, Nair A. Patient and provider attitudes toward the use of patient portals for the management of chronic disease: a systematic review. J Med Internet Res 2015 Feb 20;17(2):e40 [FREE Full text] [doi: 10.2196/jmir.3703] [Medline: 25707035]
25. Varsi C, Ekstedt M, Gammon D, Ruland CM. Using the consolidated framework for implementation research to identify barriers and facilitators for the implementation of an internet-based patient-provider communication service in five settings: a qualitative study. J Med Internet Res 2015 Nov 18;17(11):e262 [FREE Full text] [doi: 10.2196/jmir.5091] [Medline: 26582138]
26. Børøsund E, Ruland CM, Moore S, Ekstedt M. Nurses' experiences of using an interactive tailored patient assessment tool one year past implementation. Int J Med Inform 2014 Jul;83(7):23-34. [doi: 10.1016/j.ijmedinf.2013.10.010] [Medline: 24268722]
27. Chang F, Paramsothy T, Roche M, Gupta NS. Patient, staff, and clinician perspectives on implementing electronic communications in an interdisciplinary rural family health practice. Prim Health Care Res Dev 2017 Mar;18(2):149-160. [doi: 10.1017/S1463423616000414] [Medline: 27995826]
28. Antoun J. Electronic mail communication between physicians and patients: a review of challenges and opportunities. Fam Pract 2016 Apr;33(2):121-126. [doi: 10.1093/fampra/cmv101] [Medline: 26711957]
29. Kummervold PE, Johnsen JK. Physician response time when communicating with patients over the internet. J Med Internet Res 2011 Nov 01;13(4):e79 [FREE Full text] [doi: 10.2196/jmir.1583] [Medline: 22044909]
30. Varsi C, Nes LS, Kristjansdottir OB, Kelders SM, Stenberg U, Zangi HA, et al. Implementation strategies to enhance the implementation of ehealth programs for patients with chronic illnesses: realist systematic review. J Med Internet Res 2019 Sep 27;21(9) [FREE Full text] [doi: 10.2196/14255] [Medline: 31573934]
31. Elbert NJ, van Os-Medendorp H, van Renselaar W, Ekeland AG, Hakkaart-van Roijen L, Raat H, et al. Effectiveness and cost-effectiveness of ehealth interventions in somatic diseases: a systematic review of systematic reviews and meta-analyses. J Med Internet Res 2014 Apr 16;16(4):e110 [FREE Full text] [doi: 10.2196/jmir.2790] [Medline: 24739471]
32. Powell BJ, Waltz TJ, Chinman MJ, Damschroder LJ, Smith JL, Matthieu MM, et al. A refined compilation of implementation strategies: results from the Expert Recommendations for Implementing Change (ERIC) project. Implement Sci 2015 Feb 12;10:21 [FREE Full text] [doi: 10.1186/s13012-015-0209-1] [Medline: 25889199]

33. Ross J, Stevenson F, Lau R, Murray E. Factors that influence the implementation of e-health: a systematic review of systematic reviews (an update). Implement Sci 2016 Oct 26;11(1):146 [FREE Full text] [doi: 10.1186/s13012-016-0510-7] [Medline: 27782832]
34. Chaudhry B, Wang J, Wu S, Maglione M, Mojica W, Roth E, et al. Systematic review: impact of health information technology on quality, efficiency, and costs of medical care. Ann Intern Med 2006 May 16;144(10):742-752 [FREE Full text] [doi: 10.7326/0003-4819-144-10-200605160-00125] [Medline: 16702590]
35. Nilsen P. Making sense of implementation theories, models and frameworks. Implement Sci 2015 Apr 21;10:53 [FREE Full text] [doi: 10.1186/s13012-015-0242-0] [Medline: 25895742]
36. Proctor EK, Powell BJ, McMillen JC. Implementation strategies: recommendations for specifying and reporting. Implement Sci 2013 Dec 01;8:139 [FREE Full text] [doi: 10.1186/1748-5908-8-139] [Medline: 24289295]
37. Waltz TJ, Powell BJ, Matthieu MM, Damschroder LJ, Chinman MJ, Smith JL, et al. Use of concept mapping to characterize relationships among implementation strategies and assess their feasibility and importance: results from the Expert Recommendations for Implementing Change (ERIC) study. Implement Sci 2015 Aug 07;10:109 [FREE Full text] [doi: 10.1186/s13012-015-0295-0] [Medline: 26249843]
38. Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. Implement Sci 2009 Aug 07;4:50 [FREE Full text] [doi: 10.1186/1748-5908-4-50] [Medline: 19664226]
39. Keith RE, Crosson JC, O'Malley AS, Cromp D, Taylor EF. Using the Consolidated Framework for Implementation Research (CFIR) to produce actionable findings: a rapid-cycle evaluation approach to improving implementation. Implement Sci 2017 Feb 10;12(1):15 [FREE Full text] [doi: 10.1186/s13012-017-0550-7] [Medline: 28187747]
40. Kirk MA, Kelley C, Yankey N, Birken SA, Abadie B, Damschroder L. A systematic review of the use of the Consolidated Framework for Implementation Research. Implement Sci 2016 May 17;11:72 [FREE Full text] [doi: 10.1186/s13012-016-0437-z] [Medline: 27189233]
41. Seljelid B, Varsi C, Solberg Nes L, Stenehjem A, Bollerslev J, Børøsund E. Content and system development of a digital patient-provider communication tool to support shared decision making in chronic health care: InvolveMe. BMC Med Inform Decis Mak 2020 Mar 04;20(1):46 [FREE Full text] [doi: 10.1186/s12911-020-1065-8] [Medline: 32131808]
42. Jagosh J, Macaulay AC, Pluye P, Salsberg J, Bush PL, Henderson J, et al. Uncovering the benefits of participatory research: implications of a realist review for health research and practice. Milbank Q 2012 Jun;90(2):311-346 [FREE Full text] [doi: 10.1111/j.1468-0009.2012.00665.x] [Medline: 22709390]
43. Robertson T, Wagner I. Ethics: engagement, representation and politics-in-action. In: Simonsen J, Robertson T, editors. Routledge International Handbook of Participartory Design - 1st Edition. London: Routledge; 2013.
44. Morgan DL. Focus group interviewing. In: Gubrium JF, Holstein JA, editors. Handbook of Interview Research. Thousand Oaks, CA: SAGE Publications Ltd; 2001:141-159.
45. Stickdorn M, Schneider J. This is service design thinking: basics, tools, cases. Hoboken, New Jersey, United States: Wiley; 2012:1-384.
46. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psycho 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
47. Clarke V, Braun V. Thematic analysis. J Posit Psychol 2016 Dec 09;12(3):297-298. [doi: 10.1080/17439760.2016.1262613]
48. Acts of legislation. Regional Committees for Medical and Health Research Ethics. 2020. URL: https://helseforskning.etikkom.no/reglerogrutiner/loverogregler?p_dim=34770&_ikbLanguageCode=us [accessed 2021-03-22]
49. Seljelid B, Varsi C, Stenehjem AE, Bollerslev J, Solberg Nes L, Børøsund E. Content development of an electronic shared decision making tool for patients with chronic illness; mapping symptoms and needs. Society of Behavioral Medicine 2019 Annual Meeting Abstracts. Annals of Behavioral Medicine 2019 Mar;53(1):S1-S842 [FREE Full text] [doi: https://doi.org/10.1093/abm/kaz007]
50. Schreiweis B, Pobiruchin M, Strotbaum V, Suleder J, Wiesner M, Bergh B. Barriers and facilitators to the implementation of ehealth services: systematic literature analysis. J Med Internet Res 2019 Nov 22;21(11) [FREE Full text] [doi: 10.2196/14197] [Medline: 31755869]
51. Nilsen P, Bernhardsson S. Context matters in implementation science: a scoping review of determinant frameworks that describe contextual determinants for implementation outcomes. BMC Health Serv Res 2019 Mar 25;19(1):189 [FREE Full text] [doi: 10.1186/s12913-019-4015-3] [Medline: 30909897]
52. Kruse CS, Bolton K, Freriks G. The effect of patient portals on quality outcomes and its implications to meaningful use: a systematic review. J Med Internet Res 2015 Feb 10;17(2):e44 [FREE Full text] [doi: 10.2196/jmir.3171] [Medline: 25669240]
53. Albers B, Metz A, Burke K, Bührmann L, Bartley L, Driessen P, et al. Implementation support skills: findings from a systematic integrative review. Research on Social Work Practice 2020 Oct 27;31(2):147-170 [FREE Full text] [doi: 10.1177/1049731520967419]
54. Chapman E, Haby MM, Toma TS, de Bortoli MC, Illanes E, Oliveros MJ, et al. Knowledge translation strategies for dissemination with a focus on healthcare recipients: an overview of systematic reviews. Implement Sci 2020 Mar 04;15(1):14 [FREE Full text] [doi: 10.1186/s13012-020-0974-3] [Medline: 32131861]

55. Solem IKL, Varsi C, Eide H, Kristjansdottir OB, Børøsund E, Schreurs KMG, et al. A user-centered approach to an evidence-based electronic health pain management intervention for people with chronic pain: design and development of EPIO. J Med Internet Res 2020 Jan 21;22(1) [FREE Full text] [doi: 10.2196/15889] [Medline: 31961331]
56. Børøsund E, Mirkovic J, Clark MM, Ehlers SL, Andrykowski MA, Bergland A, et al. A stress management app intervention for cancer survivors: design, development, and usability testing. JMIR Form Res 2018 Sep 06;2(2):e19 [FREE Full text] [doi: 10.2196/formative.9954] [Medline: 30684438]
57. Li S, Jeffs L, Barwick M, Stevens B. Organizational contextual features that influence the implementation of evidence-based practices across healthcare settings: a systematic integrative review. Syst Rev 2018 May 05;7(1):72 [FREE Full text] [doi: 10.1186/s13643-018-0734-5] [Medline: 29729669]
58. Cranley LA, Cummings GG, Profetto-McGrath J, Toth F, Estabrooks CA. Facilitation roles and characteristics associated with research use by healthcare professionals: a scoping review. BMJ Open 2017 Aug 11;7(8) [FREE Full text] [doi: 10.1136/bmjopen-2016-014384] [Medline: 28801388]
59. Moullin JC, Sabater-Hernández D, Fernandez-Llimos F, Benrimoj SI. A systematic review of implementation frameworks of innovations in healthcare and resulting generic implementation framework. Health Res Policy Syst 2015 Mar 14;13:16 [FREE Full text] [doi: 10.1186/s12961-015-0005-z] [Medline: 25885055]

Abbreviations

CFIR: Consolidated Framework for Implementation Research
EPR: electronic patient record
HCP: health care provider
HRQoL: health-related quality of life

Edited by G Eysenbach, R Kukafka; submitted 13.07.20; peer-reviewed by C Pithara, C Jacob, G Cohen; comments to author 04.08.20; revised version received 09.10.20; accepted 18.03.21; published 08.04.21.

Please cite as:
Seljelid B, Varsi C, Solberg Nes L, Øystese KA, Børøsund E
A Digital Patient-Provider Communication Intervention (InvolveMe): Qualitative Study on the Implementation Preparation Based on Identified Facilitators and Barriers
J Med Internet Res 2021;23(4):e22399
URL: https://www.jmir.org/2021/4/e22399
doi:10.2196/22399
PMID:33830063

Original Paper

Perspectives of Inpatients With Cirrhosis and Caregivers on Using Health Information Technology: Cross-sectional Multicenter Study

Chathur Acharya[1,2], MD; Tejasav S Sehrawat[3], MD; Deborah B McGuire[1,2], PhD; Jawaid Shaw[1,2], MD; Andrew Fagan[1,2], BS; Sara McGeorge[1,2], BS; Amy Olofson[3], RN; Melanie B White[1,2], RN; Edith Gavis[1,2], RN; Patrick S Kamath[3], MD; Lori Bergstrom[3], BSc; Jasmohan Singh Bajaj[1,2], MD

[1]Division of Gastroenterology, Hepatology and Nutrition, Virginia Commonwealth University, Richmond, VA, United States

[2]McGuire VA Medical Center, Richmond, VA, United States

[3]Division of Gastroenterology and Hepatology, Mayo Clinic School of Medicine, Rochester, MN, United States

Corresponding Author:
Jasmohan Singh Bajaj, MD
Division of Gastroenterology, Hepatology and Nutrition
Virginia Commonwealth University
1201 Broad Rock Boulevard
Richmond, VA, 23249
United States
Phone: 1 804 675 5802
Email: jasmohan.bajaj@vcuhealth.org

Abstract

Background: Health information technology (IT) interventions to decrease readmissions for cirrhosis may be limited by patient-associated factors.

Objective: The aim of this study was to determine perspectives regarding adoption versus refusal of health IT interventions among patient-caregiver dyads.

Methods: Inpatients with cirrhosis and their caregivers were approached to participate in a randomized health IT intervention trial requiring daily contact with research teams via the Patient Buddy app. This app focuses on ascites, medications, and hepatic encephalopathy over 30 days. Regression analyses for characteristics associated with acceptance were performed. For those who declined, a semistructured interview was performed with themes focused on caregivers, protocol, transport/logistics, technology demands, and privacy.

Results: A total of 349 patient-caregiver dyads were approached (191 from Virginia Commonwealth University, 56 from Richmond Veterans Affairs Medical Center, and 102 from Mayo Clinic), 87 of which (25%) agreed to participate. On regression, dyads agreeing included a male patient (odds ratio [OR] 2.08, *P*=.01), gastrointestinal bleeding (OR 2.3, *P*=.006), or hepatic encephalopathy admission (OR 2.0, *P*=.01), whereas opioid use (OR 0.46, *P*=.03) and alcohol-related etiology (OR 0.54, *P*=.02) were associated with refusal. Race, study site, and other admission reasons did not contribute to refusing participation. Among the 262 dyads who declined randomization, caregiver reluctance (43%), perceived burden (31%), technology-related issues (14%), transportation/logistics (10%), and others (4%), but not privacy, were highlighted as major concerns.

Conclusions: Patients with cirrhosis admitted with hepatic encephalopathy and gastrointestinal bleeding without opioid use or alcohol-related etiologies were more likely to participate in a health IT intervention focused on preventing readmissions. Caregiver and study burden but not privacy were major reasons to decline participation. Reducing perceived patient-caregiver burden and improving communication may improve participation.

Trial Registration: ClinicalTrials.gov NCT03564626; https://www.clinicaltrials.gov/ct2/show/NCT03564626

(J Med Internet Res 2021;23(4):e24639) doi:10.2196/24639

KEYWORDS

hepatic encephalopathy; cirrhosis; outcomes; acceptance; PatientBuddy; ascites; readmissions; hepatic; encephalopathy

Introduction

Patients with cirrhosis often require expensive inpatient and outpatient care, mainly centered around complications related to hepatic encephalopathy, variceal bleeding, ascites, and medication management [1]. A critical component of optimal care is the education of patients and their caregivers, and communication with the clinical team [2]. Educational strategies for and involvement of caregivers in care processes are often neglected in favor of dispensing medications and scheduling appointments at the time of discharge [3,4]. Tools that could enhance delivery of care and maintain a link between health care providers and patients/caregivers are needed to improve clinical management and outcomes. In a prior single-center study, we evaluated the acceptability and performance of the Patient Buddy app for preventing readmissions in patients with cirrhosis [5]. The app involves two components recorded by patients and their caregivers on EncephalApp, a separate app designed to assess cognitive function in patients with cirrhosis: (1) daily postdischarge smartphone entry for 30 days of weight, abdominal girth, temperature, and medication adherence; and (2) weekly results of responses to questions related to orientation in time, place, and person [6,7]. The data are transmitted via a Health Insurance Portability and Accountability Act (HIPAA)-compliant server to central iPads, which the study team then review for assessment of alarm values that are used for communication with patients and caregivers and to determine if action is required. The central iPads are reviewed multiple times a day. In our single-center study, we found that use of the app alerted health care providers to early evolution of hepatic encephalopathy, and was largely favorably received by the patients and caregivers using the app [5].

We are currently performing a multicenter randomized trial of the Patient Buddy app at three clinical sites in patients with cirrhosis and their caregivers. The purpose of this study was to examine the perspectives of both patients and caregivers when considering use of this intensive daily health information technology (IT). Their willingness to use the app has direct bearing on patient participation in this and other multisite trials, and ultimately on the health care team's ability to monitor clinical status, reduce readmissions, and improve outcomes in cirrhosis.

Methods

We designed this study to assess the reasons for potential refusal to better inform recruitment approaches. We performed semistructured interviews in patient-caregiver dyads in the ongoing multicenter randomized study of cirrhosis inpatients and their caregivers on use of the Patient Buddy app to prevent 30-day readmissions. The randomized study is being performed in tertiary-care hospitals serving three different populations: veterans at Veterans Affairs (VA) Medical Center (Richmond, VA), uninsured/underinsured patients at a state safety net institution (Virginia Commonwealth University [VCU], Richmond, VA), and insured patients receiving care at a private group practice (Mayo Clinic, Rochester, MN). Regardless of the setting, the standard of care for all three hospitals requires setting up a postdischarge follow-up appointment within 1 month and a phone call within 1 week.

We screened for inpatients who met the following eligibility criteria: (1) diagnosis of cirrhosis as defined by liver biopsy, overt hepatic decompensation (hepatic encephalopathy, ascites, variceal bleeding, or jaundice), or radiological features of cirrhosis and endoscopic features of varices in patients with chronic liver disease; and (2) those who were admitted nonelectively and had adult cohabitating caregivers. Patients excluded for screening were those with an active alcohol use disorder, without caregivers or a stable home, on hemodialysis, discharged to a facility or hospice, inadequate internet connection at home, or an unclear diagnosis of cirrhosis. For patients who met the eligibility criteria, a member of the study team approached the patient and caregiver in person or via telephone while the patient was still hospitalized and explained the study in detail. This explanation included a discussion with the patient-caregiver dyad that emphasized the necessity of being in contact with the study team after discharge, and the potential educational value of the app and its implications for communication with the team on clinical issues. Another aspect of the discussion involved a demonstration of the app and the need for data entry of critical medications (ie, hepatic encephalopathy medications, antibiotics, beta blockers, diuretics, and other cirrhosis-related medications), weight and abdominal girth, and orientation with EncephalApp.

The intervention consisted of the Patient Buddy app coupled with EncephalApp. For the dyad to participate in the multisite randomized trial postdischarge, they had to agree to be trained individually while the patient was hospitalized, including instructions for data entry in the app; measuring body weight, temperature, and girth (using scales, measuring tapes, and thermometers provided by the study team, respectively); assessing orientation; and training to caregivers in administration of EncephalApp. If the dyad consented to participate, they were randomly assigned to one of three groups (standard of care, health IT intervention with as-needed follow-up, or health IT intervention with scheduled outpatient visits and calls within 30 days of discharge). If assigned to either of the health IT groups, the app was activated upon discharge on a smartphone provided as part of the study.

In the randomized study, we collect data from all dyads who are approached for participation, including demographic and clinical characteristics, reasons for admission, and current patient medications. Dyads indicating that they did not want to be randomized were then asked if they would participate in a semistructured qualitative interview that was designed and supervised by an expert (DM) to explore reasons for deciding not to participate and to help ascertain issues related to use of health IT as opposed to research as a whole that could guide future similar studies.

The patients and caregivers were asked about their reasons for unwillingness divided into five categories (demands related to technology, caregiver reluctance, perceived burden of the study, transport and logistics, privacy-related, or other) and other reasons through the interview questions prompted by the coordinators with qualitative responses allowed. All potential

inpatients were approached in person. The study coordinators carried out the interview, and every effort was made to perform these analyses within 1 day for both potential caregivers and patients. The responses were transcribed on an interview form by the study coordinator explaining the study, which also had space for expressing other potential reasons that informed their decision not to participate further. The other reasons included refusal to participate in studies of any kind regardless of the health IT intervention, unspecified, or unwillingness to elaborate further.

In the analysis, we compared the demographic and clinical variables such as cirrhosis characteristics, especially hepatic encephalopathy, current medications, and reasons for admission, in dyads who agreed to participate with those of dyads who did not agree to participate using unpaired *t* tests or χ^2 tests as appropriate. Finally, a logistic regression analysis was performed with agreement to participate in the randomization process as the dependent variable and disease characteristics as independent variables.

Results

A total of 349 patient-caregiver dyads were approached at the three sites (191 at VCU, 56 at VA, and 102 at Mayo Clinic), 87 of which (24.9%) agreed to participate. Table 1 shows the characteristics of these dyads and indicates the significantly different variables between those who agreed and declined participation.

Table 1. Demographic and clinical characteristics of patients who agreed and did not agree to randomization for the health information technology intervention study.

Characteristics	Agreed to randomization (n=87)	Did not agree to randomization (n=262)	*P* value
Age (years), mean (SD)	58.8 (11.3)	59.2 (11.4)	.78
Race, n (%)			.06
Caucasian	64 (74)	207 (79.0)	
African American	15 (17)	47 (17.9)	
Other	8 (9)	8 (3.1)	
Men, n (%)	62 (71)	150 (57.3)	.02
Comorbidity, n (%)			.03
HCV[a]	5 (6)	28 (10.7)	
HCV+alcohol-related	26 (30)	117 (44.7)	
Alcohol-related	9 (10)	15 (5.7)	
NAFLD[b]	25 (29)	60 (22.9)	
Other	22 (25)	42 (16.0)	
Prior HE[c], n (%)	56 (64)	151 (57.6)	.27
Lactulose use, n (%)	50 (57)	152 (58.0)	.93
Rifaximin use, n (%)	40 (46)	101 (38.5)	.22
Taking psychoactive medications, n (%)	12 (14)	57 (21.8)	.09
Taking opioids, n (%)	11 (13)	71 (27.1)	.004
MELD[d] score on admission, mean (SD)	19.5 (6.9)	18.4 (8.4)	.24
Reason for admission, n (%)			
Infection	29 (33)	81 (30.9)	.68
HE	38 (44)	82 (31.3)	.03
Renal or metabolic disease	29 (33)	97 (37.0)	.58
Gastrointestinal bleeding	28 (32)	50 (19.1)	.01
Other	21 (24)	144 (55.0)	<.001

[a]HCV: hepatitis C virus.

[b]NAFLD: nonalcoholic fatty liver disease.

[c]HE: hepatic encephalopathy.

[d]MELD: Model for End-stage Liver Disease.

Of these 87 dyads, 40 were from VCU, 22 from VA, and 25 from Mayo Clinic. The 262 dyads (75.1%) who decided not to move forward to the randomization process all agreed to participate in the semistructured interview. Of this group, 151

were from VCU, 34 from VA, and 77 from Mayo Clinic. Results from the semistructured interview revealed that 206 dyads had only one reason for not participating, 50 had two reasons, and 6 had three or more reasons. Of these, the majority of the dyads (198/262, 75.6%) agreed to having the interviews performed in person, while the remaining 8 (3.1%) had the potential caregiver interviewed on the phone within 1 day of initial in-person contact with the patient. The most common reason was caregiver reluctance (114/262, 43.5%), followed by perceived study burden (n=82, 31.3%), technology-related issues (n=36, 13.7%), transportation and logistics (n=26, 9.9%), and no specific reason (n=11, 4.2%). Importantly, none of the dyads listed privacy as a concern for their refusal.

Logistic regression was performed using dyads that said "yes" to being randomized and therefore were included in the multisite trial. The significant factors associated with dyads saying yes (Table 1) included male gender (odds ratio [OR] 2.08, *P*=.01), admission for gastrointestinal bleeding (OR 2.3, *P*=.006), or hepatic encephalopathy (OR 2.0, *P*=.01), whereas opioids (OR 0.46, *P*=.03) and alcohol-related etiology (OR 0.54, *P*=.02) showed the reverse pattern. Race, site of recruitment, other reasons for admission, whether or not the potential caregiver was a spouse, education, and use of other medications were not significantly associated with saying "yes" to be randomized.

In the 262 dyads who declined to be randomized, we compared characteristics of patients who refused for specific reasons (Tables 2 and 3). We did not compare those who had some part of this interview performed in person versus telephone due to the small number (n=8) of those interviewed solely by telephone. Of the potential caregivers, the majority were wives (n=90, 34.4%), followed by husbands (n=55, 21.0%), children (n=30, 11.5%), significant others (n=8, 3.1%), parents (n=18, 6.9%), siblings (n=16, 6.1%), and other people (n=45, 17.2%). Of the 262 patients who did not want to participate, 116 (44.3%) had an education below or at high school level, 113 (43.1%) patients had completed undergraduate college education, and 50 (19.1%) patients had masters or higher degrees.

For analyses, we combined spouse/significant others (n=153, 58.4%) and those with high school education or below for comparison with the others. Apart from technology-related demands, there were no significant differences in the demographic variables, cirrhosis severity or etiology, medication use, hepatic encephalopathy status, or reason for admission among patients who refused compared to the rest. Dyads who refused for technology-related reasons were more likely to be older, with lower education in the patient, had a higher Model for End-stage Liver Disease (MELD) score on admission, higher frequency of hepatic encephalopathy, and were less likely to be admitted for gastrointestinal bleeding or liver-unrelated (other) reasons compared with those who did not refuse for technology-related reasons (Table 2). Caregiver reluctance was lower when the potential caregiver was a spouse, but the reverse was found for perceived study burden. Very few dyads expressed that they were refusing randomization because they were not interested in research, which was included in the "no reason" category.

Table 2. Reasons for refusal to participate in relation to patient characteristics (N=262).

Characteristic	Caregiver reluctance		Perceived study burden		Technology demands	
	Yes (n=114)	No (n=148)	Yes (n=82)	No (n=180)	Yes (n=36)	No (n=226)
Age (years)[a], mean (SD)	58.8 (11.8)	59.5 (11.1)	58.0 (11.9)	59.7 (11.2)	62.9 (9.3)	58.7 (11.4)
Race, n (%)						
Caucasian	88 (77.2)	89 (60.1)	68 (83)	139 (77.2)	26 (72)	181 (80.1)
African American	26 (22.8)	21 (14.2)	11 (13)	36 (20.0)	10 (28)	37 (16.4)
Other	4 (3.5)	4 (2.7)	3 (4)	5 (2.8)	0 (0)	8 (3.5)
Men, n (%)	66 (57.9)	84 (56.8)	45 (55)	105 (58.3)	19 (53)	131 (58.0)
High school education or above, n (%)[b]	44 (38.6)	55 (37.2)	32 (39)	84 (46.7)	20 (56)	86 (38.1)
Potential caregiver not spouse, n (%)[c]	49 (43.0)	104 (70.3)	58 (71)	95 (52.8)	25 (69)	128 (56.6)
Alcohol etiology, n (%)	61 (53.5)	71 (48.0)	39 (48)	93 (51.7)	16 (44)	116 (51.3)
Prior HE[d], n (%)	66 (57.9)	85 (57.4)	43 (52)	108 (60.0)	25 (69)	125 (55.3)
Lactulose use, n (%)	65 (57.0)	87 (58.8)	45 (55)	107 (59.4)	24 (67)	128 (56.6)
Rifaximin use, n (%)	38 (33.3)	63 (42.6)	34 (41)	67 (37.2)	15 (42)	85 (37.6)
Taking psychoactive medications, n (%)	25 (21.9)	32 (21.6)	16 (20)	41 (22.8)	9 (25)	48 (21.2)
Taking opioids, n (%)	28 (24.6)	43 (29.1)	25 (30)	46 (25.6)	9 (25)	62 (27.4)
MELD[e] score on admission, mean (SD)[f]	17.9 (7.8)	18.8 (8.9)	18.4 (9.3)	18.4 (8.0)	21.5 (8.4)	17.9 (8.3)
Reason for admission, n (%)						
Infection	37 (32.5)	44 (29.7)	27 (33)	54 (30.0)	14 (39)	67 (29.6)
HE[g]	34 (29.8)	48 (32.4)	24 (29)	58 (32.2)	17 (47)	64 (28.3)
Renal or metabolic disease	40 (35.1)	57 (38.5)	27 (33)	70 (38.9)	18 (50)	79 (35.0)
Gastrointestinal bleeding[g]	22 (19.3)	28 (18.9)	18 (22)	32 (17.8)	3 (8)	46 (20.4)
Other[h]	63 (55.3)	81 (54.7)	45 (55)	99 (55.0)	13 (36)	131 (58.0)

[a]Significant difference for technology demands (*P*=.02).

[b]Significant difference for technology demands (*P*=.046).

[c]Significant difference for caregiver reluctance (*P*<.001) and perceived study burden (*P*=.006).

[d]MELD: Model for End-stage Liver Disease.

[e]HE: hepatic encephalopathy.

[f]Significant difference for technology demands (*P*=.02).

[g]Significant difference for technology demands (*P*=.001).

[h]Significant difference for technology demands (*P*=.04).

Table 3. Further reasons for refusal to participate in relation to patient characteristics (N=262).

Characteristic	Transport/logistics		No specific reason	
	Yes (n=26)	No (n=235)	Yes (n=11)	No (n=251)
Age (years), mean (SD)	62.4 (10.3)	59.0 (11.3)	59.9 (10.8)	59.2 (11.4)
Race, n (%)				
Caucasian	19 (73)	188 (80.0)	10 (91)	197 (78.5)
African American	6 (23)	41 (17.4)	1 (9)	46 (18.3)
Other	1 (4)	7 (3.0)	0 (0)	8 (3.2)
Men, n (%)	16 (62)	134 (57.0)	6 (55)	144 (57.4)
High school education or above, n (%)	11 (42)	105 (44.7)	6 (55)	110 (43.8)
Potential caregiver not spouse, n (%)	15 (58)	137 (58.3)	5 (45)	148 (59.0)
Alcohol etiology, n (%)	10 (38)	122 (51.9)	7 (64)	125 (49.8)
Prior HE[a], n (%)	14 (54)	136 (57.9)	6 (55)	145 (57.8)
Lactulose use, n (%)	15 (58)	137 (58.3)	6 (55)	146 (58.2)
Rifaximin use, n (%)	7 (27)	93 (39.6)	5 (45)	96 (38.2)
Taking psychoactive medications, n (%)	4 (15)	53 (22.6)	4 (36)	53 (21.1)
Taking opioids, n (%)	5 (19)	66 (28.1)	4 (36)	67 (26.7)
MELD[b] score on admission, mean (SD)	16.3 (6.3)	18.6 (8.6)	17.9 (8.7)	18.4 (8.4)
Reason for admission, n (%)				
Infection	7 (27)	74 (31.5)	2 (18)	79 (31.5)
HE	11 (42)	70 (29.8)	2 (18)	80 (31.9)
Renal or metabolic disease	9 (35)	88 (37.4)	3 (27)	94 (37.5)
Gastrointestinal bleeding	5 (19)	44 (18.7)	3 (27)	47 (18.7)
Other[c]	7 (27)	137 (58.3)	4 (36)	140 (55.8)

[a]MELD: Model for End-stage Liver Disease.

[b]HE: hepatic encephalopathy.

[c]Significant difference for transport/logistics (*P*=.002).

As shown in Table 4, there were some differences detected between the patients across the three sites with respect to demographic characteristics, whereas MELD score and prior hepatic encephalopathy rates were similar. Veterans were older, likely to be men, and more likely to be on rifaximin, whereas the VCU cohort was more likely to be on psychoactive medications and to be admitted for infection or renal/metabolic reasons. Mayo Clinic patients were more likely to be Caucasian and to be admitted for gastrointestinal bleeding. Although there were differences in the reason for admission of patients at each site, differences between sites were ultimately not significant in the multivariable logistic regression analysis.

Table 4. Differences in patient characteristics among the three sites.

Characteristics	VCU[a] (n=191)	VA[b] (n=56)	Mayo Clinic (n=102)	*P* value
Age (years), mean (SD)	57.6 (10.6)	62.8 (10.8)	59.9 (12.6)	.006
Race, n (%)				<.001
Caucasian	140 (73.3)	34 (61)	97 (95.1)	
African American	44 (23.0)	17 (30)	1 (0.9)	
Other	8 (4.2)	5 (9)	4 (3.9)	
Men, n (%)	103 (53.9)	52 (93)	58 (56.9)	<.001
Comorbidity, n (%)				.03
HCV[c]	22 (11.5)	7 (13)	4 (3.9)	
HCV+alcohol-related	74 (38.7)	19 (34)	50 (49.0)	
Alcohol-related	16 (8.4)	6 (11)	2 (2.0)	
NAFLD[d]	40 (20.9)	18 (32)	27 (26.5)	
Other	39 (20.4)	6 (11)	19 (18.6)	
Prior HE[e], n (%)	118 (61.8)	37 (66)	52 (51.0)	.11
Lactulose use, n (%)	118 (61.8)	33 (59)	51 (50.0)	.15
Rifaximin use, n (%)	72 (37.7)	32 (57)	37 (36.3)	.02
Taking psychoactive medications, n (%)	51 (26.7)	8 (14)	10 (9.8)	.001
Taking opioids, n (%)	50 (26.2)	9 (16)	23 (22.5)	.26
MELD[f] score on admission, mean (SD)	19.3 (8.1)	17.4 (7.1)	18.2 (8.5)	.26
Reason for admission, n (%)				
Infection	80 (41.9)	24 (43)	6 (5.9)	<.001
HE	71 (37.2)	21 (38)	28 (27.5)	.21
Renal or metabolic disease	98 (51.3)	10 (18)	18 (17.6)	<.001
Gastrointestinal bleeding	40 (20.9)	7 (13)	31 (30.4)	.02
Other	87 (45.5)	23 (41)	55 (53.9)	.23

[a]VCU: Virginia Commonwealth University

[b]VA: Rochester Veteran Affairs Medical Center.

[c]HCV: hepatitis C virus.

[d]NAFLD: nonalcholic fatty liver disease.

[e]HE: hepatic encephalopathy.

[f]MELD: Model for End-stage Liver Disease.

Discussion

Our results demonstrate that agreeing to participate in an intensive health IT regimen aimed at preventing 30-day readmissions for inpatients with cirrhosis and their caregivers may depend in part on the cirrhosis etiology and reasons for inpatient admission. However, the major reasons cited by dyads for not participating appeared to focus on caregivers, study burden, and technological demands rather than on privacy issues. The goal of this multisite trial is to reduce readmissions in cirrhosis, an intractable issue that has medical, psychosocial, and financial consequences [2,8,9]. Our preliminary study demonstrated reductions in hepatic encephalopathy–related readmissions at 30 days when using an intensive health IT intervention that focused on daily communications among patients, family caregivers, and the clinical team [5]. However, given that this previously tested intervention involved several steps in the daily recording of data that could potentially become onerous for the respondents, we included this semistructured interview study within the multisite trial to examine dyads' perspectives on the use of health IT interventions to better streamline our approach.

Only 25% of the patient-caregiver dyads who were approached in the hospital agreed to be randomly assigned to standard care or to one of the two health IT interventions. This rate is relatively lower than that of individuals who ultimately decided to download another app, EncephalApp, but was similar to the rate of individuals that ultimately used the technology in a recent study of outpatients with cirrhosis [10]. Unlike EncephalApp, which links to the Patient Buddy app, the regimen tested in our

multisite trial involves daily entry of data related to medication adherence, weight, and a multipronged approach to orientation that unfolds over 30 days and includes an educational component. This time frame requires a much longer commitment from patient-caregiver dyads along with training from health care providers at the time of discharge. Nevertheless, the intervention includes valuable educational material and offers the potential to continue communication with the treating teams after discharge for most of the clinical issues that underlie readmission, such as hepatic encephalopathy, gastrointestinal bleeding, ascites, and medication management [3,9]. This approach was tested by Bloom et al [11] in a cross-sectional survey where theoretical acceptance of an app that would require similar communications as currently being tested with Patient Buddy was considered to be acceptable to most patients. This finding is in contrast to the results of this study, likely because unlike the theoretical constructs in the prior study, the dyads had to agree to be randomized for a 30-day trial in real life in our study.

The fact that patients admitted with hepatic encephalopathy and gastrointestinal bleeding were more likely to agree to participate in randomization may be related to the patient-caregiver dyad wanting to avoid such occurrences in the future. By contrast, patients with alcohol-related liver disease etiology and opioid use, which often coexist with lower socioeconomic and education status (and potentially unfamiliarity with apps), were less likely to participate [12]. The burden of cirrhosis, and especially hepatic encephalopathy, is shared by patients and their caregivers [13]. Since the most important reasons behind patient-caregiver dyads not agreeing to proceed further were related to caregiver reluctance, perceived burden of the study, and issues involving transport and logistics, the design and components of the health IT intervention may need reconsideration.

The findings of this study are instructive. Although the overall goal of the study team and the multisite trial is to reduce burden for both patients and their families by preventing readmission, the amount of time and effort required to participate in this study may have mitigated the dyads' desire or willingness to engage in the activities needed to keep up with data entry in the app [14]. The Patient Buddy app being tested in the multisite trial was streamlined from the version tested in our preliminary work based on patient and caregiver responses (specifically items designed to reduce gait issues, fall risk, and sodium content of the diet, and the complexity of the questions asked). However, given the multiple demands on the patient-caregiver dyads' time during hospitalization and just prior to discharge, and the complexities of the current Patient Buddy app (even with its improvements), it appears that a better balance between gaining the information required and minimizing the burden on patient-caregiver dyads is warranted.

Apart from the barriers mentioned above, reasons for not wanting to be randomized to using a health IT intervention could also be centered around privacy and technological demands. However, of interest, privacy issues were not cited as a reason for the decision not to undergo randomization, which might signify trust in the HIPAA compliance of these apps or the presence of other more pervasive issues.

With respect to refusals on technological grounds, we found that patients admitted with hepatic encephalopathy and those who were older, less educated, and with higher MELD scores were more likely to refuse participation. Similar results have been shown in prior studies in patients with and without cirrhosis, and may represent a lack of individual or family capability to perform the tasks [15]. Given that we anticipated enrolling patient-caregiver dyads from a lower socioeconomic status in the multisite trial, we specifically designed a study protocol that provided smartphones individually to patients and their caregivers. However, it is still possible that unfamiliarity with a smartphone, including the use of apps, remains a barrier to engagement in health IT interventions in patients with liver disease. Although the potential issues with use of one-time apps, particularly in the outpatient setting, may be overcome with teaching, it is possible that sustained use over 30 days for the Patient Buddy app could have been construed as too demanding for older patients with hepatic encephalopathy and more advanced cirrhosis.

Finally, an important area for discussion is that perspectives of the study team and patient-caregiver dyads may have influenced our results. Although some studies focus on specific areas of cirrhosis case management postdischarge, the Patient Buddy app incorporates a more customized approach to cirrhosis and is also more broadly focused. For instance, factors such as hepatic encephalopathy, ascites, renal and metabolic issues, and gastrointestinal bleeding could all potentially influence readmission risk. Therefore, the comprehensive nature of this app could have resulted in the perception that it was more difficult to use compared to other apps focused on a single complication. Nonetheless, a broadly focused and well-rounded app addressing the impact of all complications of cirrhosis is needed to better educate patients and their family members about the disease and ultimately reduce readmission rates.

Our work with the Patient Buddy app and its impact on readmissions is continuing in the multisite trial, but this interim exploration of reasons for why dyads declined randomization in the trial has provided information to further refine our understanding on how to better address barriers to using health IT interventions. This will result in streamlining our approach toward the dyads and focusing on the time spent and potential benefits as we approach them. The rate of participation was lower compared with that reported in some other app-based trials in gastroenterology that focused on patients undergoing colonoscopy and those with inflammatory bowel disease [16,17]. These differences could be due to the inpatient setting, which selects for advanced disease and cognitive demands of the cirrhosis disease process, likely being greater than those involved for other diseases studied. Moreover, we interviewed patients and caregivers together, which makes it difficult to evaluate individual responses, and we did not specifically evaluate socioeconomic factors. We asked questions regarding technological familiarity rather than providing dedicated questionnaires since the dyads were not interested in performing further study-related work apart from a brief interview.

In conclusion, the reasons for patients with cirrhosis and their caregivers declining to participate in a trial using an intensive app that requires 30 days of feedback regardless of

demographics and clinical settings were mainly related to caregiver and study demands, but importantly not to privacy concerns. Those admitted with hepatic encephalopathy and gastrointestinal bleeding issues were more likely to agree to participate, whereas those with alcohol etiology and opioid use were less likely to participate in the health IT study. Further research should address the careful balance of patient and caregiver burdens with clinicians' needs for accurate and timely information that enables good disease management.

Acknowledgments

This work is supported by an Agency for Healthcare Research and Quality grant (RO1HS025412) to JB and PK.

Conflicts of Interest

None of the authors have a conflict of interest.

References

1. Kanwal F, Tansel A, Kramer JR, Feng H, Asch SM, El-Serag HB. Trends in 30-day and 1-year mortality among patients hospitalized with cirrhosis from 2004 to 2013. Am J Gastroenterol 2017 Aug;112(8):1287-1297. [doi: 10.1038/ajg.2017.175] [Medline: 28607480]
2. Tapper EB, Volk M. Strategies to reduce 30-day readmissions in patients with cirrhosis. Curr Gastroenterol Rep 2017 Jan;19(1):1. [doi: 10.1007/s11894-017-0543-3] [Medline: 28101791]
3. Volk ML, Tocco RS, Bazick J, Rakoski MO, Lok AS. Hospital readmissions among patients with decompensated cirrhosis. Am J Gastroenterol 2012 Mar;107(2):247-252 [FREE Full text] [doi: 10.1038/ajg.2011.314] [Medline: 21931378]
4. Tapper EB, Finkelstein D, Mittleman MA, Piatkowski G, Chang M, Lai M. A quality improvement initiative reduces 30-day rate of readmission for patients with cirrhosis. Clin Gastroenterol Hepatol 2016 May;14(5):753-759 [FREE Full text] [doi: 10.1016/j.cgh.2015.08.041] [Medline: 26407750]
5. Ganapathy D, Acharya C, Lachar J, Patidar K, Sterling RK, White MB, et al. The Patient Buddy app can potentially prevent hepatic encephalopathy-related readmissions. Liver Int 2017 Dec;37(12):1843-1851. [doi: 10.1111/liv.13494] [Medline: 28618192]
6. Allampati S, Duarte-Rojo A, Thacker LR, Patidar KR, White MB, Klair JS, et al. Diagnosis of minimal hepatic encephalopathy using Stroop EncephalApp: a multicenter US-based, norm-based study. Am J Gastroenterol 2016 Jan;111(1):78-86. [doi: 10.1038/ajg.2015.377] [Medline: 26644276]
7. Bajaj JS, Heuman DM, Sterling RK, Sanyal AJ, Siddiqui M, Matherly S, et al. Validation of EncephalApp, smartphone-based Stroop test, for the diagnosis of covert hepatic encephalopathy. Clin Gastroenterol Hepatol 2015 Oct;13(10):1828-1835 [FREE Full text] [doi: 10.1016/j.cgh.2014.05.011] [Medline: 24846278]
8. Tapper EB, Halbert B, Mellinger J. Rates of and reasons for hospital readmissions in patients with cirrhosis: a multistate population-based cohort study. Clin Gastroenterol Hepatol 2016 Aug;14(8):1181-1188. [doi: 10.1016/j.cgh.2016.04.009] [Medline: 27085758]
9. Bajaj JS, Reddy KR, Tandon P, Wong F, Kamath PS, Garcia-Tsao G, North American Consortium for the Study of End-Stage Liver Disease. The 3-month readmission rate remains unacceptably high in a large North American cohort of patients with cirrhosis. Hepatology 2016 Jul;64(1):200-208 [FREE Full text] [doi: 10.1002/hep.28414] [Medline: 26690389]
10. Louissaint J, Lok AS, Fortune BE, Tapper EB. Acceptance and use of a smartphone application in cirrhosis. Liver Int 2020 Jul;40(7):1556-1563. [doi: 10.1111/liv.14494] [Medline: 32343467]
11. Bloom PP, Marx M, Wang TJ, Green B, Ha J, Bay C, et al. Attitudes towards digital health tools for outpatient cirrhosis management in patients with decompensated cirrhosis. BMJ Innov 2020 Jan 14;6(1):18-25. [doi: 10.1136/bmjinnov-2019-000369]
12. Probst C, Rehm J. Alcohol use, opioid overdose and socioeconomic status in Canada: A threat to life expectancy? Can Med Assoc J 2018 Nov 05;190(44):E1294-E1295 [FREE Full text] [doi: 10.1503/cmaj.180806] [Medline: 30397155]
13. Bajaj JS, Wade JB, Gibson DP, Heuman DM, Thacker LR, Sterling RK, et al. The multi-dimensional burden of cirrhosis and hepatic encephalopathy on patients and caregivers. Am J Gastroenterol 2011 Oct;106(9):1646-1653 [FREE Full text] [doi: 10.1038/ajg.2011.157] [Medline: 21556040]
14. Fabrellas N, Moreira R, Carol M, Cervera M, de Prada G, Perez M, et al. Psychological burden of hepatic encephalopathy on patients and caregivers. Clin Transl Gastroenterol 2020 Apr;11(4):e00159 [FREE Full text] [doi: 10.14309/ctg.0000000000000159] [Medline: 32352686]
15. Nunes A, Limpo T, Castro SL. Acceptance of mobile health applications: examining key determinants and moderators. Front Psychol 2019;10:2791. [doi: 10.3389/fpsyg.2019.02791] [Medline: 31920836]
16. Desai M, Nutalapati V, Bansal A, Buckles D, Bonino J, Olyaee M, et al. Use of smartphone applications to improve quality of bowel preparation for colonoscopy: a systematic review and meta-analysis. Endosc Int Open 2019 Mar 18;7(2):E216-E224 [FREE Full text] [doi: 10.1055/a-0796-6423] [Medline: 30705956]

17. Con D, De Cruz P. Mobile phone apps for inflammatory bowel disease self-management: a systematic assessment of content and tools. JMIR Mhealth Uhealth 2016 Mar 01;4(1):e13 [FREE Full text] [doi: 10.2196/mhealth.4874] [Medline: 26831935]

Abbreviations

HIPAA: Health Insurance Portability and Accountability Act
IT: information technology
MELD: Model for End-stage Liver Disease
OR: odds ratio
VA: Veterans Affairs
VCU: Virginia Commonwealth University

Edited by R Kukafka; submitted 28.09.20; peer-reviewed by P Bloom, J Louissaint; comments to author 09.10.20; revised version received 13.10.20; accepted 21.03.21; published 09.04.21.

Please cite as:
Acharya C, Sehrawat TS, McGuire DB, Shaw J, Fagan A, McGeorge S, Olofson A, White MB, Gavis E, Kamath PS, Bergstrom L, Bajaj JS
Perspectives of Inpatients With Cirrhosis and Caregivers on Using Health Information Technology: Cross-sectional Multicenter Study
J Med Internet Res 2021;23(4):e24639
URL: https://www.jmir.org/2021/4/e24639
doi:10.2196/24639
PMID:33744844

Original Paper

Social Network Analysis of the Effects of a Social Media–Based Weight Loss Intervention Targeting Adults of Low Socioeconomic Status: Single-Arm Intervention Trial

Ran Xu[1], PhD; David Cavallo[2], MPH, PhD

[1]Department of Allied Health Sciences, College of Agriculture, Health and Natural Resources, University of Connecticut, Storrs, CT, United States
[2]Department of Nutrition, School of Medicine, Case Western Reserve University, Cleveland, OH, United States

Corresponding Author:
Ran Xu, PhD
Department of Allied Health Sciences
College of Agriculture, Health and Natural Resources
University of Connecticut
Koons Hall 326
Storrs, CT,
United States
Phone: 1 860 486 2945
Email: Ran.2.xu@uconn.edu

Abstract

Background: Obesity is a known risk factor for cardiovascular disease risk factors, including hypertension and type II diabetes. Although numerous weight loss interventions have demonstrated efficacy, there is considerably less evidence about the theoretical mechanisms through which they work. Delivering lifestyle behavior change interventions via social media provides unique opportunities for understanding mechanisms of intervention effects. Server data collected directly from web-based platforms can provide detailed, real-time behavioral information over the course of intervention programs that can be used to understand how interventions work.

Objective: The objective of this study was to demonstrate how social network analysis can facilitate our understanding of the mechanisms underlying a social media–based weight loss intervention.

Methods: We performed secondary analysis by using data from a pilot study that delivered a dietary and physical activity intervention to a group of participants via Facebook. We mapped out participants' interaction networks over the 12-week intervention period and linked participants' network characteristics (eg, in-degree, out-degree, network constraint) to participants' changes in theoretical mediators (ie, dietary knowledge, perceived social support, self-efficacy) and weight loss by using regression analysis. We also performed mediation analyses to explore how the effects of social network measures on weight loss could be mediated by the aforementioned theoretical mediators.

Results: In this analysis, 47 participants from 2 waves completed the study and were included. We found that increases in the number of posts, comments, and reactions significantly predicted weight loss (β=–.94, *P*=.04); receiving comments positively predicted changes in self-efficacy (β=7.81, *P*=.009), and the degree to which one's network neighbors are tightly connected with each other weakly predicted changes in perceived social support (β=7.70, *P*=.08). In addition, change in self-efficacy mediated the relationship between receiving comments and weight loss (β=–.89, *P*=.02).

Conclusions: Our analyses using data from this pilot study linked participants' network characteristics with changes in several important study outcomes of interest such as self-efficacy, social support, and weight. Our results point to the potential of using social network analysis to understand the social processes and mechanisms through which web-based behavioral interventions affect participants' psychological and behavioral outcomes. Future studies are warranted to validate our results and to further explore the relationship between network dynamics and study outcomes in similar and larger trials.

(J Med Internet Res 2021;23(4):e24690) doi:10.2196/24690

KEYWORDS

weight loss intervention; social media intervention; electronic health; social network analysis

Introduction

Obesity is a common disease with a high economic burden, and it is a known risk factor for several chronic diseases, including cancer, cardiovascular diseases, and diabetes [1]. Individuals with low socioeconomic status (SES) are disproportionately more likely to be obese and develop related diseases [2,3]. While there are a vast number of behavioral weight loss interventions established in previous literature [4], relatively few studies of high quality have targeted participants of low SES. There is some evidence that groups with low SES experience lower efficacy in weight loss interventions, especially those that focus only on individual-level determinants [5-7]. This highlights the need for the development of effective behavioral weight loss strategies for individuals with low SES.

A potentially effective way of delivering behavioral weight loss interventions to participants of low SES is the use of social media platforms (eg, Facebook, Twitter). Social media is widely used by internet users of low SES [8]. Compared with face-to-face interventions, web-based delivery of social media offers advantages such as continuous availability and remote access, which could reduce barriers to intervention use among participants of low SES, such as limited time or inadequate transportation [9,10]. In addition, social media users are accustomed to sharing information about their health experiences and opinions [11,12]. This sharing may be enhanced by the communication features of social media websites, which can facilitate sharing and communication among users about weight loss behaviors, goals, and experiences [13,14]. These features make social media platforms particularly attractive for delivering behavioral weight loss interventions. Indeed, some preliminary evidence suggests that social media–based interventions can be efficacious in producing weight loss and increasing the frequency of social interaction and social support [15]. Results from a handful of social media–based behavioral weight loss interventions targeting participants of low SES show that participants have similar receptiveness to the intervention format and weight loss outcome when comparing in-person and social media intervention delivery [16,17]. However, as the evidence of the efficacy of social media–based weight loss intervention accumulates, our understanding of the mechanisms underlying social media–based weight loss interventions is still somewhat limited. Previous literature suggests that constructs such as social support and social comparison are important weight loss predictors. Some studies have linked social media–based intervention components with increased social support, positive behavior change, and health outcomes [18,19]. Studies have also reported greater weight loss for participants in teams or those reporting greater social influence [20]. This evidence, however, is scattered in the context of social media–based weight loss interventions and is largely based on cross-sectional or post-hoc studies. The mechanisms through which social media–based weight loss interventions change participants' interactions, theoretical mediators (eg, knowledge, social support, self-efficacy), and subsequent weight loss behavior and outcomes are still largely unclear.

A promising approach to increase our understanding of the mechanisms underlying social media–based weight loss interventions is to study the user-generated content and interactions during the intervention by using social network analysis. Previous research has shown that interactions generated through social networks serve as important sources for information and knowledge [21-23], social and physical resources [24,25], and social support [26,27]. In addition, social networks influence individual behavior and perception through various mechanisms such as sensemaking, norms, and learning [28,29]. Thus, it is possible to leverage social network analysis to better understand users' position and role in the interaction networks and how that affects the theoretical mediators and subsequent weight loss behavior and outcomes.

In this study, using data from a 12-week pilot behavioral intervention that assessed the feasibility of a weight loss program delivered via social media to adults of low SES [30], we mapped out participants' interaction networks and conducted social network analysis to assess how participants' network characteristics were associated with various psychological and behavioral study outcomes, as well as the possible mediating mechanisms that lead to participants' weight loss.

Methods

Design

We conducted secondary analysis of data from the INSHAPE CLE study. INSHAPE CLE was a single-arm, pre-post intervention trial designed to assess the feasibility of delivering a social media–based weight loss intervention to low-income residents of a large midwestern urban area of the United States of America. INSHAPE CLE delivered a 12-week behavioral weight loss intervention that included goal setting and self-monitoring using Fitbit devices and the Fitbit self-monitoring platform. Participants were also enrolled in a private Facebook group moderated by study personnel (Moderator), who posted planned communications to the group that delivered nutrition and physical activity education and encouraged participants to exchange social support and model desired behaviors. Moderators also responded to participant questions and monitored the group for inappropriate communications. All procedures performed in this study involving human participants were in accordance with the ethical standards of the institutional and national research committee and were approved by the Case Western Reserve University Institutional Review Board and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards [31-34]. This paper does not contain any studies with animals performed by any of the authors.

Participants

INSHAPE CLE participants were recruited using social media advertisements, flyers posted in strategic locations, and in-person recruitment events at federally sponsored nutrition education events. Primary inclusion criteria were age of 35-65 years, regular internet use, BMI of 25-40, less than 30 minutes per day of moderate or vigorous physical activity, residence in the target metropolitan area, income less than 185% of the federal poverty based on family size, and no affirmative answers on the Physical Activity Readiness Questionnaire [35] or written clearance by a physician. Participants were recruited in 2 waves

and enrolled in 2 separate Facebook groups with largely identical moderator content.

Study Procedures

Participants were screened online. Eligible participants were provided a web-based consent form and those who accepted were provided a web-based baseline questionnaire and invited to an in-person orientation meeting where study instructions were provided and anthropometric measures were collected. Participants attending an orientation meeting were considered enrolled in the study. Enrollees were provided a web-based follow-up questionnaire at the end of the 12-week intervention and invited to a meeting where follow-up anthropometric measures were taken. Additional details of the INSHAPE CLE intervention and study procedures can be found elsewhere [30].

Measures

Demographics

As a part of the baseline questionnaire, participants were asked to report their birth year, gender, race, ethnicity, educational status, and employment status.

Anthropometrics

All anthropometrics were collected by study staff who attended a 1-hour training session. We measured height by using a stadiometer (Detecto PHR Portable Mechanical Height Rod for DR400C) and weight by using a floor scale (Detecto DR550C) calibrated using a 10-kg weight (Ohaus 80850302). Measurement procedures were based on the National Health and Nutrition Examination Survey Anthropometry Procedures Manual.

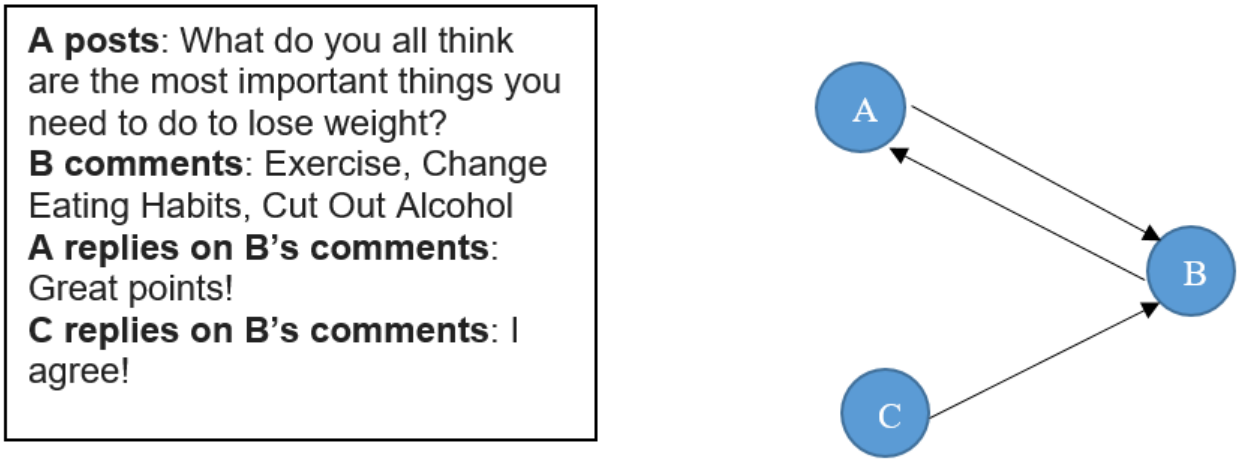
Theoretical Mediators

Several psychological and behavioral constructs were theorized to be important determinants of the weight loss outcome. Specifically, dietary knowledge was assessed by the sum of correct answers from a 36-item questionnaire adapted from a validated nutrition knowledge questionnaire and was scored as the percentage of questions answered correctly. The questionnaire assessed several domains of nutrition knowledge, including national dietary recommendations, the nutrient content of foods, and the relationship between dietary behavior and chronic disease. Perceived social support was measured using the 5-item Friend Social Support and Eating Habits Scale that asks participants to indicate how often their friends (1=never, 5=very often) communicated positive messages about dietary behavior (eg, discussed my eating habit changes with me [asked me how I'm doing with my eating changes]). Self-efficacy was assessed using a validated 4-item scale measure. This scale asked participants "How confident are you that you can lose weight?" with a series of conditional statements (eg, Even if you have to try several times until it works) rated on a sliding scale from 0% (not at all confident) to 100% (completely confident).

Social Network/Engagement Measures

Each post, comment, and reaction (eg, emojis) on the Facebook group during the intervention was recorded. Two data collection methods were used to obtain Facebook data. The Grytics analytical platform [36] provided data directly from the Facebook application programming interface. We also collected select data manually by reviewing and logging information directly from the Facebook group. This was required due to changes to Facebook's privacy policy that occurred during the second group's intervention that deidentified individual Facebook user data. However, the Grytics platform did not provide data on the reactions for comments with the exception of likes. Data were collected in both ways and validated by comparing aggregate data provided by Grytics, where available, to the totals from the manual data collection. In cases where totals did not agree, Facebook was manually checked a second time to ensure that the data collection methods produce consistent data across the entire intervention period. To obtain a more nuanced understanding of subjects' different levels/types of participation and engagement in the intervention, a social network was constructed using the comment relationship. A network tie from i to j represents i has sent out a comment/reply to j's post or comment, with the weight representing the number of comments across the whole intervention. To further illustrate this, in the simple example shown in Figure 1, the communication stream on the left can be represented as the network on the right.

Figure 1. A simple example illustrating the network construction process, wherein the link represents the comment/reply relationship.



We derived several centrality measures from the social network to indicate the type or level of treatment an individual received during the intervention. Specifically, in-degree was calculated as the total number of comments each individual received during the intervention. Network constraint was calculated to represent the degree to which one's interaction partners also interact with each other. Previous studies have linked network constraint to individual's social capital, social support, and access to diverse information [24]. The interaction partners/network neighbors of an individual with higher network constraint are more tightly connected with each other, while the interaction partners/network neighbors of an individual with lower network constraint are more disconnected from each other. Finally, we calculated out-degree to represent participants' engagement. As the network constructed excluded posts and reactions that one created, instead of out-degree based on the constructed network, we calculated out-degree as the sum of posts, reactions, and comments for each individual. As out-degree and in-degree are highly skewed, we used the log-transformed variables in the final analysis.

Data Analyses

Data analyses were completed using STATA 16.0 (StataCorp). Data were screened for deviations from assumptions required for the statistical analyses used. Analyses were performed on participants who completed baseline and follow-up measures on variables used in the analysis (ie, study completers). There were no significant differences in the demographic characteristics between completers (N=47) and those lost to follow-up (n=8). We calculated descriptive statistics for baseline characteristics, outcome, and theoretical mediator measures, as well as social network measures. To examine how social network measures were associated with the change in outcome and theoretical mediators, we used multivariate regression analysis with the change in weight, dietary knowledge, social support, and self-efficacy as the outcome (ie, difference between the outcome after the intervention and at baseline) and each aforementioned social network measure as the predictor in separate regression models. In each regression analysis, we controlled for the baseline value of the outcome, the treatment group indicator (group 1 or group 2) as well as other baseline characteristics such as age, race, education status, and BMI. We reported robust standard error to account for possible heteroscedasticity. In addition, to rule out the possible bias of network construction in our results, we also performed 2 sensitivity analyses: (1) we calculated out-degree as the total number of comments each individual made during the intervention and reran the aforementioned analyses and (2) we constructed the comment networks and associated network measures by excluding moderator-related interactions (ie, participant-only network) and reran the aforementioned analysis. Detailed results are reported in Multimedia Appendix 1. Furthermore, while the original intervention was not powered for the mediation analysis, to explore the possible mediating roles of changes in dietary knowledge, social support, and self-efficacy on the effect of social network measures on weight change, we conducted separate mediation analyses by using structural equation modeling, with changes in each of the theoretical mediators as the mediator and each social network measure as the predictor. In each structural equation modeling model, we controlled for the same covariates as previously mentioned.

Results

Participants' Demographics and Intervention Outcomes

Table 1 presents the baseline characteristics of the participants (N=47) included in the analysis (n=34 for wave I [group 1], n=13 for wave II [group 2]). The mean age of the participants was 46 years with mean baseline BMI of 34.09 kg/m^2. Participants were predominantly female (44/47, 94%) and mostly African American (32/47, 68%). Almost half (22/47, 47%) reported completing college or graduate school. The baseline characteristics of the participants in the 2 waves/groups were similar and we did not observe noticeable differences in these variables between the waves/groups.

Table 2 reports the change in the main study outcomes during the intervention. Compared with the baseline values, participants experienced an average weight loss of 1.25 kg (P=.049), significant increases in dietary knowledge of 2.28 (P=.02), and positive dietary social support of 2.87 (P<.001), but a nonsignificant decrease in average self-efficacy of 3.07 (P=.28).

Table 1. Demographics of INSHAPE CLE participants who completed the study.

Characteristics	Group 1 (n=34)	Group 2 (n=13)	Total (N=47)
Age (years), mean (SD)	45.59 (8.91)	48.38 (10.52)	46.36 (9.35)
Gender, n (%)			
Female	31 (91)	13 (100)	44 (94)
Male	2 (6)	0 (0)	2 (4)
Transgender	1 (3)	0 (0)	1 (2)
Race, n (%)			
White	9 (27)	3 (23)	12 (26)
Black or African American	23 (68)	9 (69)	32 (68)
More than one race	2 (6)	1 (8)	3 (6)
Education, n (%)			
College graduate or more	19 (56)	3 (23)	22 (47)
Some college	13 (38)	7 (54)	20 (43)
High school graduate	2 (6)	3 (23)	5 (11)
BMI, mean (SD)	33.79 (3.88)	34.87 (3.86)	34.09 (3.86)

Table 2. Study outcomes of INSHAPE CLE participants who completed the study.

Outcome	Baseline values, mean (SD)	Follow-up values, mean (SD)	Mean change (95% CI)	*P* value
Weight (kg)	94.63 (12.85)	93.38 (13.76)	–1.25[a] (–2.51 to 0.00)	.049
Dietary knowledge (scale 0-100)	27.64 (7.57)	30.08 (8.19)	2.28[a] (0.44 to 4.13)	.02
Positive dietary social support (scale 5-25)	10.48 (5.27)	13.26 (4.83)	2.87[b] (1.25 to 4.49)	<.001
Weight loss self-efficacy (scale 0-100)	86.24 (15.30)	83.03 (20.48)	–3.07 (–8.67 to 2.54)	.28

[a]Significant at $P<.05$.

[b]Significant at $P<.001$.

Social Network Measures and Association With Changes in the Study Outcome

Social network was constructed based on participants' comment relationships during the intervention. Figure 2A and Figure 2B present the comment network among participants in group 1 and group 2, respectively. A link from node 1 to node 2 represents that node 1 has made comments to node 2, with thickness representing frequency. Node color represents the number of comments one made to others, with darker color indicating more comments. Node size represents the number of posts one created, with larger size indicating more posts. Nodes were laid out by a multilevel force-directed algorithm [37]. The network graph intuitively shows there are different levels of engagement among participants—some participants were positioned at the center of the network with many posts or comments, while many participants were positioned at the periphery of the network with very few comments or posts. The structure is more evident in group 1 (where the group size is larger), where k-core analysis [38] showed 14 people formed a core—each person in the core received comments from at least 10 other people in the core. This figure also shows there are different types of engagement among participants—some participants had few original posts but commented on others frequently, some participants did not make many comments to others but received many comments, and some participants were more embedded in the network with their network neighbors tightly connecting to each other. This observation is evident in Table 3, which shows there were large variations in both out-degree (mean 186.32, SD 178.24) and in-degree (mean 25.42, SD 30.83) among the participants. Table 3 also reports the results from regression analyses on how different types of engagement (indicated by social network measures) were associated with changes in study outcomes. Among them, increases in the number of posts, comments, and reactions made (out-degree) significantly predicted weight loss (β=–.94, $P=.04$)—1% increase in the number of posts, comments, and reactions made was associated with 0.0094 kg in weight loss. Increase in the number of comments received (in-degree) significantly predicted increase in self-efficacy (β=7.81, $P=.009$)—1% increase in the number of comments received was associated with 0.0781 units increase in positive self-efficacy change (self-efficacy was measured on a 0-100 scale). In addition, while not significant at .05 level, out-degree was also likely to be positively associated with changes in self-efficacy (β=3.44, $P=.08$), and the degree to which one's network neighbors are tightly connected with each other (network constraint) was likely to be positively associated with changes in positive dietary social support (β=7.70, $P=.08$).

Sensitivity analyses reported in Multimedia Appendix 1 (Table S1 and Table S2) showed that the network structures and the regression results related to in-degree and out-degree were consistent and robust with different ways of constructing networks.

Figure 2. Comment network during the intervention in group 1 (A) and group 2 (B). Link from node 1 to node 2 represents node 1 has made comments to node 2, with thickness representing frequency. Node color represents the number of comments one made to others, with darker color indicating more comments. Node size represents the number of posts one created, with larger size indicating more posts.

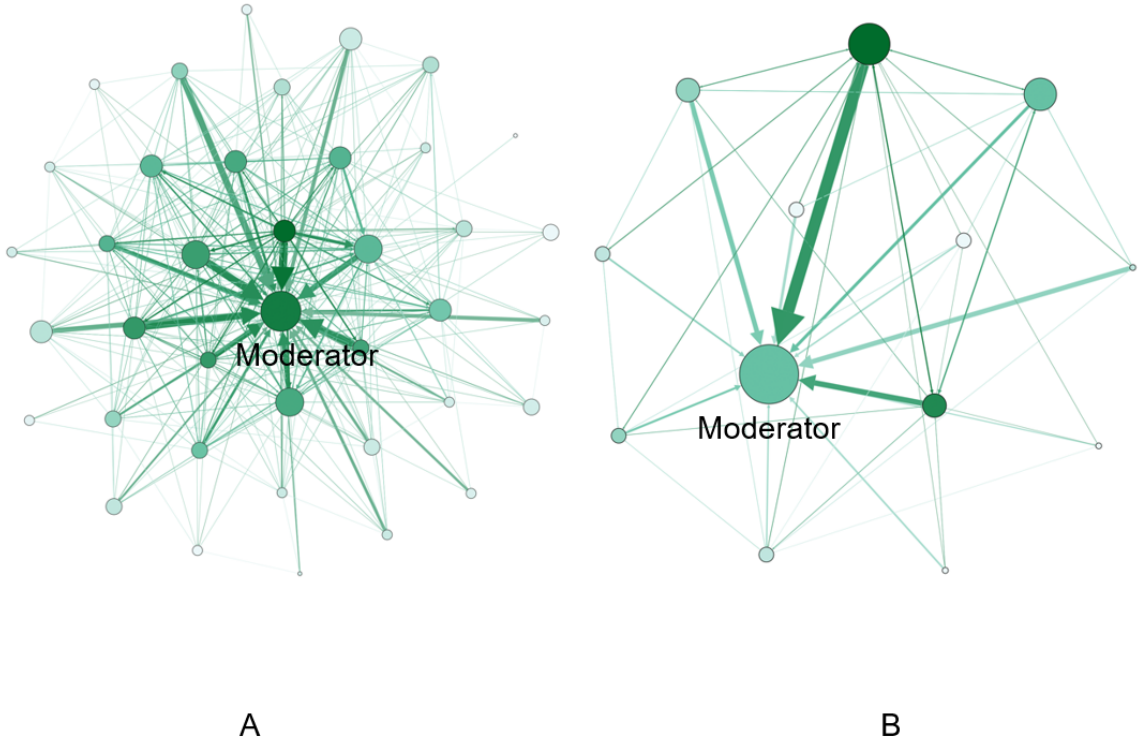


Table 3. Associations between social network measures and change in study outcomes for INSHAPE CLE participants.[a]

Predictors/outcome	Mean (SD)	Median (IQR)	Min-Max	Weight change		Dietary knowledge change		Social support change		Self-efficacy change	
				β (95% CI)	*P* value	β (95% CI)	*P* value	β (95% CI)	*P* value	β (95% CI)	*P* value
Out-degree	186.32 (178.24)	127 (238)	1-643	–.94[b] (–1.85 to –.04)	.04	.06 (–1.13 to 1.24)	.93	.69 (–.37 to 1.75)	.20	3.44[c] (–.38 to 7.26)	.08
In-degree	25.42 (30.83)	15 (29)	0-174	–.72 (–2.08 to .64)	.29	.94 (–.57 to 2.46)	.21	–.22 (–1.58 to 1.12)	.73	7.81[d] (2.06 to 13.57)	.009
Network constraint	0.62 (0.24)	0.52 (0.46)	0.23-1.05	–6.16 (–15.99 to 3.66)	.21	–7.78 (–18.36 to 2.79)	.14	7.70[c] (–.98 to 16.39)	.08	–21.96 (–63.12 to 19.20)	.29

[a]In the analysis, out-degree and in-degree were log-transformed. All models controlled for the outcome before the intervention, the treatment group indicator, age, race, education status, and BMI.

[b]Significant at *P*<.05.

[c]Significant at *P*<.10.

[d]Significant at *P*<.01.

Mediation Analysis

As an exploratory step, we also investigated the possible mediating effects of the theoretical mediators on the effects of social network measures on weight loss, which is presented in Table 4. Structural equation modeling results show that there was a significant indirect effect from in-degree to weight loss that went through a change in self-efficacy (β=–.89, *P*=.02). Figure 3 shows that receiving more comments was positively associated with changes in one's self-efficacy during the intervention (β=7.81, *P*<.001), which was subsequently associated with more weight loss (β=–.11, *P*=.001).

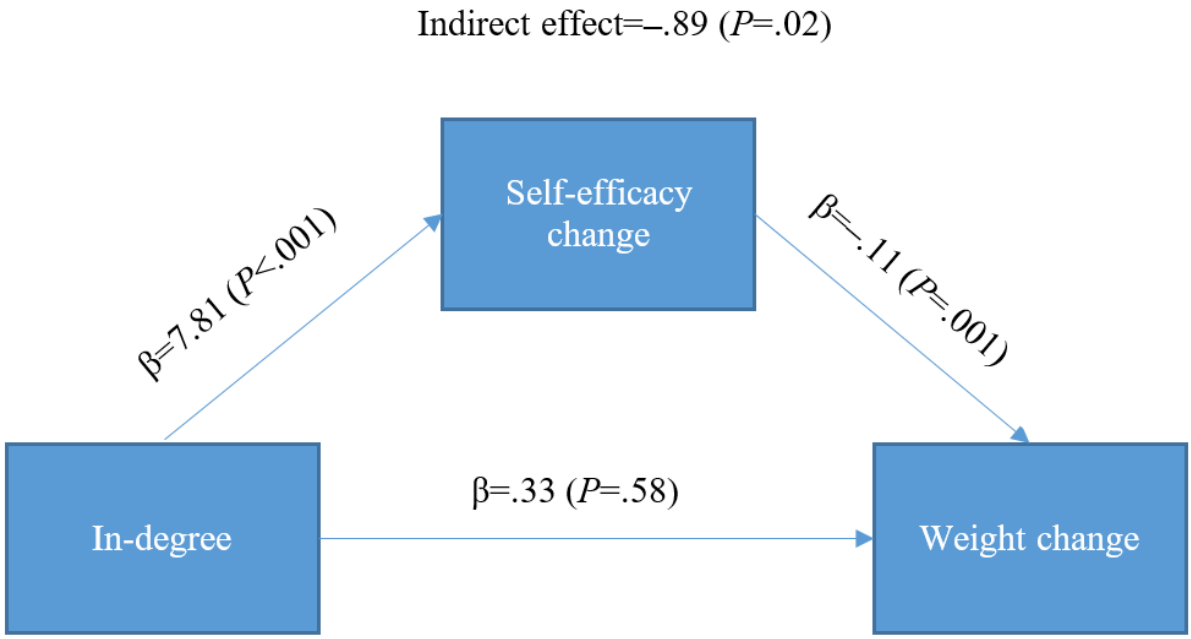
Table 4. Mediation analysis from social network measures to changes in theoretical mediators to weight change for INSHAPE CLE participants.[a]

Predictors/mediators	Dietary knowledge change		Social support change		Self-efficacy change	
	β (95% CI)	*P* value	β (95% CI)	*P* value	β (95% CI)	*P* value
Out-degree	–.01 (–.06 to .05)	.83	–.09 (–.32 to .14)	.43	–.29 (–.64 to .06)	.11
In-degree	–.03 (–.37 to .31)	.88	.04 (–.20 to .29)	.73	–.89[b] (–1.62 to –.16)	.02
Network constraint	1.22 (–1.52 to 3.95)	.38	–1.21 (–3.84 to 1.40)	.36	2.43 (–1.99 to 6.86)	.28

[a]In the analysis, out-degree and in-degree were log-transformed. All models controlled for the outcome before the intervention, the treatment group indicator, age, race, education status, and BMI.

[b]Significant at $P<.05$.

Figure 3. Estimated indirect effect from in-degree to change in self-efficacy to weight loss.



Discussion

Principal Results

In this study, we used social network analysis to analyze the interaction data from a pilot study that delivered a dietary and physical activity intervention to a group of participants of low SES via Facebook. We mapped out participants' interaction networks over the 12-week intervention period and linked participants' network characteristics to their behavioral and psychological outcomes such as weight change, self-efficacy, dietary knowledge, as well as perceived social support. Our findings suggest that there is great heterogeneity in the ways and degree to which participants engage in the intervention. Participants engaged at different levels—some participants posted and commented much more than other participants during the intervention, creating a core-periphery structure in the interaction network. Participants also engaged in different ways—some participants had few original posts but commented on others frequently, some participants did not make many comments to others but received many comments, and some participants were more embedded in the network with their network neighbors tightly connecting to each other. Individuals' network characteristics in the interaction network are predictive of their various study outcomes. Specifically, we found that the number of posts, comments, and reactions one made was directly associated with weight loss, the number of comments one received significantly predicted change in self-efficacy, and the degree to which one's network neighbors are tightly connected with each other also weakly predicted change in perceived social support. The interactions generated through the intervention may change participants' weight loss outcomes by affecting their psychological mediators. One mechanism we found was that receiving comments/replies is associated with weight loss through increase in a participant's self-efficacy.

Implications

Here, we demonstrate the potential of using social network analysis to understand the social processes of behavior change within the web-based weight loss intervention. Our findings shed light on the social processes participants experienced during the web-based weight loss intervention and how these processes might affect their psychological and behavioral outcomes, which can inform the design of future interventions to achieve better outcomes. For example, the interventionist can monitor the interaction network throughout the intervention

and deliberately reach out to those who are less likely to engage. The interventionist could also encourage participants to comment and reply to others more often to boost other participants' self-efficacy. Moreover, the interventionist could focus on design features to facilitate the formation of cohesive groups, which might increase perceived social support. Finally, the interventionist could train actively engaging participants to be a peer leader who can help with engaging other participants.

Comparison With Prior Work

While there are ample studies applying the concept of social network analysis in health and health care settings [39-42], not many have been used in the context of web-based behavioral interventions and fewer have specifically focused on weight loss interventions. Our study contributes to this body of literature by improving our understanding of the mechanisms through which social processes generated through intervention affect participants' psychological and behavioral outcomes. While participant engagement in the Facebook group used in this study exceeded that in previous studies using similar formats targeting diet or physical activity [43,44], similar to previous studies and social media research in other domains, we found great heterogeneity in ways and levels that participants engage in Facebook discussions [45-48]. Consistent with previous studies, we also found that engagement during the intervention was positively associated with weight loss [49,50]. Furthermore, we found that an individual's various network characteristics are associated with other important psychological outcomes—receiving comments is positively associated with participants' changes in self-efficacy, and embedding in a more cohesive network (one's network neighbors more tightly connected with each other) is likely to have a positive association with changes in perceived social support. While not studied in the web-based behavioral intervention context, this is consistent with social network literature from other fields [51,52]. Finally, our mediation analysis shows an interesting path explaining how the intervention might affect participants' weight loss outcome—receiving comments (in-degree) during the intervention is likely to boost one's self-efficacy, which resulted in greater weight loss. While the separate links from in-degree to self-efficacy or from self-efficacy to weight loss have been established and validated in other contexts [53,54], to our knowledge, this is the first study to establish this path in web-based behavioral weight loss intervention, which further points to the potential of using social network analysis to understand the mechanisms through which web-based behavioral weight loss intervention affect participants' psychological and behavioral outcomes.

Limitations and Future Work

This study has several limitations that point to avenues for future research. First, our sample size was small (N=47) and our participants were predominantly female (44/47, 94%). This limits the statistical power to detect the intervention effects as well as the generalizability of our results to a larger or more gender-balanced population, following a longstanding pattern in weight loss studies of difficulty recruiting male participants [55]. Similarly, we used income as a proxy measure for SES in the inclusion criteria. Given that 47% (22/47) of our participants reported attaining a college degree or advanced degree, we cannot generalize our results to individuals with low levels of education. Future studies should include more male participants and participants with low levels of education to further explore the efficacy of the intervention and how subjects may respond differently. Second, this study did not fully tease out all possible confounding factors and thus cannot establish causality. For instance, participants with higher baseline levels of motivation may be more successful at losing weight and more likely to post comments. Future studies may include more baseline characteristics and utilize randomized controlled designs to better establish causality. Third, in our analysis, the interaction network and associated network measures were aggregated over a 12-week intervention period; thus, we did not fully explore the interaction dynamics during the intervention. Future studies could study how networks change over time and unpack the temporal dynamics between networks and the study outcomes. Fourth, the network in our analysis is primarily constructed from posts and comments and we did not fully explore other relationships such as reactions and views (ie, we did not consider the number of reactions or views one received). While comments and posts have been considered more valuable than other engagement (eg, likes and "lurking") as they are more cognitively demanding [30], other engagements potentially comprise a substantial proportion of social media use and thus warrant more careful consideration in future studies [56]. Finally, when constructing the interaction network, we did not consider the content of the conversation, which is hypothesized to have different effects on the participants (eg, informational conversation may increase dietary knowledge while emotional support conversation may increase perceived social support). Future studies may employ qualitative analysis and natural language processing to further distinguish the network ties with different content.

Conclusions

In this pilot study, we constructed participants' interaction networks by using data from a feasibility trial of a web-based weight loss intervention delivered via Facebook and linked their network characteristics with changes in several important study outcomes of interest such as self-efficacy, social support, and weight. Our results point to the potential of using social network analysis to understand the social processes and mechanisms through which web-based behavioral interventions affect participants' psychological and behavioral outcomes. Future studies are warranted to validate our results and further explore the relationship between network dynamics and study outcomes in similar and larger trials.

Acknowledgments

This study was funded by the National Cancer Institute via the Case Comprehensive Cancer Center (CON501837).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Sensitivity analyses.
[DOCX File , 642 KB - jmir_v23i4e24690_app1.docx]

References

1. Managing overweight and obesity in adults: systematic evidence review from the obesity expert panel. National Heart, Lung, and Blood Institute. 2013 Nov. URL: https://www.nhlbi.nih.gov/health-topics/managing-overweight-obesity-in-adults [accessed 2020-11-03]
2. McLaren L. Socioeconomic status and obesity. Epidemiol Rev 2007;29:29-48. [doi: 10.1093/epirev/mxm001] [Medline: 17478442]
3. Sommer I, Griebler U, Mahlknecht P, Thaler K, Bouskill K, Gartlehner G, et al. Socioeconomic inequalities in non-communicable diseases and their risk factors: an overview of systematic reviews. BMC Public Health 2015 Sep 18;15:914 [FREE Full text] [doi: 10.1186/s12889-015-2227-y] [Medline: 26385563]
4. Stephens SK, Cobiac LJ, Veerman JL. Improving diet and physical activity to reduce population prevalence of overweight and obesity: an overview of current evidence. Prev Med 2014 May;62:167-178. [doi: 10.1016/j.ypmed.2014.02.008] [Medline: 24534460]
5. Beauchamp A, Backholer K, Magliano D, Peeters A. The effect of obesity prevention interventions according to socioeconomic position: a systematic review. Obes Rev 2014 Jul;15(7):541-554. [doi: 10.1111/obr.12161] [Medline: 24629126]
6. Hillier-Brown FC, Bambra CL, Cairns J, Kasim A, Moore HJ, Summerbell CD. A systematic review of the effectiveness of individual, community and societal-level interventions at reducing socio-economic inequalities in obesity among adults. Int J Obes (Lond) 2014 Dec;38(12):1483-1490 [FREE Full text] [doi: 10.1038/ijo.2014.75] [Medline: 24813369]
7. McGill R, Anwar E, Orton L, Bromley H, Lloyd-Williams F, O'Flaherty M, et al. Erratum to: Are interventions to promote healthy eating equally effective for all? Systematic review of socioeconomic inequalities in impact. BMC Public Health 2015 Sep 15;15:894 [FREE Full text] [doi: 10.1186/s12889-015-2162-y] [Medline: 26373840]
8. Social media fact sheet. Pew Research Center Internet and Technology. 2019 Jun 12. URL: http://www.pewinternet.org/fact-sheet/socialmedia/ [accessed 2020-11-04]
9. Baruth M, Sharpe PA, Parra-Medina D, Wilcox S. Perceived barriers to exercise and healthy eating among women from disadvantaged neighborhoods: results from a focus groups assessment. Women Health 2014;54(4):336-353 [FREE Full text] [doi: 10.1080/03630242.2014.896443] [Medline: 24617795]
10. Griffiths F, Lindenmeyer A, Powell J, Lowe P, Thorogood M. Why are health care interventions delivered over the internet? A systematic review of the published literature. J Med Internet Res 2006 Jun 23;8(2):e10 [FREE Full text] [doi: 10.2196/jmir.8.2.e10] [Medline: 16867965]
11. Fox S. The social life of health information. Pew Research Center. 2014 Jan 15. URL: https://www.pewresearch.org/fact-tank/2014/01/15/the-social-life-of-health-information/ [accessed 2020-11-15]
12. Heo Y, Park J, Kim J, Park H. The emerging viewertariat in South Korea: The Seoul mayoral TV debate on Twitter, Facebook, and blogs. Telematics and Informatics 2016 May;33(2):570-583. [doi: 10.1016/j.tele.2015.08.003]
13. Rus HM, Cameron LD. Health Communication in Social Media: Message Features Predicting User Engagement on Diabetes-Related Facebook Pages. Ann Behav Med 2016 Oct;50(5):678-689. [doi: 10.1007/s12160-016-9793-9] [Medline: 27059761]
14. Nam Y, Lee Y, Park HW. Measuring web ecology by Facebook, Twitter, blogs and online news: 2012 general election in South Korea. Qual Quant 2014 Mar 16;49(2):675-689. [doi: 10.1007/s11135-014-0016-9]
15. Waring ME, Jake-Schoffman DE, Holovatska MM, Mejia C, Williams JC, Pagoto SL. Social Media and Obesity in Adults: a Review of Recent Research and Future Directions. Curr Diab Rep 2018 Apr 18;18(6):34. [doi: 10.1007/s11892-018-1001-9] [Medline: 29671135]
16. Herring SJ, Cruice JF, Bennett GG, Davey A, Foster GD. Using technology to promote postpartum weight loss in urban, low-income mothers: a pilot randomized controlled trial. J Nutr Educ Behav 2014;46(6):610-615 [FREE Full text] [doi: 10.1016/j.jneb.2014.06.002] [Medline: 25069621]
17. Herring SJ, Cruice JF, Bennett GG, Rose MZ, Davey A, Foster GD. Preventing excessive gestational weight gain among African American women: A randomized clinical trial. Obesity (Silver Spring) 2016 Jan;24(1):30-36 [FREE Full text] [doi: 10.1002/oby.21240] [Medline: 26592857]
18. Turner-McGrievy GM, Tate DF. Weight loss social support in 140 characters or less: use of an online social network in a remotely delivered weight loss intervention. Transl Behav Med 2013 Sep;3(3):287-294 [FREE Full text] [doi: 10.1007/s13142-012-0183-y] [Medline: 24073180]

19. Cavallo DN, Tate DF, Ward DS, DeVellis RF, Thayer LM, Ammerman AS. Social support for physical activity-role of Facebook with and without structured intervention. Transl Behav Med 2014 Dec;4(4):346-354 [FREE Full text] [doi: 10.1007/s13142-014-0269-9] [Medline: 25584083]
20. Leahey T, Kumar R, Weinberg B, Wing R. Teammates and social influence affect weight loss outcomes in a team-based weight loss competition. Obesity (Silver Spring) 2012 Jul;20(7):1413-1418 [FREE Full text] [doi: 10.1038/oby.2012.18] [Medline: 22310234]
21. Hansen MT. The Search-Transfer Problem: The Role of Weak Ties in Sharing Knowledge across Organization Subunits. Administrative Science Quarterly 1999 Mar;44(1):82. [doi: 10.2307/2667032]
22. Alexiev, AS, Jansen, JJ, Van den Bosch, FAJ, Volberda, HW. Top management team advice seeking and exploratory innovation: The moderating role of TMT heterogeneity. Journal of Management Studies 2010 Jan 29;47(7):1343-1364. [doi: 10.1111/j.1467-6486.2010.00919.x]
23. Stewart SA, Abidi SSR. Applying social network analysis to understand the knowledge sharing behaviour of practitioners in a clinical online discussion forum. J Med Internet Res 2012 Dec 04;14(6):e170 [FREE Full text] [doi: 10.2196/jmir.1982] [Medline: 23211783]
24. Burt RS. Structural Holes: The Social Structure of Competition. Cambridge, MA: Harvard University Press; Aug 11, 1995:1-324.
25. Lin N. Building a theory of social capital. Connections 1999;22(1):28-51. [doi: 10.4324/9781315775357-27]
26. Hall A, Wellman B. Social networks and social support. Social support and health. Cambridge, MA: Academic Press; 1985. URL: https://psycnet.apa.org/record/1985-97489-002 [accessed 2020-11-03]
27. Wellman B, Gulia M. The network basis of social support: A network is more than the sum of its ties. In: Networks in the Global Village. Philadelphia, PA: Routledge; 1999:83-118.
28. Kelman HC. Compliance, identification, and internalization three processes of attitude change. Journal of Conflict Resolution 2016 Jul;2(1):51-60. [doi: 10.1177/002200275800200106]
29. Friedkin NE. A Structural Theory of Social Influence. Cambridge, United Kingdom: Cambridge University Press; 2006:1-254.
30. Cavallo D, Martinez R, Hooper M, Flocke S. Feasibility of a social media-based weight loss intervention designed for low-SES adults. Transl Behav Med 2020:ibaa070. [doi: 10.1093/tbm/ibaa070]
31. McKee HA, Porter JE. The Ethics of Internet Research: A Rhetorical, Case-Based Process. New York, USA: Peter Lang Inc., International Academic Publishers; New edition; 2009.
32. Sams S, Lim YS, Park HW. e-Research applications for tracking online socio-political capital in the Asia-Pacific region. Asian Journal of Communication 2011 Oct;21(5):450-466. [doi: 10.1080/01292986.2011.594897]
33. Pagoto S, Nebeker C. How scientists can take the lead in establishing ethical practices for social media research. J Am Med Inform Assoc 2019 Apr 01;26(4):311-313 [FREE Full text] [doi: 10.1093/jamia/ocy174] [Medline: 30698793]
34. Pagoto S, Waring ME, Xu R. A Call for a Public Health Agenda for Social Media Research. J Med Internet Res 2019 Dec 19;21(12):e16661 [FREE Full text] [doi: 10.2196/16661] [Medline: 31855185]
35. Thomas S, Reading J, Shephard R. Revision of the Physical Activity Readiness Questionnaire (PAR-Q). Can J Sport Sci 1992 Dec;17(4):338-345. [Medline: 1330274]
36. Grytics analytical platform. URL: https://grytics.com/ [accessed 2021-03-29]
37. Hu Y. Efficient, high-quality force-directed graph drawing. Mathematica Journal 2005;10(1):37-71. [doi: 10.1007/978-1-4939-2864-4_648]
38. Batagelj V, Zaversnik M. An O(m) algorithm for cores decomposition of networks. Advances in Data Analysis and Classification. 2011. URL: https://arxiv.org/abs/cs/0310049 [accessed 2021-03-29]
39. Chambers D, Wilson P, Thompson C, Harden M. Social network analysis in healthcare settings: a systematic scoping review. PLoS One 2012;7(8):e41911 [FREE Full text] [doi: 10.1371/journal.pone.0041911] [Medline: 22870261]
40. Valente TW. Network interventions. Science 2012 Jul 06;337(6090):49-53. [doi: 10.1126/science.1217330] [Medline: 22767921]
41. Kaonga NN, Labrique A, Mechael P, Akosah E, Ohemeng-Dapaah S, Sakyi Baah J, et al. Using social networking to understand social networks: analysis of a mobile phone closed user group used by a Ghanaian health team. J Med Internet Res 2013 Apr 03;15(4):e74 [FREE Full text] [doi: 10.2196/jmir.2332] [Medline: 23552721]
42. Takahashi Y, Uchida C, Miyaki K, Sakai M, Shimbo T, Nakayama T. Potential benefits and harms of a peer support social network service on the internet for people with depressive tendencies: qualitative content analysis and social network analysis. J Med Internet Res 2009 Jul 23;11(3):e29 [FREE Full text] [doi: 10.2196/jmir.1142] [Medline: 19632979]
43. Cavallo DN, Sisneros JA, Ronay AA, Robbins CL, Jilcott Pitts SB, Keyserling TC, et al. Assessing the Feasibility of a Web-Based Weight Loss Intervention for Low-Income Women of Reproductive Age: A Pilot Study. JMIR Res Protoc 2016 Feb 26;5(1):e30 [FREE Full text] [doi: 10.2196/resprot.4865] [Medline: 26920252]
44. Pagoto SL, Waring ME, Schneider KL, Oleski JL, Olendzki E, Hayes RB, et al. Twitter-Delivered Behavioral Weight-Loss Interventions: A Pilot Series. JMIR Res Protoc 2015 Oct 23;4(4):e123 [FREE Full text] [doi: 10.2196/resprot.4864] [Medline: 26500186]
45. Valle CG, Tate DF. Engagement of young adult cancer survivors within a Facebook-based physical activity intervention. Transl Behav Med 2017 Dec;7(4):667-679 [FREE Full text] [doi: 10.1007/s13142-017-0483-3] [Medline: 28374211]

46. Merchant G, Weibel N, Patrick K, Fowler JH, Norman GJ, Gupta A, et al. Click "like" to change your behavior: a mixed methods study of college students' exposure to and engagement with Facebook content designed for weight loss. J Med Internet Res 2014 Jun 24;16(6):e158 [FREE Full text] [doi: 10.2196/jmir.3267] [Medline: 24964294]
47. Napolitano MA, Whiteley JA, Mavredes MN, Faro J, DiPietro L, Hayman LL, et al. Using social media to deliver weight loss programming to young adults: Design and rationale for the Healthy Body Healthy U (HBHU) trial. Contemp Clin Trials 2017 Sep;60:1-13 [FREE Full text] [doi: 10.1016/j.cct.2017.06.007] [Medline: 28611007]
48. Barnett GA, Xu WW, Chu J, Jiang K, Huh C, Park JY, et al. Measuring international relations in social media conversations. Government Information Quarterly 2017 Jan;34(1):37-44. [doi: 10.1016/j.giq.2016.12.004]
49. Turner-McGrievy GM, Tate DF. Weight loss social support in 140 characters or less: use of an online social network in a remotely delivered weight loss intervention. Transl Behav Med 2013 Sep;3(3):287-294 [FREE Full text] [doi: 10.1007/s13142-012-0183-y] [Medline: 24073180]
50. Hales SB, Davidson C, Turner-McGrievy GM. Varying social media post types differentially impacts engagement in a behavioral weight loss intervention. Transl Behav Med 2014 Dec;4(4):355-362 [FREE Full text] [doi: 10.1007/s13142-014-0274-z] [Medline: 25584084]
51. Vardaman JM, Amis JM, Dyson BP, Wright PM, Van de Graaff Randolph R. Interpreting change as controllable: The role of network centrality and self-efficacy. Human Relations 2012 Jun 28;65(7):835-859. [doi: 10.1177/0018726712441642]
52. Burt RS, Lin N, Cook K. Closure as social capital. In: Social capital: Theory and research. Philadelphia, PA: Routledge; 2001:31-55.
53. Linde JA, Rothman AJ, Baldwin AS, Jeffery RW. The impact of self-efficacy on behavior change and weight change among overweight participants in a weight loss trial. Health Psychol 2006 May;25(3):282-291. [doi: 10.1037/0278-6133.25.3.282] [Medline: 16719599]
54. Roach JB, Yadrick M, Johnson JT, Boudreaux L, Forsythe WA, Billon W. Using self-efficacy to predict weight loss among young adults. Journal of the American Dietetic Association 2003 Oct;103(10):1357-1359. [doi: 10.1016/s0002-8223(03)01072-1]
55. Franz MJ, VanWormer JJ, Crain AL, Boucher JL, Histon T, Caplan W, et al. Weight-loss outcomes: a systematic review and meta-analysis of weight-loss clinical trials with a minimum 1-year follow-up. J Am Diet Assoc 2007 Oct;107(10):1755-1767. [doi: 10.1016/j.jada.2007.07.017] [Medline: 17904936]
56. Edelmann N. What is lurking? A literature review of research on lurking. In: The Psychology of Social Networking. Berlin, Boston: Sciendo Migration; 2015:159-174.

Abbreviations

SES: socioeconomic status

Edited by R Kukafka; submitted 30.09.20; peer-reviewed by S Stewart, HW Park; comments to author 01.11.20; revised version received 19.11.20; accepted 12.03.21; published 09.04.21.

Please cite as:
Xu R, Cavallo D
Social Network Analysis of the Effects of a Social Media–Based Weight Loss Intervention Targeting Adults of Low Socioeconomic Status: Single-Arm Intervention Trial
J Med Internet Res 2021;23(4):e24690
URL: https://www.jmir.org/2021/4/e24690
doi:10.2196/24690
PMID:33835033

Original Paper

Perceptions of and Opinions on a Computerized Behavioral Activation Program for the Treatment of Depression in Young People: Thematic Analysis

Lucy Tindall[1], MSc; Paul Toner[2], PhD; Antonina Mikocka-Walus[3], PhD; Barry Wright[1], MD

[1]Department of Health Sciences, University of York, York, United Kingdom
[2]School of Psychology, Queen's University Belfast, Belfast, United Kingdom
[3]School of Psychology, Deakin University, Victoria, Australia

Corresponding Author:
Lucy Tindall, MSc
Department of Health Sciences
University of York
Heslington
York, YO105DD
United Kingdom
Phone: 44 7976774636
Email: lucy.tindall@york.ac.uk

Abstract

Background: Depression is one of the leading causes of illness and disability in young people, with approximately 20% having experienced a depressive episode by the age of 18 years. Behavioral activation (BA), a National Institute for Health and Care Excellence–recommended treatment for adults with depression, has shown preliminary support for its use with young people. BA may have the potential to be adapted and delivered in a computerized format to address the barriers often associated with young people accessing support. Despite the benefits of adopting computerized therapy delivery, the limited effectiveness of some programs has been attributed to a failure to tailor interventions to patients and practices. Therefore, while developing new treatments, it is important that target users be involved in the intervention design.

Objective: This qualitative study aims to explore the views and preferences of young people and health care professionals regarding the development of a new computerized BA therapy for young people with low mood or depression, to ensure that the therapy was suitable for the target user.

Methods: Semistructured focus groups and individual interviews were conducted with young people (those with experience in accessing support and those without) and health care professionals regarding the development of a new computerized BA therapy for young people with low mood or depression. The data were analyzed using thematic analysis.

Results: A total of 27 individuals, comprising both health care professionals and young people, participated in this study. Vital information pertaining to the important components of a new therapy, including its presentation, delivery, and content, was collected.

Conclusions: Variations in perspectives highlighted the need to adopt a systemic approach in therapy development by considering the opinions of young people with and without experience in accessing mental health support and health care professionals.

(J Med Internet Res 2021;23(4):e19743) doi:10.2196/19743

KEYWORDS

depression; qualitative; thematic analysis; young people; health care professionals; computerized therapies

Introduction

Background

Approximately 20% of young people experience at least one depressive episode by the age of 18 years [1], making depression one of the leading causes of illness and disability in this group [2]. Despite this, help-seeking is limited only to approximately 35% of young people [1] because of factors including stigma [3,4], negative attitudes about help-seeking [4], accessibility [3], and reluctance to engage one-to-one with a therapist [5]. Much attention has been placed on computerized delivery of

therapy, in particular cognitive behavioral therapy (CBT) [6-8], which removes some of these barriers. One alternative to CBT is behavioral activation (BA), which encourages increased engagement in adaptive activities and decreased engagement in activities that might maintain or increase depression risk [9]. Owing to its empirical support [10-13], BA is a National Institute for Health and Care Excellence–recommended treatment for adults with depression [14]. Given its demonstrated effectiveness and simple delivery, BA may have the potential to be adapted for use with young people and delivered in a computerized format.

Despite the evident benefits of adopting a computerized approach to therapy delivery, including increasing therapy accessibility, availability [15], and anonymity [16]; reducing stigmatization [17]; and providing those reluctant to engage one-to-one with a therapist with access to care [18], the limited effectiveness of some programs has been attributed to a failure to tailor interventions to patients and practices [19,20]. Therefore, in the development of new treatments, it is important that those for whom an intervention is being designed are involved in its development [21,22], and qualitative research is particularly useful in this context.

When designing a computerized BA program, those deemed the best to inform the development include not only those who may use such a program but also those involved in treatment delivery. Research suggests that one important process that may enhance treatment effectiveness is the incorporation of young people's views in treatment design to ensure that the developmental needs of the users are considered [23]. However, some depression and anxiety treatments for young people are simply downward extensions of adult therapies, as reflected in both their design and delivery [24,25]. Therefore, the needs and capacities of young people are not always considered when such information is being collected from adults, including parents and pediatric health care professionals [26]. This omission to collect information firsthand may result in the misinterpretation of young peoples' needs and a focus on the needs of adult representatives instead [27], highlighting the need to include young people when developing interventions designed for them.

In research that has incorporated the views of young people, differences have been evident between their perceptions and those of health care providers. When reviewing research related to the experiences of care systems, health care professionals' priorities contrasted with the key points highlighted by young people [27]. Furthermore, an examination of illness experiences through interviews with 7 children who had been hospitalized showed that illnesses had different meanings for young people compared with those for professional caregivers [28]. Here, both the illnesses and the mechanisms of recovery were seen in very different ways. Although young people had subjective and personal knowledge of illnesses derived from their own experiences, health care professionals based their knowledge on easily measured objective symptoms. Therefore, there is a need to encapsulate the perspectives of both health care professionals and young people in the development of a computerized BA program.

Objectives

This qualitative study explores the views and preferences of young people and health care professionals regarding the development of a new computerized BA therapy for young people with low mood or depression. The information gathered about the treatment components that such a therapy should include and how it should be presented and delivered inform the development of the new treatment.

Methods

Overview

This research adopted an exploratory, qualitative design using a critical realist approach [29] and was conducted within 4 UK National Health Service (NHS) trusts. Ethical approval was granted by the Health Research Authority (reference: 16/EM/0420) and the Department of Health Sciences Research Ethics Committee at the University of York.

Participants, Sampling, and Recruitment

In total, 3 groups of individuals participated: young people from a community sample, young people from a service user sample, and health care professionals with experience of working with young people with low mood or depression. Participants were identified using a combination of purposive and snowball sampling approaches.

Young People

As many young people experiencing a depressive disorder do not seek help [1], a community sample of young people from a local school within the remits of one of the participating NHS trusts was invited to a focus group. A purposive sampling approach was used, stratifying those identified by age and gender to increase the representativeness of the participants. In addition, a sample of young people receiving support for depression within 3 local child and adolescent mental health services (CAMHS) were identified by health care professionals and invited to attend face-to-face, semistructured interviews with the researcher. Recruiting young people from both a community and a health care sample allowed maximum variation in the views collected. The recruited participants were males and females aged between 11 and 16 years. This ensured that those recruited represented a sample similar to the target group that the eventual program was to be designed for.

All prospective participants received information about the study, with those expressing interest providing written assent alongside consent from a parent or guardian.

Health Care Professionals

A sample of health care professionals with experience of working with young people with low mood or depression from 2 CAMHS was invited to a focus group or an individual interview if more convenient. Anyone expressing interest provided written consent for participation.

Data Collection

All focus groups and interviews were conducted by the lead researcher within the schools and CAMHS sites where recruitment occurred. Before all interviews and focus groups,

attendees completed a short questionnaire—young people were asked to supply basic demographic information and health care professionals provided information about their clinical practice.

The interviews and focus groups closely followed 1 of 2 interview schedules (one for young people and one for health care professionals; Multimedia Appendix 1), comprising semistructured, open-ended questions. The interview schedules for young people were based on collecting information pertaining to 4 different domains: previous experiences of web-based help and opinions regarding the new program development, the content and treatment-related activities to be considered for inclusion in the new program, delivery-specific information (eg, the number and structure of sessions, access, and program presentation), and parental involvement. Although a similar structure was adopted in the interview schedules for health care professionals, they were also asked questions related to the new program's positioning in the care pathway and about perceived support requirements for program completion. With permission, all interviews and focus groups were audio recorded digitally and transcribed verbatim.

During the focus groups and interviews, participants completed 2 bespoke activities specifically developed for the research. The first activity, which was completed by all, was used to elicit further information regarding the content to be considered for inclusion in the new program, with individuals asked to indicate what they deemed to be important from a list of potential options (eg, activity scheduling, activity monitoring, and relapse prevention). Definitions of terms were provided as part of this activity to aid in participants' understanding. This list was generated from a review of the treatment components often included in BA [30] and the findings from a previous systematic review and meta-analysis of BA used in young people [31]. The second activity, which was completed by young people only, was used as an additional tool to collect delivery-specific information. This involved young people ranking 10 factors that they felt would make the program more attractive to young people (interactivity, colorfulness, etc) in order of importance. The activities allowed for methodological triangulation [32], made participation more enjoyable, and allowed individuals to provide information in various ways.

Data Analysis

Data were analyzed using thematic analysis, which was selected to provide rich and detailed insights about a given topic [33]. The analysis was both inductive and deductive, with themes or subthemes generated from both the raw data provided by participants and from theory and previous research. Thematic analysis comprises 6 phases: familiarization with the data, generating initial codes, searching for themes, reviewing themes, defining and naming themes, and reporting production [33]. By closely following these phases, themes were identified from what participants perceived to be the important components of a computerized BA program for use with young people experiencing low mood or depression. All analyses were completed by hand using Microsoft Word, and no specific qualitative software packages were used.

Results

Overview

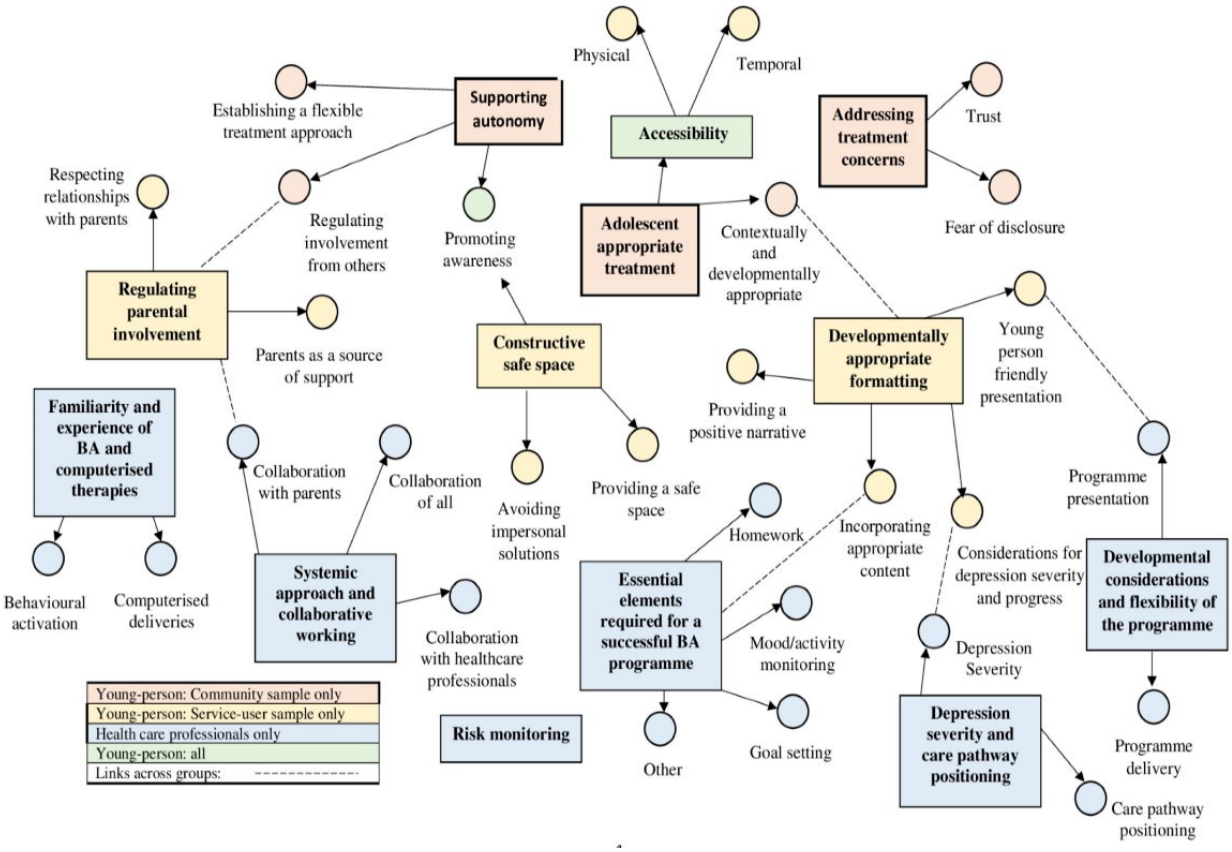
In total, 27 individuals comprising young people from a community sample, young people from a service user sample, and health care professionals attended interviews and focus groups with the lead researcher. Participant demographic information is presented in Table 1, whereas the themes identified in the thematic analysis and their interrelations across groups are presented in Figure 1.

Table 1. Participant sample characteristics (N=27).

Participant characteristics	Young person: community sample (n=9)	Young person: service use sample (n=9)	Health care professional sample (n=9)
Gender, n (%)			
Male	5 (55)	1 (11)	2 (22)
Female	4 (45)	8 (89)	7 (78)
Ethnicity, n (%)			
White-British	8 (88)	9 (100)	8 (88)
White-Irish	N/A[a]	N/A	1 (11)
Black-African	1 (11)	N/A	N/A
Young-person participants only			
Age (years), mean (SD; range)	13.8 (0.83; 13-15)	15.2 (1.09; 13-16)	N/A
Experience of low mood or depression, n (%)			
Yes	2 (22)	9 (100)	N/A
No	7 (78)	N/A	N/A
Health care professional participants only			
Age group (years), n (%)			
18-30	N/A	N/A	2 (22)
31-50	N/A	N/A	2 (22)
51-70	N/A	N/A	5 (56)
Highest education level, n (%)			
Secondary education	N/A	N/A	1 (11)
Trade or technical or vocational	N/A	N/A	1 (11)
Professional degree	N/A	N/A	1 (11)
Bachelor's degree	N/A	N/A	1 (11)
Master's degree	N/A	N/A	3 (34)
Doctorate	N/A	N/A	2 (22)
Job role, n (%)			
Nurse	N/A	N/A	4 (45)
Clinical psychologist	N/A	N/A	2 (22)
Student mental health worker	N/A	N/A	1 (11)
Support worker	N/A	N/A	1 (11)
Art therapist	N/A	N/A	1 (11)
Practice years, n (%)			
0-5	N/A	N/A	3 (33)
11-15	N/A	N/A	1 (11)
20+	N/A	N/A	5 (56)
Computer proficiency, n (%)			
Fair	N/A	N/A	2 (22)
Good	N/A	N/A	4 (45)
Very good	N/A	N/A	3 (33)

[a]N/A: not applicable.

Figure 1. Thematic map of young person participants and health care professionals' perspectives. BA: behavioral activation.



Young People: Community Sample

In total, 3 main themes were identified: supporting autonomy, addressing treatment concerns, and adolescent-appropriate treatment.

Supporting Autonomy

Most young people in the community sample were unsure of how to access support if required, with only 2 having ever received any support—one for bullying support and another for dealing with their parents' divorce. Most assumed that, if they needed any support, their parents would access it on their behalf. None of the users had ever accessed any web-based support, with the majority unaware of its existence, thus highlighting the need for increased awareness of its availability:

> *I think that's why a lot of kids nowadays are feeling more pressured into committing suicide and things like that because it's like "oh I can't get any help" because they don't know what to do.* [Male, community, aged 13 years]

Young people wanted the opportunity for autonomy if using the new program, highlighting the need for flexibility and wanting to make treatment decisions. Therefore, they wanted much of the program to be optional, with users choosing what to view and when. Examples included deciding whether to disclose information about experiences of low mood or depression, selecting which activities to participate in, and deciding whether to complete additional follow-up sessions if available. They also wanted to decide how long they spent on sessions and the frequency of logging on. However, if flexibility with session length was not possible, approximately 30 minutes per session was regarded as reasonable.

Wanting to be independent and able to make treatment-related decisions, many young people did not want parents to be involved, reporting that such involvement could be detrimental to improving mood. Particular concerns included feeling that parents would not understand young person–specific issues:

> *I'm not being disrespectful to them, but they'll think they know everything about kids and stuff, but they really like, they don't.* [Male, community, aged 13 years]

Friends were suggested as an alternative support to parents, but shortcomings of this approach were identified, including concerns that this could place great pressure on friends who might worry too much, who may not take things seriously, or who could be patronizing. Therefore, if any risk was presumed, young people felt that parents needed to be informed. Consequently, they contended that the new program should include some information for parents but that this should be both minimal and optional, so young people could retain autonomy.

Addressing Treatment Concerns

Concerns were raised about how young people would be treated if they accessed support. In particular, they worried that information disclosed would be reported back to parents who may then treat their child differently and place restrictions on them that could worsen their low mood:

> *...your mum and dad would be like "right I'm not letting you out the house, I'm not letting you do this, I'm not letting you do that" because that kind of puts you down more.* [Male, community, aged 13 years]

Such concerns were not limited to parents, with some worried about what would happen if they were to report low mood to health care professionals, particularly general practitioners. For most participants, the prospect of computerized therapy delivery allayed some of these concerns. However, for one individual, even accessing therapy in a computerized format could lead to others finding out about their low mood and treating them differently. Specifically, this individual was worried that if they felt particularly low and accessed computerized therapy, people might arrive at their house to tell them "don't kill yourself."

Young people wanted assurance that, if accessing computerized therapy, confidentiality would be maintained. They suggested having good security with individual usernames and passwords required for access, incorporating automatic log-offs after short periods of inactivity, and ensuring no personal information requests were made. They also wanted the program's name to be discrete and not indicative of its purpose.

A consensus emerged that young people would only disclose information about how they were feeling to another person if trust was both established and retained. One young person had concerns about trust if therapy was delivered in a computerized format, stating that "anyone can set up a website." This was different for others, with one individual highlighting that if therapy was set up by a reputable source, trust could be gained. Second, the anonymity of a computerized approach meant that establishing trust was not an issue.

Adolescent-Appropriate Treatment

The community sample felt that the adolescent period was where computerized therapy could be of most importance, particularly for those already experiencing low mood or depression. They outlined specific stressors, including the transition to secondary school, changes to curriculum and workload, and working toward and completing the General Certificate of Secondary Education.

Overall, they felt that 2 or 3 different presentations were needed to suit the target age range, with developmental level guiding the program presentation and chronological age guiding the content (ie, to ensure the therapy did not place greater pressure on young people at already stressful times, such as during examination periods):

> *A low level sixteen-year-old might have the same ability as a high level twelve-year-old.* [Female, community, aged 14 years]

The community sample identified communication, goal setting, activity scheduling, and identifying barriers as the most important components for inclusion. In particular, they felt that communication would help users to develop their social skills and described how the inclusion of links to clubs would aid activity scheduling. Additional content for consideration included games or puzzles, if these were not *childish*, and videos, if these were nonpatronizing, modern, showed depression in a "serious way," and included actors that users could identify with. Several young people wanted homework to be included but only if it were called something else and minimal, thus considering other demands on users' time. If information about depression was included, young people wanted this to incorporate depression statistics, information about coping strategies, and current treatments. Advice from health care professionals was requested as part of any problem solving. Although recording activities and moods was the least favored component for inclusion (with some participants concerned that this could lead to information disclosure), others wanted this to be included to allow connections to be made between periods and moods.

The colorfulness of the new program, including information about depression and having additional follow-up sessions, was regarded as the most important component in making the new program attractive to users, whereas having printable handouts and homework activities was considered less important. Participants also wanted the program to be sensitive, fun, and relaxed and to provide adequate information—not too much so that users would be overwhelmed or too little so that they would not take it seriously.

Concerns were expressed that not all young people would be able to access a computerized program with proposals made to place it in accessible locations (eg, general practice [GP] surgeries) to ensure equal access. Furthermore, young people wanted the program to be available at all times and therefore accessible not only when most convenient but also when most required. Consequently, young people wanted the program content to be downloadable to ensure it could be viewed even if an internet connection was unavailable. Program access also had to be simple, with no requirements for young people to trawl through terms or conditions or have to "sign up now," which they described would dissuade them from using the program.

Young People: Service User Sample

A total of 4 main themes were identified: constructive safe space, developmentally appropriate formatting, accessibility, and regulation of parental involvement.

Constructive Safe Space

In total, 4 service users had accessed web-based support, searching the internet and browsing websites, generally when they felt particularly low and when alternative support was unavailable. Although 2 users had found web-based help useful, shortcomings of the approach were described, including a lack of personal rapport, automated responses described as "robotic," and a feeling that information was more directed at parents:

> *I found the [name of website] one to be a lot of phone numbers and it was a lot more directed towards parents than the child.* [Female, service user, aged 16 years]

All the service users had received face-to-face support for low mood or depression, with the majority describing how hard they had found this, particularly talking with another person about their feelings. Similar to the community sample, their main concern related to how they would be perceived, with several reporting feelings of embarrassment or worry that others would

not understand. Consequently, they felt that a computerized treatment approach addressed some of these difficulties, providing young people with anonymity that allowed them to talk freely without judgment, in a format that they were both familiar with and comfortable using:

> *...our safe space is talking through a phone, through a screen because that seems easier than talking to people.* [Female, service user, aged 15 years]

Other perceived benefits included enabling young people to access support without having to ask for it and the opportunity to receive it within their own surroundings where they felt most comfortable.

None of the service users had ever had any web-based therapies recommended to them or been directed to it as part of therapy. They were therefore keen to ensure that young people were aware of its availability and suggested advertising the new program in health care settings and schools and via social media.

Developmentally Appropriate Formatting

The consensus was that 2 versions of the new program were needed. A simpler version was proposed for younger users aged 11-13 years, with more in-depth information presented for older individuals aged 14-16 years.

The service users wanted the new program to be fun, relaxed, and nonpatronizing while being simple, so that young people would not disengage:

> *If an eleven-year-old who is anxious, low mood, stressing out about themselves is answering questions or reading words that they don't understand, it is only going to make it worse for them.* [Female, service user, aged 15 years]

Furthermore, it was highlighted that many young people, especially those with low mood or depression, tend to procrastinate; therefore, the simpler and more focused the new program, the more likely young people would be to use it.

The activities considered most important for inclusion included learning about coping, playing games or puzzles, practicing new skills, and having information about relapse prevention. Several service users had previously used mood monitoring, finding it useful in allowing them to monitor their feelings over time; therefore, this was suggested for inclusion by some users. Additional components identified included information about exercise and other conditions (eg, eating disorders), a chat option so that users could communicate with similar others, useful telephone numbers, and satisfying videos (videos designed to reduce anxiety and aid relaxation). Homework was considered the least important for inclusion, with the requisite that, if included, it had to be minimal to increase adherence and avoid people *getting bored*.

Some users wanted notifications included to remind users to complete sessions and the option to print content. However, these were not favored by all. Some users felt that as long as program content was downloadable, printable content was not required, and it would be easier to keep all information in one place, that is, on the computer. Although some users could see the benefits of notifications, they felt that these needed to be optional, with users controlling how and when to receive them.

To make the program more attractive to users, its interactivity and its availability for completion were considered the most important components, whereas homework and having printable handouts were considered less important.

Although the majority of service users wanted information about depression presented in the new program, several concerns were raised. Some worried that information of this type could make some people feel worse, and therefore, this needed to be optional. Similar concerns were also discussed regarding the recording of activities and moods and watching videos:

> *I know the videos are to show that people have overcome it, but the only other thing I would say is that I remember looking at people and stuff like that and then I'd feel bad because I'd feel like "well they've got it worse than me, so I'm just being dramatic for no reason.*" [Female, service user, aged 16 years]

The service users felt that the program length—both its session number and the time needed to complete them—required flexibility. Although they had made suggestions about this, they felt that more or less time could be spent on the program relative to how users were feeling and how they were progressing through it. Therefore, anyone noticing positive changes and feeling better would cease program use, whereas those who did not would use it for longer. They also proposed that anyone experiencing a relapse should be offered a follow-up session or be able to repeat the new program in its entirety.

Accessibility

The service users felt that through computerized therapy delivery, more young people would be aware of and more likely to access support. They reported that "nowadays everyone has access to the internet" and therefore wanted the new program to be widely available and not restricted to specific locations (eg, GP surgeries or CAMHS sites). This was deemed especially important for those who might struggle to access face-to-face support (eg, those who are too anxious or have transportation issues) and would help to address lengthy waiting lists in CAMHS:

> *The process of getting into CAMHS is like a bit of a pain with waiting lists and then maybe not having enough sessions and getting discharged before you're ready and stuff because the fact that there are so many people that need to use the service, but if it's internet, like, as many people can use it as they need I'm guessing.* [Female, service user, aged 15 years]

Although the service users felt that internet access would not be a problem for most young people, they did however recommend that all program content should be downloadable so that individuals could ensure program access even without an internet connection.

Although some users felt that it should be available to all, others felt that, to monitor risk, only those given access by a clinician should be able to use it. One individual suggested that clinicians

could send program log-in details to anyone they deemed suitable or their parents.

Although variation in the number of sessions was proposed, the consensus was that this should be dependent on the level of support required. Furthermore, some form of follow-up, to review progress and ensure therapy did not simply stop, was regarded as important. There was variation in how this could be incorporated, with some users suggesting a single session after therapy completion to be sufficient and others wanting monthly follow-ups for up to 6 months. One individual suggested making follow-ups optional and aimed at those who might not have processed all of the program content.

The service users felt that the program length needed to consider the attention spans of young people, with most deeming between 30 and 45 minutes per session as sufficient. They felt that young people might *lose focus* if sessions were too long and proposed that larger topics could be covered in multiple sessions to keep session lengths shorter.

Some users felt that weekly sessions with time to practice skills would be appropriate, with concern that users may forget what they had done in a session if longer than a week was left until the next. Others wanted a higher frequency of sessions, with both completing 2 sessions per week and completing sessions on alternate days. It was also suggested that program content should remain accessible following completion so that it could be repeated if needed.

Regulating Parental Involvement

Several service users felt that there should be no parental involvement in the new program with concerns that those not having a close relationship with their parents would not engage with a parent present. Some felt that parental involvement should be decided by the user, whereas one service user felt that parents should be informed if their child was using the program simply to show them that they were using the computer in an appropriate way.

In contrast, parental involvement was regarded positively by some users who felt that the program could be completed more successfully if parents were involved. These individuals described how this could result in parents gaining a better understanding of how their child was feeling and how parents could learn techniques to be able to provide more suitable support:

> *If parents have techniques for the child, they could then show them techniques, you know that could help them, or maybe learn not to maybe act a certain way with them and stuff.* [Male, service user, aged 16 years]

Handouts providing information about low mood or depression and coping techniques were suggested as a way of implementing parental involvement. There were also suggestions about including parents in some sessions with variations, including young people completing sessions together with parents, having parent-only and young person-only sessions completed separately, or combining these approaches with parents attending sessions at various points.

Health Care Professionals

In total, 6 main themes were identified: familiarity and experience of BA and computerized therapies, essential elements required for a successful BA program, developmental considerations and flexibility of the program, depression severity and care pathway positioning, risk monitoring and systemic approach, and collaborative working.

Familiarity and Experience of BA and Computerized Therapies

Most health care professionals were familiar with the components of BA, having used some of them—mainly mood and activity monitoring—in therapy. However, most health care professionals, unaware of the term, did not refer to this as BA. Those who were familiar with BA described how their knowledge of the treatment was derived from an adult model. Therefore, if they had applied BA principles to young people, this was through them adapting their knowledge of BA with adults and creating a bespoke treatment approach.

All health care professionals were familiar with, and valued, the increased use of computerized therapies, with some having referred young people to web-based resources while awaiting face-to-face therapy. Despite this, limitations of the approach were discussed, particularly the lack of therapist presence during therapy delivery:

> *Instilling the hope, the computer cannot instill hope.* [Health care professional, male]

Furthermore, they felt that computerized therapy delivery was not suitable for all, especially in high-risk situations, and expressed concern about what would happen to young people if they felt particularly low following completion of computerized treatment sessions.

Essential Elements Required for a Successful BA Program

Goal setting, activity scheduling, coping skills, mood or activity recording, and monitoring feelings were deemed the most important criteria for inclusion in the new program. Although goal setting was regarded as a *vital* component in assisting therapists to gauge treatment direction, health care professionals also felt it was important for young people to monitor their mood regularly and identify factors affecting it. Most users also wanted to see information about depression, problem solving, and quizzes included. However, there was concern about how components such as goal setting and problem solving could be implemented in a computerized format, given that they are person-centered and individualized approaches. Some users wanted homework to be included, but they stipulated that this needed to be presented under a different name and monitored with reminders sent to support engagement.

To increase inclusivity, health care professionals suggested presenting information in different modes and allowing users to adjust program settings to enhance ease of use. They also wanted users to be able to ask questions about anything they did not understand and have key points revisited in each session to assist understanding.

Developmental Considerations and Flexibility of the Program

It was felt that more than one presentation of the new program, based on developmental level, was needed. There was concern that some young people would not understand particular words and others, especially those with low mood or depression and lacking energy, would struggle if presented with a lot of text. Therefore, breaking words down to aid understanding, using combinations of words and pictures in the program, and embedding an audio option so that content could be read aloud were suggested.

The health care professionals wanted both session frequency and length to be tailored to each user with consideration given to energy levels, depression severity, developmental level, and the number of questions a young person had. One individual, highlighting that those with low moods might have shorter attention spans, proposed basing sessions upon graded exposure. Overall, the consensus was that approximately 8-12 sessions, delivered weekly and lasting between 30 and 45 minutes, would be optimal.

Depression Severity and Care Pathway Positioning

It was felt that those with severe low mood or depression might have difficulties using a computerized program; therefore, it would be most suited to those experiencing mild-to-moderate low mood or depression. Aligning to this, tiers 2 (early help and targeted services) and 3 (specialized services) of CAMHS and step 2 of the children and young people's improving access to psychological therapy stepped care model (low-intensity services) were suggested as areas where the new program would be best placed.

Health care professionals felt that the new program could be positioned anywhere in the care pathway. First, it could be used by young people awaiting treatment, thus reducing the number of subsequent face-to-face appointments and allowing more time to be allocated to those with more severe depression. Another recommendation was for the program to be used alongside usual care, either between or during face-to-face sessions with the therapist present. Particular benefits of this included providing young people with a new medium within their care and providing therapists with an additional tool to monitor progress. Finally, it was proposed that the new program could be offered as a *further treatment* and used before discharge from CAMHS.

Risk Monitoring

Several recommendations were made about how to identify anyone deemed at risk following computerized therapy and how to manage this situation. These included incorporating live chat or crisis numbers into the new program, allowing young people to request a call back from a health care professional if they were feeling at risk and encouraging users to take control and agree to seek additional support if required. To implement this, one health care professional suggested using contracts:

> *You could have it in the contract...so just something that they signed to say that "if you do feel at risk or you feel that you need to talk to someone that you are willing to contact the crisis team and speak to someone," even the Samaritans.* [Health care professional, female]

Systemic Approach and Collaborative Working

Therapeutic alliance was particularly important to health care professionals. They felt that during the treatment process, clinicians were required to instill hope and monitor both progress and deterioration with individual clients, providing support accordingly. They were therefore concerned about the lack of human contact inherent to computerized therapy delivery and, although supportive of the new program, wanted treatment delivery to incorporate a balance between the computer and therapist:

> *I think for me though there is nothing more powerful that the connection between the therapist and the client, you know.* [Health care professional, female]

Health care professionals felt that parental involvement was also essential, particularly regarding risk monitoring, and reported that therapies were often more successful if parents were included. However, they acknowledged that relationships with parents can be difficult, and therefore, parental support within the new program would need to be adapted based on user-parent relationships.

Finally, it was highlighted that, for treatment success, a systemic approach to therapy delivery was required with the cooperation of multiple agencies. They contended that everyone had a role during therapy delivery, and therefore, it was important to be flexible and ensure roles were clearly defined:

> *Everybody has a role to play from the preventative early intervention, the mild end, the moderate end, the severe end so it will be quite clear who does what, when, and how, and then that maxes our public money, so we're not stop starting.* [Health care professional, male]

Through collaboration, it was felt that multilevel monitoring would be in place throughout treatment completion, with young people having a variety of support contacts when required, something especially important if the risk was presumed.

Discussion

Principal Findings

This study allowed vital information to be collected from various perspectives to inform the development of a new computerized BA program for young people. Previous research suggests that differences exist between the views of young people and health care professionals [27,28], with the limited effectiveness of some programs attributed to a failure to tailor interventions to patients and practices [19,20]. Therefore, the views of both young people and health care professionals were collected and incorporated into the development of the new intervention.

The findings from this qualitative work demonstrated that both young people and health care professionals endorsed the use of computerized therapy delivery, and the advantages of the approach were identified. These included enhanced anonymity, treatment accessibility, and awareness; these findings were

concordant with previous research in the area [15,16,18]. Through the focus group and interview discussions and the associated activities completed, the contention that BA may be an acceptable treatment for delivery in this context was supported, as was support for future clinical effectiveness research. When presented with potential content for inclusion in the new intervention, based on techniques commonly delivered within BA, all content was selected by at least one-third of participants in each group. Furthermore, two of the main approaches consistently delivered within the BA, activity scheduling and activity monitoring [30], were selected for inclusion by most participants.

Apart from providing support for the development of a new computerized therapy program based on BA, this qualitative work also provided vital information regarding the presentation, delivery, and content of the new program to ensure that it meets the needs of the target user.

Although there were individual differences across the 3 participant groups, with novel themes emerging from each (Figure 1), agreements were evident about several issues pertaining to the presentation and delivery of the program, which were used to inform its development. For example, all groups felt that more than one presentation was required to suit the target age range, and it was generally felt that program sessions should last approximately 30-45 minutes with appropriate time in between sessions, and all groups wanted the program to be easily accessible.

However, opinions regarding some of the program content were more complex to reconcile, both in incorporating the views of the 3 participant groups where differences were apparent and also being mindful of the need to include evidence-based components of therapy.

Across all groups, concern was expressed about the inclusion of homework, with this component regarded as less important than others. Despite these concerns, research has demonstrated correlations between homework compliance and clinical improvements when delivered within CBT [34]. Including homework in the new program was therefore important, despite being unpopular. To ensure a balance between incorporating evidence-based components within therapy but being sensitive to the opinions of the target users, both were considered. Although they would generally have preferred homework to be omitted from the new program, the participants stipulated that, if included, it required a different name and had to be minimal. Therefore, in the development of the new program, these stipulations guided the inclusion of a homework component.

Discussions also occurred regarding the involvement of others in the completion of a computerized therapy, particularly parents. As reported elsewhere in a systematic review of young people's experiences of using technology-assisted CBT [17], both positive and negative opinions regarding parental involvement were expressed. Although the community sample would have preferred no involvement from others, they agreed that in instances of risk, parents needed to be notified; the service users were generally more positive and identified how treatment effectiveness could be enhanced if parents were involved. In contrast, health care professionals regarded the involvement of others, both parents and health care professionals, as essential, contending that both needed to take an active role in the new program. Research has highlighted that the need for autonomy among young people can be a barrier to seeking mental health support [35]. Therefore, this clearly needs to be addressed in the development of the new program while considering the important perspectives captured within this study. Researchers argue that interventions need to address the autonomous needs of young people to increase their likelihood of seeking help and suggest helping young people to know how and when to seek support from others [35]. Therefore, in the new program, development focus needed to be placed upon supporting help-seeking while remaining sensitive to young people's need for autonomy.

The needs of young people can be better met if there is knowledge about where they seek help when they are distressed [36]. Despite different experiences, the young-person participants expressed concerns related to accessing support, particularly concerning information disclosure and how they would be treated and perceived as a result. They felt that a computerized therapy would allay some of these concerns while allowing them to access therapy at times when most needed and in their own surroundings where they feel most comfortable. Similar findings have been reported in previous research [37,38], which highlights the benefits of adopting a computerized therapy delivery approach in ameliorating some of the barriers faced by young people in accessing services. However, although the health care professionals in this study were familiar with the availability of computerized therapies, few young people were, as none were directed to any resources in the past. This is represented in Figure 1, which shows that although familiarity with computerized therapies emerged as the main theme for health care professionals, this was not the case for young-person participants who instead discussed the requirement of promoting awareness. This demonstrates the need to sufficiently promote the support available to young people.

Similar to previous research [27,28] and as presented in Figure 1, the differences in the experiences, opinions, and priorities across the participant groups in this work highlight the need to incorporate different stakeholders in the development of new therapies. Although it is clear that the opinions of young people need to be considered within their own care, the views of health care professionals also need acknowledgment, as those who participated in this study provided salient information from a clinical perspective. Furthermore, the opinions of those who do not routinely access services but for whom a treatment may be useful need to be considered in therapy design. All young-person participants were from similar age ranges and ethnic groups and lived within relative geographical proximity. Despite this, clear differences in opinion were evident based on whether they had previously accessed support. As discussed, 65% of young people with low mood or depression do not access care [1] because of factors including stigma [3,4], negative attitudes about help-seeking [4], accessibility [3], and reluctance to engage one-to-one with a therapist [5]. Thus, although it is undoubtedly important to tailor treatments to target users [39], the views of those with limited experience of CAMHS also need to be considered.

Therefore, a systemic approach to the development of new interventions needs to be adopted to ensure that they meet the needs of the target user. Such an approach should incorporate the views of those with treatment experience, who are experts on what works and what does not; those who have not accessed treatment, who are the experts on the barriers and how these may be overcome; and health care professionals, who are experts on the practical application of therapy and its integration into services.

Limitations

Owing to several limitations of this study, the results need to be interpreted with caution. To ensure that rich and trustworthy information was collected as part of this study, 3 different participant groups were recruited. Despite this, several factors may have affected the transferability of the findings. As an opt-in method was used, the young-person participants were more likely to be those motivated to engage and thus might not have expressed views demonstrative of the general age range population. Only one service user participant was male, which, although unsurprising with research suggesting that males are less likely to access mental health services [40], meant that the views of males were underrepresented within this sample. Furthermore, although recruitment occurred within 4 NHS trusts, these were all located in close proximity to each other and were culturally similar, with little ethnic diversity in the overall sample. Therefore, the views expressed might not be typical of health care professionals based within the NHS trusts or young people receiving support from other NHS trusts.

Conclusions

Vital information was collected to inform the development of a new computerized therapy. Although similarities in the opinions of the participant groups were evident, differences highlighted the need to adopt a systemic approach in therapy development. In the context of young-person therapy, the opinions of young people with and without experience in accessing mental health support and health care professionals need to be incorporated.

Acknowledgments

This work was supported by the Economic and Social Research Council (grant number: 1500498). The authors would like to thank the young people and health care professionals who participated in this research.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Young person participant and health care professional interview schedules.
[DOC File , 55 KB - jmir_v23i4e19743_app1.doc]

References

1. Gladstone TG, Marko-Holguin M, Rothberg P, Nidetz J, Diehl A, DeFrino DT, et al. An internet-based adolescent depression preventive intervention: study protocol for a randomized control trial. Trials 2015 May 01;16:203 [FREE Full text] [doi: 10.1186/s13063-015-0705-2] [Medline: 25927539]
2. World Health Organization. Mental Health Action Plan 2013-2020. Geneva, Switzerland: WHO Document Production Services; 2020.
3. Gulliver A, Griffiths KM, Christensen H. Perceived barriers and facilitators to mental health help-seeking in young people: a systematic review. BMC Psychiatry 2010;10:113 [FREE Full text] [doi: 10.1186/1471-244X-10-113] [Medline: 21192795]
4. Rickwood D, Deane FP, Wilson CJ, Ciarrochi J. Young people's help-seeking for mental health problems. Advances in Mental Health 2005 Dec;4(3):218-251. [doi: 10.5172/jamh.4.3.218]
5. Rickwood DJ. Promoting Youth Mental Health through Computer-Mediated Communication. International Journal of Mental Health Promotion 2010 Aug;12(3):32-44. [doi: 10.1080/14623730.2010.9721817]
6. March S, Spence SH, Donovan CL. The efficacy of an internet-based cognitive-behavioral therapy intervention for child anxiety disorders. J Pediatr Psychol 2009 Jun;34(5):474-487 [FREE Full text] [doi: 10.1093/jpepsy/jsn099] [Medline: 18794187]
7. Abeles P, Verduyn C, Robinson A, Smith P, Yule W, Proudfoot J. Computerized CBT for adolescent depression ("Stressbusters") and its initial evaluation through an extended case series. Behav Cogn Psychother 2009 Mar;37(2):151-165. [doi: 10.1017/S1352465808005067] [Medline: 19364416]
8. Wright B, Tindall L, Littlewood E, Adamson J, Allgar V, Bennett S, et al. Computerised cognitive behaviour therapy for depression in adolescents: study protocol for a feasibility randomised controlled trial. BMJ Open 2014;4(10):e006488 [FREE Full text] [doi: 10.1136/bmjopen-2014-006488] [Medline: 25361841]
9. Dimidjian S, Barrera M, Martell C, Muñoz RF, Lewinsohn PM. The origins and current status of behavioral activation treatments for depression. Annu Rev Clin Psychol 2011;7:1-38. [doi: 10.1146/annurev-clinpsy-032210-104535] [Medline: 21275642]

10. Chartier IS, Provencher MD. Behavioural activation for depression: efficacy, effectiveness and dissemination. J Affect Disord 2013 Mar 5;145(3):292-299. [doi: 10.1016/j.jad.2012.07.023] [Medline: 22884236]
11. Ekers D, Richards D, Gilbody S. A meta-analysis of randomized trials of behavioural treatment of depression. Psychol Med 2008 May;38(5):611-623. [doi: 10.1017/S0033291707001614] [Medline: 17903337]
12. Ekers D, Webster L, van Straten SA, Cuijpers P, Richards D, Gilbody S. Behavioural activation for depression; an update of meta-analysis of effectiveness and sub group analysis. PLoS One 2014;9(6):e100100 [FREE Full text] [doi: 10.1371/journal.pone.0100100] [Medline: 24936656]
13. Cuijpers P, van Straten A, Warmerdam L. Behavioral activation treatments of depression: a meta-analysis. Clinical Psychology Review 2007 Apr;27(3):318-326. [doi: 10.1016/j.cpr.2006.11.001]
14. Depression: The treatment and management of depression in adults. National Institute for Health and Care Excellence. URL: https://pubmed.ncbi.nlm.nih.gov/22132433/ [accessed 2021-03-28]
15. Stallard P, Velleman S, Richardson T. Computer Use and Attitudes Towards Computerised Therapy Amongst Young People and Parents Attending Child and Adolescent Mental Health Services. Child Adolesc Ment Health 2010 May;15(2):80-84. [doi: 10.1111/j.1475-3588.2009.00540.x] [Medline: 32847246]
16. Pretorius C, Chambers D, Cowan B, Coyle D. Young People Seeking Help Online for Mental Health: Cross-Sectional Survey Study. JMIR Ment Health 2019 Aug 26;6(8):e13524 [FREE Full text] [doi: 10.2196/13524] [Medline: 31452519]
17. McCashin D, Coyle D, O'Reilly G. Qualitative Synthesis of Young People's Experiences With Technology-Assisted Cognitive Behavioral Therapy: Systematic Review. J Med Internet Res 2019 Nov 12;21(11):e13540 [FREE Full text] [doi: 10.2196/13540] [Medline: 31714251]
18. Fleming TM, Cheek C, Merry SN, Thabrew H, Bridgman H, Stasiak K, et al. Juegos serios para el tratamiento o la prevención de la depresión: una revisión sistemática. RPPC 2015 Jan 15;19(3):227. [doi: 10.5944/rppc.vol.19.num.3.2014.13904]
19. Jolly K, Bradley F, Sharp S, Smith H, Thompson S, Kinmonth AL, et al. Randomised controlled trial of follow up care in general practice of patients with myocardial infarction and angina: final results of the Southampton heart integrated care project (SHIP). The SHIP Collaborative Group. BMJ 1999 Mar 13;318(7185):706-711 [FREE Full text] [doi: 10.1136/bmj.318.7185.706] [Medline: 10074017]
20. Wiles R. Patient's perceptions of their heart attack and recovery: the influence of epidemiological "evidence" and personal experience. Social Science and Medicine. 1998. URL: https://doi.org/10.1016/S0277-9536(97)10140-X [accessed 2020-07-17]
21. Scholten H, Granic I. Use of the Principles of Design Thinking to Address Limitations of Digital Mental Health Interventions for Youth: Viewpoint. J Med Internet Res 2019 Jan 14;21(1):e11528 [FREE Full text] [doi: 10.2196/11528] [Medline: 31344671]
22. Orlowski S, Matthews B, Bidargaddi N, Jones G, Lawn S, Venning A, et al. Mental Health Technologies: Designing With Consumers. JMIR Hum Factors 2016 Jan 28;3(1):e4 [FREE Full text] [doi: 10.2196/humanfactors.4336] [Medline: 27026210]
23. Sauter FM, Heyne D, Michiel WP. Cognitive behavior therapy for anxious adolescents: developmental influences on treatment design and delivery. Clin Child Fam Psychol Rev 2009 Dec;12(4):310-335 [FREE Full text] [doi: 10.1007/s10567-009-0058-z] [Medline: 19568935]
24. Hudson JL, Kendall PC, Coles ME, Robin JA, Webb A. The other side of the coin: using intervention research in child anxiety disorders to inform developmental psychopathology. Dev Psychopathol 2002;14(4):819-841. [doi: 10.1017/s095457940200408x] [Medline: 12549705]
25. Holmbeck G, O'Mahar K, Abad M, Colder C, Updegrove A. Cognitive-behavior therapy with adolescents: Guides from developmental psychology. In P. C. Kendall (Ed.), Child and adolescent therapy: Cognitive-behavioral procedures. New York: Guilford Press; 2006:419-464.
26. Irwin LG, Johnson J. Interviewing young children: explicating our practices and dilemmas. Qual Health Res 2005 Jul;15(6):821-831. [doi: 10.1177/1049732304273862] [Medline: 15961878]
27. Fox A, Berrick JD. A Response to No One Ever Asked Us: A Review of Children's Experiences in Out-of-Home Care. Child Adolesc Soc Work J 2006 Dec 13;24(1):23-51. [doi: 10.1007/s10560-006-0057-6]
28. Forsner M, Jansson L, Sørlie V. The experience of being ill as narrated by hospitalized children aged 7-10 years with short-term illness. J Child Health Care 2005 Jun;9(2):153-165. [doi: 10.1177/1367493505051406] [Medline: 15961369]
29. Danermark B, Ekström M, Karlsson J. C. Explaining society: Critical realism in the social sciences, London: Routledge; 2019.
30. Kanter JW, Manos RC, Bowe WM, Baruch DE, Busch AM, Rusch LC. What is behavioral activation? A review of the empirical literature. Clin Psychol Rev 2010 Aug;30(6):608-620. [doi: 10.1016/j.cpr.2010.04.001] [Medline: 20677369]
31. Tindall L, Mikocka-Walus A, McMillan D, Wright B, Hewitt C, Gascoyne S. Is behavioural activation effective in the treatment of depression in young people? A systematic review and meta-analysis. Psychol Psychother 2017 Dec;90(4):770-796 [FREE Full text] [doi: 10.1111/papt.12121] [Medline: 28299896]
32. Denzin N. The research act: A theoretical introduction to sociological methods. New York, NY: Aldine Transaction; 2009.
33. Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Research in Psychology 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
34. Simons AD, Marti CN, Rohde P, Lewis CC, Curry J, March J. Does Homework "Matter" in Cognitive Behavioral Therapy for Adolescent Depression? J Cogn Psychother 2012 Nov 01;26(4):390-404. [doi: 10.1891/0889-8391.26.4.390]

35. Wilson CJ, Deane FP. Brief report: Need for autonomy and other perceived barriers relating to adolescents' intentions to seek professional mental health care. J Adolesc 2012 Mar;35(1):233-237. [doi: 10.1016/j.adolescence.2010.06.011] [Medline: 20650512]
36. Krasnova TN, Samokhodskaya LM, Ivanitsky LV, Korogodina AD, Borisov EN, Nikiforova NV, et al. [Impact of interleukin-10 and interleukin-28 gene polymorphisms on the development and course of lupus nephritis]. Ter Arkh 2015 Dec;87(6):40-44. [doi: 10.17116/terarkh201587640-44] [Medline: 26281194]
37. Carlbring P, Andersson G. Internet and psychological treatment. How well can they be combined? Computers in Human Behavior 2006 May;22(3):545-553. [doi: 10.1016/j.chb.2004.10.009]
38. Sheffield J, Fiorenza E, Sofronoff K. Adolescents' Willingness to Seek Psychological Help: Promoting and Preventing Factors. Journal of Youth and Adolescence 2004 Dec;33(6):495-507 [FREE Full text] [doi: 10.1023/B:JOYO.0000048064.31128.c6]
39. Bradley F, Wiles R, Kinmonth AL, Mant D, Gantley M. Development and evaluation of complex interventions in health services research: case study of the Southampton heart integrated care project (SHIP). The SHIP Collaborative Group. BMJ 1999 Mar 13;318(7185):711-715 [FREE Full text] [Medline: 10074018]
40. Rickwood DJ, Mazzer KR, Telford NR. Social influences on seeking help from mental health services, in-person and online, during adolescence and young adulthood. BMC Psychiatry 2015;15:40 [FREE Full text] [doi: 10.1186/s12888-015-0429-6] [Medline: 25886609]

Abbreviations

BA: behavioral activation
CAMHS: child and adolescent mental health services
CBT: cognitive behavioral therapy
GP: general practice
NHS: National Health Service

Edited by G Eysenbach; submitted 30.04.20; peer-reviewed by M Eder; comments to author 13.07.20; revised version received 21.07.20; accepted 11.11.20; published 13.04.21.

Please cite as:
Tindall L, Toner P, Mikocka-Walus A, Wright B
Perceptions of and Opinions on a Computerized Behavioral Activation Program for the Treatment of Depression in Young People: Thematic Analysis
J Med Internet Res 2021;23(4):e19743
URL: https://www.jmir.org/2021/4/e19743
doi:10.2196/19743
PMID:33847594

Original Paper

Predictors of Parental Barriers to Reduce Excessive Child Screen Time Among Parents of Under-Five Children in Selangor, Malaysia: Cross-sectional Study

Elliza Mansor[1,2], MPH; Norliza Ahmad[1], PhD; Diana Raj[1,2], MPH; Nor Afiah Mohd Zulkefli[1], PhD; Zalilah Mohd Shariff[3], PhD

[1]Department of Community Health, Faculty of Medicine and Health Sciences, Serdang, Malaysia

[2]Ministry of Health Malaysia, Putrajaya, Malaysia

[3]Department of Nutrition and Dietetics, Faculty of Medicine and Health Sciences, Universiti Putra Malaysia, Serdang, Malaysia

Corresponding Author:
Norliza Ahmad, PhD
Department of Community Health
Faculty of Medicine and Health Sciences
Universiti Putra Malaysia
Serdang, 43400
Malaysia
Phone: 60 019 271 0577
Email: lizaahmad@upm.edu.my

Abstract

Background: Globally, there is an increasing prevalence of excessive screen time exposure among young children, including in Malaysia. Parents are advised to limit this exposure, but there are barriers for many of them to follow this recommendation. To date, there is a lack of research on the factors that cause these parental barriers.

Objective: This study aimed to determine the parental barrier toward the reduction of excessive child screen time and its predictors among parents of children aged younger than 5 years in the Petaling District, Selangor, Malaysia.

Methods: A cross-sectional study was conducted from April 2019 to June 2020 among 789 parent-child dyads attending child health clinics in the Petaling District. Validated self-administered questionnaires were used to capture information on sociodemographic, parental, child-related, and environmental factors and parental barriers. Stratified sampling with probability proportionate to size was employed. Data were analyzed using SPSS Statistics version 25 (IBM Corp). Descriptive analysis and bivariable analysis were performed before multiple linear regression was used to identify predictors of parental barriers.

Results: The overall mean score of parental barriers was 3.51 (SD 0.83), indicating that the average numbers of barriers experienced by parents were more than 3. The multivariable analysis showed that the predictors of parental barriers included monthly household income (adjusted β=–.03, 95% CI –0.05 to –0.02), parents who worked in public sectors (adjusted β=.18, 95% CI 0.06 to 0.29), positive parental attitude on screens (adjusted β=.68, 95% CI 0.58 to 0.79), low parent self-efficacy to influence child's physical activity (adjusted β=–.32, 95% CI –0.43 to –0.20), and child screen time (adjusted β=.04, 95% CI 0.02 to 0.06).

Conclusions: The strongest predictor of parental barriers to reduce excessive child screen time was the positive parental attitude on screen time which could contribute to their abilities to limit child screen time. Thus, future intervention strategies should aim to foster correct parental attitudes toward screen time activities among young children.

(J Med Internet Res 2021;23(4):e25219) doi:10.2196/25219

KEYWORDS

child; self-efficacy; screen time; Malaysia; parent-child relations; public sector; children; screen; parental

Introduction

Screen time refers to the total amount of time a person spends passively on any screen-based technology such as smartphone, tablet, video game, computer, television, or any wearable device [1,2]. The World Health Organization recommends that children aged younger than 2 years should have no access to screen time whereas for those aged 2 to 5 years, sedentary screen time should not exceed 1 hour per day, and less is better [1]. In recent years, the increasing prevalence of excessive screen time among children has become a public health issue of global importance [1,2]. Research studies showed a high percentage of children and youth who exceeded the recommended screen time worldwide [3-5]. For example, a study in the United States reported that on average, children aged 8 months to 8 years were exposed up to 4 hours of background television in a day [6]. Likewise, a study in the United Kingdom reported a high proportion of children aged 6 to 36 months were using screen media, and their screen time increased significantly with age [7]. In addition, studies from two developed countries in Asia (ie, Singapore and Japan) also reported that children aged younger than 5 years were exposed to both television and other devices an average of 4 hours per day [8,9]. Locally, two Malaysian surveys found that 27% to 52% of children aged younger than 6 years were exposed to television and other devices (computers, tablets, and smartphones) for more than 2 hours per day [10,11]. Furthermore, about 68% of older Malaysian children (aged 7 to 12 years) spent an average of 3 hours on screen time daily [12]. Even more worrying is the fact that as high as 74% of Malaysian children aged younger than 2 years have been exposed to screen time, in contrast to the World Health Organization recommendation [11].

Multiple studies have highlighted the undesirable effects of excessive screen time on the developmental, psychosocial, and physical health of children. Among these negative effects were visual symptoms after prolonged screen use [13]; speech delay; and unfavorable development of physical, cognitive, and academic abilities of children [14-18]. Moreover, studies also report increased body fat and obesity as a result of increased food consumption and exposure to unhealthy food advertisements during screen viewing [19]. In view of these adverse effects, parents were advised to limit the screen time of their children based on the appropriate duration recommended for the child's age [1,2].

During childhood and adolescence periods, parents play a crucial role in preventing the development of sedentary behavior [20]. Several studies have demonstrated the role of parental influences on a child's screen time behavior [21,22]. However, it was also reported that parents could be facing various barriers in the efforts to reduce excessive child screen time [23-26]. Among these barriers were the lack of affordable alternative activities and poor accessibility to them secondary to weather conditions and transportation issues [23,24,26]. Additionally, parental issues such as parental fatigue, time constraints, and a desire to have time away from children to complete personal tasks also hinder the participation of children in physical activities [24,25,27]. Available studies suggested it is vital to identify such barriers experienced by parents for the establishment of effective intervention in promoting healthy behaviors among children [28]. However, a limited number of studies report on factors associated with parental barriers to reduce a child's screen time in Malaysia. This study aimed to investigate parental barriers toward the reduction of a child's screen time and relevant predictors from the aspect of sociodemographic, parental, child-related, and environmental factors among parents of children aged younger than 5 years who attended child health clinics in the Petaling District, Selangor, Malaysia.

Methods

Study Design and Location

A cross-sectional study was conducted in 9 health clinics located in the Petaling District from April 2019 to June 2020. This study was conducted in health clinics because all government health clinics in Malaysia offer free child health services such as health screening, developmental and nutritional status assessment, immunizations, and dental health services that are accessible for all local citizens [29,30]. Due to this fact, Malaysian parents, especially those of middle and lower income groups, predominantly used these health clinics. Petaling District was selected as the study site was because it is the most populated district in Selangor, Malaysia, with a high number of young children [31]. Petaling District is urbanized with one of the highest median household incomes [32].

Study Participants

The study population was Malaysian parents of children aged younger than 5 years attending the child health clinics in the Petaling District. Parents who were illiterate or whose children had physical or mental disabilities or chronic diseases such as heart disease and asthma were excluded. These exclusion criteria ensured that we recruited parents with healthy children whose access to screen media was not influenced by their health condition.

Sample Size and Sampling Method

The sample size was calculated using the correlation coefficient formula based on the inputs from a previous study of a similar topic [33]. Based on a 95% confidence level, 80% power, and 20% nonresponse rate, a minimum sample size of 737 participants were required. To achieve the sample size, stratified sampling with probability proportionate to sample size was adopted. The sample size for the parent-child dyad in each clinic was made to be proportionate to the average number of children attending the 9 child health clinics in the Petaling District in 2017, totaling approximately 7270. We used the attendance record from each clinic as the sampling frame, and each respondent was selected using a systematic sampling method whereby the k interval obtained was 14. The first parent-child dyad (a parent attending the clinic with a child aged younger than 5 years) was selected using a random number generator that gave all respondents an equal chance of being chosen. By repeating the same method in each clinic, a total of 789 participants were recruited for the study.

Study Instruments

The study employed a validated study instrument adapted from various studies with a good to excellent level of 1-week test-retest reliability (intraclass correlation coefficient 0.64-0.98). The original English version of the questionnaire consisted of 8 sections and was translated into the Malay language by experts from the Editing and Translation Service, Centre for the Advancement of Language Competence, at University Putra Malaysia. See Multimedia Appendix 1 for questionnaire.

Sociodemographic Factors

Information regarding age, gender, ethnicity, education level, household income, employment status, and marital status of respondents was collected. The household income in Malaysia is categorized into 3 income groups based on the median monthly household income reported by Department of Statistics Malaysia: namely, the upper 20% (T20), the middle 40% (M40), and the lower 40% (B40) [32]. These income group definitions are not fixed depending on the Malaysia's gross domestic product.

Parental Factors

Parental attitude toward screen time was evaluated based on parent opinions of statements related to screen time benefits. The scores were then categorized [34,35], and parents who scored above 32 were rated as having a positive attitude. Next, parent perceptions were evaluated regarding the influence of screen time on their child's physical, cognitive, and social well-being [36]. Higher scores represent a greater perception of positive influence of screen time toward the specific aspect of well-being. In addition, the parent's self-efficacy in influencing the child's physical activity reflected the confidence level of the parent in situations related to a child's physical activity. This variable was classified into 3 groups based on the percentile of the score [26].

Last, parenting style was categorized into 4 groups according to the median split in the Baumrind classification of parent involvement and strictness toward their child [37]. Parental restrictive practices on screen time were assessed based on their sedentary-related restrictive practices in which higher scores reflected more restrictive practices [38]. Parental screen time was defined as the average number of hours a parent spent on television, computers, cell phones, and other electronic devices for leisure purposes excluding work or school use during weekdays and weekends.

Child-Related Factors

Data on sociodemographic characteristics (age, gender, and numbers of siblings) of the child were obtained from the parents. The child's screen time was taken as the average time spent by the child on each media device per weekday and weekend [8]. Child care settings were categorized into home care by parents, home care by other than parents, and childcare center [39].

Environmental Factors

Environmental factors refer to the conditions of both the household and neighborhood. Questions on the household environment captured information on the number and type of screen devices in the home including those in the child's bedroom,, and outdoor play equipment at home. Questions on the neighborhood environment asked about the availability of physical activity facilities in the public areas and perceived safety with regard to the likelihood of criminal activities [40].

Parental Barriers to Reduce Excessive Child Screen Time

The outcome of this study was assessed based on 6 statements [34]: (1) there is pressure from society to purchase and use media-related equipment, (2) my neighborhood is not safe for my child to play outdoors, (3) poor weather limits my child's opportunities to go outside, (4) I need a coping tool to meet the demands of a busy day at work or raising multiple children, (5) I need time to do household chores, and (6) my child really enjoys screen time activities. All items were rated on a 5-point Likert scale ranging from strongly disagree to strongly agree. The mean scores of the responses represented an overall score with regard to parental barriers. The maximum score was 6, with higher scores indicated that parents were facing more barriers.

Data Collection

A validated self-administered questionnaire was distributed to all respondents. Selected parents were approached during their visits to the child health clinics in the Petaling District.

Statistical Analysis

Data entry and analysis were performed using SPSS Statistics version 25.0 (IBM Corp). Prior to the analysis, data cleaning was done. Skewness, kurtosis, and histogram were part of the normality testing conducted on continuous data. Following that, descriptive analysis was performed for all variables. Categorical data were described in frequency and percentage, whereas continuous data were shown in mean and standard deviation or median and interquartile range depending on normality test results.

In the next step, bivariable analysis was performed using simple linear regression to determine the association between parental barriers in reducing excessive child screen time and all independent variables. Dummy variables were created for all categorical variables prior to bivariable analysis. Based on the findings from simple linear regression, any variables with $P<.25$ were included in the subsequent multiple linear regression analysis to identify the predictors of parental barriers in reducing excessive child screen time [41]. All variables in the model were tested for interaction and multicollinearity to fulfill the assumption for multiple linear regression analysis. Final findings were presented as adjusted β with 95% confidence intervals with the level of significance set at $P<.05$.

Ethics

Ethical approval was obtained from the medical research and ethics committee of the Ministry of Health Malaysia (NMRR-19-41-45681 [Investigator-Initiated Research]). All participants provided written informed consent. They were notified that their information would be kept confidential and they could withdraw from the study at any time.

Results

Sociodemographic Characteristics of Participants

The majority of the 789 participants were women (662/789, 83.9%), Malays (684/789, 86.7%), and married (781/789, 99.0%). The mean age of the respondents was 31.6 (SD 4.81) years, and 70.6% (557/789) had tertiary education. The median income of parents in this study was MYR 5000 (interquartile range 3000); half (403/789, 51.1%) of the parents belonged to the B40 income group based on the categorization of the Malaysia household income groups [32]. Among the 77.7% (613/789) of parents who were employed, the proportion of parents working in the private and public sectors was similar (Table 1).

Table 1. Distribution of respondents according to sociodemographic factors (n=789).

Characteristic	Value, n (%)
Parent age in years	
<30	283 (35.9)
≥30	506 (64.1)
Gender	
Female/mother	662 (83.9)
Male/father	127 (16.1)
Ethnicity	
Malay	684 (86.7)
Chinese	31 (3.9)
Indian	36 (4.6)
Other	38 (4.8)
Parent education level	
Primary school and below	14 (1.8)
Secondary school	218 (27.6)
Preuniversity and higher	557 (70.6)
Monthly household income (MYR[a])	
<4360 (B40[b])	403 (51.1)
4360-9619 (M40[c])	310 (39.3)
>9620 (T20[d])	76 (9.6)
Employment status	
Public sector	218 (27.6)
Private sector	221 (28.0)
Self-employed	174 (22.1)
Unemployed/housewife	176 (22.3)
Marital status	
Married	781 (99.0)
Divorced/widowed/separated	8 (1.0)

[a]MYR: Malaysian ringgit. 1 MYR = 0.24 US $.

[b]B40: lower 40%, based on the median monthly household income reported by Department of Statistics Malaysia [32].

[c]M40: middle 40%, based on the median monthly household income reported by Department of Statistics Malaysia [32].

[a]T20: upper 20%, based on the median monthly household income reported by Department of Statistics Malaysia [32].

Parental Barriers to Reduce Excessive Child Screen Time

The mean score of each barrier reported by parents toward the reduction of excessive screen time as outlined in Table 2 was 3.51 (SD 0.83), and a higher mean score represented a more commonly encountered barrier among the parents. Most parents attributed the need to spend time on household chores, unpredictable weather, and lack of a safe neighborhood that restricted outdoor play as barriers for them to limit screen time exposure in their children. Table 2 shows the distribution of the

different barriers experienced by parents in the efforts to reduce screen time among children.

Table 2. Distribution of parental barriers to reduce excessive child screen time (n=789).

Barrier statement	Strongly disagree, n (%)	Disagree, n (%)	Somewhat agree, n (%)	Agree, n (%)	Strongly agree, n (%)	Mean (SD)
There is pressure from society to purchase and use media-related equipment (such as mobile devices, computers, and DVD players).	112 (14.2)	251 (31.8)	199 (25.2)	171 (21.7)	56 (7.1)	2.76 (1.15)
My neighborhood is not safe for my child to play outside.	75 (9.5)	213 (27.0)	234 (29.7)	197 (25.0)	70 (8.9)	2.97 (1.12)
Unpredictable weather (such as hot, cold, and rain) limits my child's chances of playing outside.	53 (6.7)	190 (24.1)	256 (32.4)	229 (29.0)	61 (7.7)	3.07 (1.05)
I need a coping tool to meet the demand of a busy day at work or raising multiple children.	57 (7.2)	226 (28.6)	287 (36.4)	175 (22.2)	44 (5.6)	2.90 (1.01)
I need time to do household chores (such as washing and cooking).	48 (6.1)	163 (20.7)	307 (38.9)	221 (28.0)	50 (6.3)	3.08 (0.99)
My child really enjoys screen time activities.	89 (11.3)	225 (28.5)	280 (35.5)	165 (20.9)	30 (3.8)	2.77 (1.02)

Bivariable Analysis of Parental Barriers to Reduce Excessive Child Screen Time

Table 3 shows all the factors studied that were significantly associated with parental barriers including sociodemographic, parental, child-related, and environmental factors. For sociodemographic factors, parents of Malay (β=–.29, 95% CI –0.56 to –0.01; *P*=.04), Chinese (β=–.49, 95% CI –0.88 to –0.09; *P*=.02), and Indian (β=–.40, 95% CI –0.78 to –0.02; *P*=.04) ethnicity; combined monthly household income (β=–.04, 95% CI –0.06 to –0.02, *P*<.001), and employment status in terms of parents who worked in the public sector (β=.16, 95% CI 0.00 to 0.32, *P*=.045) or were unemployed/housewife (β=.18, 95% CI 0.01 to 0.34, *P*=.04) were among the factors that were significantly associated with parental barriers.

Parental factors found to be positively associated with parental barriers included parental attitude on child screen time (β=.78, 95% CI 0.66 to 0.89; *P*<.001), perception on the influence of screen time on cognitive well-being (β=.04, 95% CI 0.02 to 0.06; *P*=.001) and social well-being (β=.05, 95% CI 0.02 to 0.08; *P*=.001), as well as parental screen time (β=.18, 95% CI 0.06 to 0.31; *P*=.003). On the contrary, a neglectful type of parenting style (β=.15, 95% CI 0.01 to 0.28; *P*=.03) and low self-efficacy to influence child's physical activity (β=–.43 95% CI –0.58 to –0.28; *P*<.001) were negatively associated with parental barriers.

In addition, 3 of the 5 child-related factors were significantly associated with parental barriers in reducing a child's screen time. The significant factors included number of children (β=.18, 95% CI 0.06 to 0.30; *P*=.003), child's age (β=.01, 95% CI 0.00 to 0.01; *P*=.001), and child's screen time (β=.07, 95% CI 0.05 to 0.09; *P*<.001).

Environmental factors significantly associated with parental barriers included the presence of screen devices in child's bedroom (β=–.19, 95% CI –0.35 to –0.03; *P*=.02) and parental perceived safety related to crime (β=.27, 95% CI 0.21 to 0.34; *P*<.001).

Table 3. Association between sociodemographic, parental, child-related, and environmental factors with parental barriers (n=789).

Variable	β	SE	*P* value	95% CI
Sociodemographic factors				
Age in years	0.01	0.01	.31	–0.01 to 0.01
Gender				
Female	Reference	—[a]	—	—
Male	0.05	0.08	.53	–0.11 to 0.21
Ethnicity				
Malay	–0.29	0.14	.04[b]	–0.56 to –0.01
Chinese	–0.49	0.20	.02[b]	–0.88 to –0.09
Indian	–0.40	0.19	.04[b]	–0.78 to –0.02
Other	Reference	—	—	—
Educational level				
Primary school	0.02	0.23	.95	–0.44 to 0.47
Secondary school	Reference	—	—	–0.24 to 0.02
Preuniversity	–0.11	0.07	.11	–0.06 to –0.02
Monthly household income (MYR[c])	–0.04	0	<.001[b]	–0.06 to –0.02
Employment status				
Private sector	Reference	—	—	—
Public sector	0.16	0.08	.045[b]	0.00 to 0.32
Self-employed	0.03	0.08	.76	–0.14 to 0.19
Unemployed/housewife	0.18	0.08	.04[b]	0.01 to 0.34
Marital status				
Married	Reference	—	—	—
Divorced/widowed/separated	–0.01	0.30	.97	–0.59 to 0.57
Parental factors				
Attitude on screen time				
Negative	Reference	—	—	—
Positive	0.78	0.06	<.001[b]	0.66 to 0.89
Parenting style				
Authoritative	Reference	—	—	—
Authoritarian	0.05	0.12	.66	–0.16 to 0.26
Indulgent	0.01	0.10	.96	–0.18 to 0.19
Neglectful	0.15	0.07	.03[b]	0.01 to 0.28
Perception on the influence of screen time on child's				
Physical well-being	0.03	0.01	.05	0.00 to 0.05
Cognitive well-being	0.04	0.01	.001[b]	0.02 to 0.06
Social well-being	0.05	0.02	.001[b]	0.02 to 0.08
Restrictive practices on screen time	–0.01	0.01	.10	–0.03 to 0.00
Self-efficacy to influence the child's physical activity				
High	Reference	—	—	—
Moderate	–0.00	0.07	.98	–0.14 to 0.13

Variable	β	SE	*P* value	95% CI
Low	–0.43	0.08	<.001[b]	–0.59 to –0.28
Parental screen time in hours				
≤2	Reference	—	—	—
>2	0.18	0.06	.003[b]	0.06 to 0.31
Child-related factors				
Number of children				
Single child	Reference	—	—	—
2 or more children	0.18	0.06	.003[b]	0.06 to 0.30
Child's age in months	0.01	0	.001[b]	0.00 to 0.01
Child's gender				
Male	Reference	—	—	—
Female	0.08	0.06	.18	–0.04 to 0.20
Child's screen time in hours	0.07	0.01	<.001[b]	0.05 to 0.09
Child care setting				
Home care by parents	Reference	—	—	—
Home care by others	–0.05	0.08	.53	–0.20 to 0.10
Childcare center	0.01	0.07	.87	–0.13 to 0.15
Environmental factors				
Additional piece of screen device at home				
<3 media devices	Reference	—	—	—
≥3 media devices	0	0.12	.97	–0.23 to 0.24
Screen device in child bedroom				
Yes	Reference	—	—	—
No	–0.19	0.08	.02[b]	–0.35 to –0.03
Outdoor play equipment at home				
Yes	Reference	—	—	—
No	0.04	0.06	.52	–0.08 to 0.16
Availability of public physical activity facilities				
Yes	Reference	—	—	—
No	0.07	0.06	.29	–0.06 to 0.18
Perceived safety related to crime	0.27	0.03	<.001	0.21 to 0.34

[a]Not applicable.

[b]Significant at $P<.05$.

[c]MYR: Malaysian ringgit. 1 MYR = 0.24 US $.

Predictors of Parental Barriers to Reduce Excessive Child Screen Time

A total of 16 factors with $P<.25$ in the simple linear regression were included in the preliminary model. One of the factors, parental perception of safety related to crime, was removed from the model due to its interaction with 2 other factors (public sector employment and monthly household income). The remaining 15 variables showed no multicollinearity and interaction. Hence, all assumptions for multiple linear regression were fulfilled. Forward variable selection method was applied in the final model. The model fits reasonably well with adjusted $R^2=.26$ and F test<0.001.

Table 4 shows that among parents of children aged younger than 5 years, those with low monthly household income encountered more barriers in the efforts to reduce their child's screen time (adjusted β=–.03, 95% CI –0.05 to –0.02). Furthermore, parents who worked in the public sector were more likely to have a higher parental barrier score when

compared with those in the private sector (adjusted β=.18, 95% CI 0.06 to –0.29). Moreover, parents with a positive attitude toward child screen time recorded a higher score of parental barriers than those with a negative attitude (adjusted β=.68, 95% CI 0.58 to 0.79). Parents with low self-efficacy to influence their child's physical activity also displayed a higher parental barrier score as compared with their counterparts (adjusted β=–.32, 95% CI –0.43 to –0.20). Last, parents whose children spent more hours on screen were expected to face more barriers in the efforts to reduce excessive child screen time (adjusted β=.04, 95% CI 0.02 to 0.06).

Table 4. Predictors of parental barriers to reduce excessive child screen time (n=789).

Variable	Adjusted β	SE	*t* test	*P* value[a]	95% CI
Monthly household income (MYR[b])	–0.03	0	–3.83	<.001	–0.05 to –0.02
Employment status					
Private sector	Reference				
Public sector	0.18	0.06	3.05	.002	0.06 to 0.29
Attitude on screen time					
Negative	Reference				
Positive	0.68	0.06	12.37	<.001	0.58 to 0.79
Self-efficacy to influence a child's physical activity					
High	Reference				
Low	–0.32	0.06	–5.45	<.001	–0.43 to –0.20
Child's screen time	0.04	0.01	4.40	<.001	0.02 to 0.06

[a]Significant at $P<.05$.

[b]MYR: Malaysian ringgit. 1 MYR = 0.24 US $.

Discussion

Principal Findings

The majority of parents who participated in this study were Malay, females, and married. More than half of them were aged older than 30 years. This finding highlighted the fact that mothers frequently dealt with all matters related to a child's health, including routine clinic appointments [42]. Moreover, a profile of the respondents also reflected most Malaysian adults, who married in their mid to late 20s [43] and would only bear a child after that age. The high numbers of Malay respondents mirrored the attendees of primary health care facilities in Malaysia, who were mostly Malays with less than one-third being Chinese and Indians [44]. Even though most parents in this study attained preuniversity and higher levels of education, the majority of them belonged to the B40 or M40 group. Furthermore, two-thirds of parents were employed, and the number of parents who worked in the private and public sectors was nearly equivalent. The remaining 22% was self-employed. This finding showed that adults living in urban areas such as the Petaling District have access to various employment sectors in Malaysia.

To the best of our knowledge, no previous studies in Malaysia have reported the barriers experienced by parents in limiting screen time among children aged younger than 5 years. From the analysis, most of the parents experienced an average of more than 3 barriers. This differed significantly from the findings of Jarvis and colleagues [45], whereby parents experienced only one type of barrier. The differences could be due to the different types of barriers presented to the respondents in the studies. Jarvis et al [45] explored barriers that were self-reported by parents through open-ended questions, whereas specific barriers were presented in a self-administered questionnaire to parents in this study. Common barriers encountered by parents in this study (ie, time needed for household chores, weather factor, and coping tools to meet the demand of a busy day) were consistent with previous research findings [23,24,27]. We have found it is a common occurrence among current generations of modern parents to provide screen devices to their young ones as an easy way out [27,46].

In this study, the strongest predictor for parental barriers was the parents' positive attitude on child screen time. A previous study showed that parents' positive attitude toward screen behavior could contribute to internal conflict and thus reduce their abilities to limit child screen time [47]. A similar finding was noted in this study. When parents displayed supportive behavior toward screen time and perceived screen time to be beneficial to their child or themselves, it became a barrier for them to place the necessary restriction on screen time activity. This is not surprising because parents' attitudes toward technology use cast a big influence on how they perceive and value screen time at home [48]. Therefore, our study findings suggested that parents with a more positive attitude toward screen behavior could be subconsciously establishing a home environment that facilitated screen-based activities for their children. Subsequently, this might become a routine habit, leading to early overexposure of children to screen time [5,49]. Eventually, parents would face an uphill battle with the worsening level of barriers toward the reduction of excessive screen time in their children [23].

Furthermore, our study reported that parents with low self-efficacy to influence child's physical activity encountered more barriers to reduce child screen time when compared with parents with high self-efficacy, as highlighted in a previous study [50]. Parents with low confidence to motivate children in engaging in physical activity might also be less likely to initiate screen time reduction in children. In other words, it would not be convincing for the children to follow the restriction if the parents failed to be the role models in supporting healthy behavior among the children [51].

Apart from that, monthly household income was also a predictor of parental barriers in our study. Parents were likely to have an 0.03 increase in the score of the parental barrier for every 1000 MYR decrement in monthly household income. Lower family income could be linked to certain cost issues and, subsequently, restrict access to healthier choices and substitute activities for screen time, thus constituting a common barrier to support healthy behavior [23,25]. In other words, this could explain why lower income families encounter more barriers: they lack the resources to substitute screen-based activities with other activities as compared with higher income families that can afford to provide alternative activities and facilities for their children to substitute screen time [52,53].

Besides family income, parents' employment status was another determinant of parental barriers toward the support of active lifestyles in children, based on a systematic review [23]. Our findings were similar in which working parents encountered more barriers to reduce excessive screen time. However, the difference in the challenges encountered by parents working in the public and private sectors needs to be explored further. A recent study showed that workers in the Malaysian public sector frequently experienced work-life conflict as they had difficulty in coping with the pressures of integrating their work and personal lives as a result of the long and inconsistent working hours [54]. This could in turn lead to parenting stress that acts as a potential parental barrier in the reduction of screen time exposure in children [45,55].

Last, previous studies have reported that children's screen times rose substantially when more barriers were reported by parents [26,34]. This was echoed by our results that demonstrated child screen time as a predictor of parental barriers in the reduction of screen time. The potential reason behind this could be the addictive nature of screen behavior that was integrated into every day routine, thus making it extremely difficult to minimize children's time spent on screen [5,56].

Limitations and Strengths

To the best of our knowledge, this is among the first quantitative studies in the local setting that tried to determine parental barriers toward the reduction of screen time in children. The data were collected with a self-administered validated questionnaire using the probability sampling method and examined the association between a dependent variable with many factors. However, this study is not without limitations. Potential recall bias could have occurred during data collection as many factors were only evaluated from the perspective of a single respondent. Furthermore, the samples were only recruited among parents attending the child health clinics in one district. Thus, the study findings should be interpreted with caution in view of the limited generalizability to other populations.

Future Direction

To overcome barriers encountered by parents, it is essential to enhance parental awareness by providing valuable information through a well-designed health education program. These efforts can help parents overcome the challenges and difficulties of screen time restriction. Health education should consist of information on the recommended screen time behavior for young children. Apart from that, the program should also aim to establish the correct parental attitude on the children's screen time behavior and provide the relevant skills to improve parental self-efficacy in supporting screen time reduction in children. Importantly, the health education program should target families with low income and parents working in the public sector. Future studies should consider intervention research that pinpoints effective strategies for targeted parental barriers.

Conclusion

In summary, this study identified 5 predictors of parental barriers to reducing excessive screen time among children, the strongest one being the positive parental attitude toward child screen time. The majority of parents with children aged younger than 5 years experienced more than 3 barriers in their efforts to reduce screen time.

Acknowledgments

We would like to thank the Director General of Health Malaysia for his permission to publish this article. We also want to acknowledge the Petaling District Health Office and all the participating health clinics under them as well as parents who participated in this study for their extraordinary cooperation and support especially during data collection. Our research has been funded by grant GP-IPM/2018/9640700 from Inisiatif Putra Muda, University Putra Malaysia. However, the execution of this research is the sole responsibility of the authors, and the content of this manuscript does not reflect the official views of the funder.

Authors' Contributions

EM was the principal researcher responsible for the study design and performing the data analysis. DR contributed to the design of this study and the collection of data. NA, NAMZ, and ZMS supervised the execution of this study and contributed to the study design. NA guided, advised, and supervised the analyses and edited the manuscript. All authors were involved in writing and authorized the submitted and published version of this manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Questionnaire.
[DOCX File , 48 KB - jmir_v23i4e25219_app1.docx]

References

1. Guidelines on physical activity, sedentary behaviour and sleep for children under 5 years of age: summary. Geneva: World Health Organization; 2019. URL: https://apps.who.int/iris/rest/bitstreams/1213838/retrieve [accessed 2020-04-28]
2. Canadian Paediatric Society, Digital Health Task Force. Screen time and young children: promoting health and development in a digital world. Paediatr Child Health 2017 Nov;22(8):461-477 [FREE Full text] [doi: 10.1093/pch/pxx123] [Medline: 29601064]
3. Tremblay MS, Barnes JD, González SA, Katzmarzyk PT, Onywera VO, Reilly JJ, Global Matrix 2.0 Research Team. Global Matrix 2.0: report card grades on the physical activity of children and youth comparing 38 countries. J Phys Act Health 2016 Dec;13(11 Suppl 2):S343-S366. [doi: 10.1123/jpah.2016-0594] [Medline: 27848745]
4. LeBlanc AG, Katzmarzyk PT, Barreira TV, Broyles ST, Chaput J, Church TS, ISCOLE Research Group. Correlates of total sedentary time and screen time in 9-11 year-old children around the world: the International Study of Childhood Obesity, Lifestyle and the Environment. PLoS One 2015;10(6):e0129622 [FREE Full text] [doi: 10.1371/journal.pone.0129622] [Medline: 26068231]
5. Inchley J, Currie D, Young T. Growing up unequal: gender and socioeconomic differences in young people?s health and well-being. Health Behaviour in School-aged Children (HBSC) Study: international report from the 2013/14 survey. 2016. URL: https://www.euro.who.int/__data/assets/pdf_file/0003/303438/HSBC-No.7-Growing-up-unequal-Full-Report.pdf [accessed 2020-06-14]
6. Lapierre MA, Piotrowski JT, Linebarger DL. Background television in the homes of US children. Pediatrics 2012 Nov;130(5):839-846. [doi: 10.1542/peds.2011-2581] [Medline: 23027166]
7. Taylor G, Monaghan P, Westermann G. Investigating the association between children's screen media exposure and vocabulary size in the UK. J Children Media 2017 Aug 29;12(1):51-65. [doi: 10.1080/17482798.2017.1365737]
8. Bernard JY, Padmapriya N, Chen B, Cai S, Tan KH, Yap F, et al. Predictors of screen viewing time in young Singaporean children: the GUSTO cohort. Int J Behav Nutr Phys Act 2017 Sep 05;14(1):112 [FREE Full text] [doi: 10.1186/s12966-017-0562-3] [Medline: 28870219]
9. Cheng S, Maeda T, Yoichi S, Yamagata Z, Tomiwa K, Japan Children's Study Group. Early television exposure and children's behavioral and social outcomes at age 30 months. J Epidemiol 2010;20 Suppl 2:S482-S489 [FREE Full text] [doi: 10.2188/jea.je20090179] [Medline: 20179364]
10. Lee ST, Wong JE, Ong WW, Ismail MN, Deurenberg P, Poh BK. Physical activity pattern of Malaysian preschoolers: environment, barriers, and motivators for active play. Asia Pac J Public Health 2016 Jul;28(5 Suppl):21S-34S. [doi: 10.1177/1010539516638155] [Medline: 27026634]
11. Institute for Public Health. National health and morbidity survey 2016: maternal and child health (MCH) volume two: maternal and child health findings. 2016. URL: http://iku.moh.gov.my/images/IKU/Document/REPORT/2016/NHMS2016ReportVolumeII-MaternalChildHealthFindingsv2.pdf [accessed 2020-04-19]
12. Sharif R, Chong KH, Zakaria NH, Ong ML, Reilly JJ, Wong JE, et al. Results From Malaysia's 2016 report card on physical activity for children and adolescents. J Phys Act Health 2016 Nov;13(11 Suppl 2):S201-S205. [doi: 10.1123/jpah.2016-0404] [Medline: 27848721]
13. Rosenfield M. Computer vision syndrome: a review of ocular causes and potential treatments. Ophthalmic Physiol Opt 2011 Sep;31(5):502-515. [doi: 10.1111/j.1475-1313.2011.00834.x] [Medline: 21480937]
14. Domingues-Montanari S. Clinical and psychological effects of excessive screen time on children. J Paediatr Child Health 2017 Apr;53(4):333-338. [doi: 10.1111/jpc.13462] [Medline: 28168778]
15. Walsh JJ, Barnes JD, Tremblay MS, Chaput J. Associations between duration and type of electronic screen use and cognition in US children. Comput Human Behav 2020 Jul;108:106312. [doi: 10.1016/j.chb.2020.106312]
16. Madigan S, Browne D, Racine N, Mori C, Tough S. Association between screen time and children's performance on a developmental screening test. JAMA Pediatr 2019 Mar 01;173(3):244-250 [FREE Full text] [doi: 10.1001/jamapediatrics.2018.5056] [Medline: 30688984]
17. van den Heuvel M, Ma J, Borkhoff CM, Koroshegyi C, Dai DWH, Parkin PC, TARGet Kids! Collaboration. Mobile media device use is associated with expressive language delay in 18-month-old children. J Dev Behav Pediatr 2019;40(2):99-104 [FREE Full text] [doi: 10.1097/DBP.0000000000000630] [Medline: 30753173]
18. Adelantado-Renau M, Moliner-Urdiales D, Cavero-Redondo I, Beltran-Valls MR, Martínez-Vizcaíno V, Álvarez-Bueno C. Association between screen media use and academic performance among children and adolescents: a systematic review

and meta-analysis. JAMA Pediatr 2019 Sep 23;173(11):1058-1067 [FREE Full text] [doi: 10.1001/jamapediatrics.2019.3176] [Medline: 31545344]
19. Robinson TN, Banda JA, Hale L, Lu AS, Fleming-Milici F, Calvert SL, et al. Screen media exposure and obesity in children and adolescents. Pediatrics 2017 Nov;140(Suppl 2):S97-S101 [FREE Full text] [doi: 10.1542/peds.2016-1758K] [Medline: 29093041]
20. Rhee K. Childhood overweight and the relationship between parent behaviors, parenting style, and family functioning. Ann Am Acad Pol Soc Sci 2008 Jan;615(1):11-37. [doi: 10.1177/0002716207308400]
21. Hamilton K, Spinks T, White KM, Kavanagh DJ, Walsh AM. A psychosocial analysis of parents' decisions for limiting their young child's screen time: an examination of attitudes, social norms and roles, and control perceptions. Br J Health Psychol 2016 May;21(2):285-301. [doi: 10.1111/bjhp.12168] [Medline: 26462456]
22. Samaha M, Hawi NS. Associations between screen media parenting practices and children's screen time in Lebanon. Telematics Informatics 2017 Feb;34(1):351-358. [doi: 10.1016/j.tele.2016.06.002]
23. Hesketh KR, Lakshman R, van Sluijs EMF. Barriers and facilitators to young children's physical activity and sedentary behaviour: a systematic review and synthesis of qualitative literature. Obes Rev 2017 Dec;18(9):987-1017 [FREE Full text] [doi: 10.1111/obr.12562] [Medline: 28589678]
24. Martin-Biggers J, Alleman G, Hongu N, Worobey J, Byrd-Bredbenner C. Beliefs and barriers to limiting screentime behaviors by parents of preschoolers. J Nutr Educ Behav 2015 Jul;47(4):S61. [doi: 10.1016/j.jneb.2015.04.161]
25. Martin-Biggers J, Spaccarotella K, Hongu N, Alleman G, Worobey J, Byrd-Bredbenner C. Translating it into real life: a qualitative study of the cognitions, barriers and supports for key obesogenic behaviors of parents of preschoolers. BMC Public Health 2015 Feb 26;15:189 [FREE Full text] [doi: 10.1186/s12889-015-1554-3] [Medline: 25886030]
26. Smith BJ, Grunseit A, Hardy LL, King L, Wolfenden L, Milat A. Parental influences on child physical activity and screen viewing time: a population based study. BMC Public Health 2010 Oct 08;10:593 [FREE Full text] [doi: 10.1186/1471-2458-10-593] [Medline: 20929587]
27. Hiniker A, Suh H, Cao S, Kientz J. Screen time tantrums: how families manage screen media experiences for toddlers and preschoolers. Proc 2016 CHI Conf Human Factors Comput Syst 2016:648-660. [doi: 10.1145/2858036.2858278]
28. Golley RK, Hendrie GA, Slater A, Corsini N. Interventions that involve parents to improve children's weight-related nutrition intake and activity patterns—what nutrition and activity targets and behaviour change techniques are associated with intervention effectiveness? Obes Rev 2011 Feb;12(2):114-130. [doi: 10.1111/j.1467-789X.2010.00745.x] [Medline: 20406416]
29. Jaafar S, Mohamad NK, Abd MK, Othman N, Healy M. Malaysia Health System Review: Health Systems in Transition, Vol 3 No 1. Geneva: World Health Organization; 2012. URL: https://apps.who.int/iris/bitstream/handle/10665/206911/9789290615842_eng.pdf [accessed 2020-10-02]
30. Thomas S, Beh L, Nordin RB. Health care delivery in Malaysia: changes, challenges and champions. J Public Health Afr 2011 Sep 05;2(2):93-97 [FREE Full text] [doi: 10.4081/jphia.2011.e23] [Medline: 28299064]
31. Children Statistics, Malaysia, 2017. Department of Statistics Malaysia. URL: https://www.dosm.gov.my/v1/index.php?r=column/pdfPrev&id=WGlmVnppZ2J6b2hGZHFQMmxWQ2UwUT09 [accessed 2020-05-21]
32. Report of Household Income and Basic Amenities Survey 2016. Department of Statistics Malaysia. URL: https://www.dosm.gov.my/v1/index.php?r=column/pdfPrev&id=RUZ5REwveU1ra1hGL21JWVlPRmU2Zz09 [accessed 2020-06-22]
33. Hullings C. A Scale to Assess Perceived Barriers to Physical Activity Among Low-Income High School Students in New Jersey [Thesis]. New Brunswick: Rutgers, The State University of New Jersey; 2015.
34. Carson V, Janssen I. Associations between factors within the home setting and screen time among children aged 0-5 years: a cross-sectional study. BMC Public Health 2012 Jul 23;12:539 [FREE Full text] [doi: 10.1186/1471-2458-12-539] [Medline: 22823887]
35. Asplund KM, Kair LR, Arain YH, Cervantes M, Oreskovic NM, Zuckerman KE. Early childhood screen time and parental attitudes toward child television viewing in a low-income Latino population attending the special Supplemental Nutrition Program for women, infants, and children. Child Obes 2015 Oct;11(5):590-599 [FREE Full text] [doi: 10.1089/chi.2015.0001] [Medline: 26390321]
36. Hinkley T, Carson V, Kalomakaefu K, Brown H. What mums think matters: a mediating model of maternal perceptions of the impact of screen time on preschoolers' actual screen time. Prev Med Rep 2017 Jun;6:339-345 [FREE Full text] [doi: 10.1016/j.pmedr.2017.04.015] [Medline: 28480158]
37. Veldhuis L, van Grieken A, Renders CM, Hirasing RA, Raat H. Parenting style, the home environment, and screen time of 5-year-old children; the 'be active, eat right' study. PLoS One 2014;9(2):e88486 [FREE Full text] [doi: 10.1371/journal.pone.0088486] [Medline: 24533092]
38. Pearson N, Salmon J, Crawford D, Campbell K, Timperio A. Are parental concerns for child TV viewing associated with child TV viewing and the home sedentary environment? Int J Behav Nutr Phys Act 2011 Sep 27:1-8 [FREE Full text] [doi: 10.1186/1479-5868-8-102] [Medline: 21951470]
39. Tandon PS, Zhou C, Lozano P, Christakis DA. Preschoolers' total daily screen time at home and by type of child care. J Pediatr 2011 Feb;158(2):297-300. [doi: 10.1016/j.jpeds.2010.08.005] [Medline: 20980020]

40. Prado CV, Rech CR, Hino AAF, Reis RS. Perception of neighborhood safety and screen time in adolescents from Curitiba, Brazil. Rev Bras Epidemiol 2017;20(4):688-701 [FREE Full text] [doi: 10.1590/1980-5497201700040011] [Medline: 29267753]
41. Bursac Z, Gauss CH, Williams DK, Hosmer DW. Purposeful selection of variables in logistic regression. Source Code Biol Med 2008 Dec 16;3:17 [FREE Full text] [doi: 10.1186/1751-0473-3-17] [Medline: 19087314]
42. Wall-Haas CL, Kulbok P, Kirchgessner J, Rovnyak V. Shared medical appointments: facilitating care for children with asthma and their caregivers. J Pediatr Health Care 2012;26(1):37-44. [doi: 10.1016/j.pedhc.2010.06.007] [Medline: 22153142]
43. Marriage and Divorce Statistics, Malaysia, 2019. Department of Statistics, Malaysia. URL: https://www.dosm.gov.my/v1/index.php?r=column/pdfPrev&id=d1BZVzBZYXVwOTBPdXhGVEJTQWl4dz09 [accessed 2020-06-27]
44. Ahmad BA, Khairatul K, Farnaza A. An assessment of patient waiting and consultation time in a primary healthcare clinic. Malays Fam Physician 2017;12(1):14-21 [FREE Full text] [Medline: 28503269]
45. Jarvis JW, Harrington DW, Manson H. Exploring parent-reported barriers to supporting their child's health behaviors: a cross-sectional study. Int J Behav Nutr Phys Act 2017 May 15;44(17):1-10 [FREE Full text] [doi: 10.1186/s12966-017-0508-9] [Medline: 28587672]
46. Ellison G, Barker A, Yougov TK. Work and care: a study of modern parents. 2009. URL: https://www.equalityhumanrights.com/sites/default/files/research-report-15-work-and-care-study-of-modern-parents.pdf [accessed 2020-07-05]
47. Solomon-Moore E, Matthews J, Reid T, Toumpakari Z, Sebire SJ, Thompson JL, et al. Examining the challenges posed to parents by the contemporary screen environments of children: a qualitative investigation. BMC Pediatr 2018 Apr 07;18(1):129 [FREE Full text] [doi: 10.1186/s12887-018-1106-y] [Medline: 29626932]
48. Lauricella AR, Wartella E, Rideout VJ. Young children's screen time: the complex role of parent and child factors. J Applied Devel Psychol 2015 Jan;36:11-17. [doi: 10.1016/j.appdev.2014.12.001]
49. Whitebread D, Bingham S. Habit formation and learning in young children. 2013 May. URL: https://mascdn.azureedge.net/cms/the-money-advice-service-habit-formation-and-learning-in-young-children-may2013.pdf [accessed 2020-06-13]
50. Lee E, Hesketh KD, Rhodes RE, Rinaldi CM, Spence JC, Carson V. Role of parental and environmental characteristics in toddlers' physical activity and screen time: Bayesian analysis of structural equation models. Int J Behav Nutr Phys Act 2018 Feb 09;15(1):17 [FREE Full text] [doi: 10.1186/s12966-018-0649-5] [Medline: 29426324]
51. Gadsden V, Ford M, Breiner H. Parenting Matters: Supporting Parents of Children Ages 0-8. Washington: National Academies Press; 2016.
52. Hales D, Vaughn AE, Mazzucca S, Bryant MJ, Tabak RG, McWilliams C, et al. Development of HomeSTEAD's physical activity and screen time physical environment inventory. Int J Behav Nutr Phys Act 2013 Dec 05;10(1):1-13 [FREE Full text] [doi: 10.1186/1479-5868-10-132] [Medline: 24313962]
53. Rodrigues D, Padez C, Machado-Rodrigues AM. Active parents, active children: the importance of parental organized physical activity in children's extracurricular sport participation. J Child Health Care 2018 Mar;22(1):159-170. [doi: 10.1177/1367493517741686] [Medline: 29166768]
54. Ibrahim D. Work-life management patterns among public sector workers in Malaysia. 2014 Presented at: Proceedings of SOCIOINT14- International Conference on Social Sciences and Humanities; 2014; Istanbul URL: http://www.ocerints.org/Socioint14_e-publication/papers/470.pdf
55. Tombeau Cost K, Korczak D, Charach A, Birken C, Maguire JL, Parkin PC, et al. Association of parental and contextual stressors with child screen exposure and child screen exposure combined with feeding. JAMA Netw Open 2020 Feb 05;3(2):e1920557 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.20557] [Medline: 32022883]
56. Minges KE, Owen N, Salmon J, Chao A, Dunstan DW, Whittemore R. Reducing youth screen time: qualitative metasynthesis of findings on barriers and facilitators. Health Psychol 2015 Apr;34(4):381-397 [FREE Full text] [doi: 10.1037/hea0000172] [Medline: 25822054]

Edited by R Kukafka; submitted 23.10.20; peer-reviewed by N Ahamd, A Hoskins; comments to author 30.11.20; revised version received 13.12.20; accepted 04.03.21; published 13.04.21.

Please cite as:
Mansor E, Ahmad N, Raj D, Mohd Zulkefli NA, Mohd Shariff Z
Predictors of Parental Barriers to Reduce Excessive Child Screen Time Among Parents of Under-Five Children in Selangor, Malaysia: Cross-sectional Study
J Med Internet Res 2021;23(4):e25219
URL: https://www.jmir.org/2021/4/e25219
doi:10.2196/25219
PMID:33847590

Original Paper

A Direct-to-Public Peer Support Program (Big White Wall) Versus Web-Based Information to Aid the Self-management of Depression and Anxiety: Results and Challenges of an Automated Randomized Controlled Trial

Richard Morriss[1], MD; Catherine Kaylor-Hughes[2], PhD; Matthew Rawsthorne[1], BA; Neil Coulson[3], PhD; Sandra Simpson[4], BSc; Boliang Guo[1], PhD; Marilyn James[3], PhD; James Lathe[5], MSc; Paul Moran[6], PhD; Laila J Tata[3], PhD; Laura Williams[1], PhD

[1]Institute of Mental Health, University of Nottingham, Nottingham, United Kingdom
[2]Department of General Practice, University of Melbourne, Melbourne, United Kingdom
[3]School of Medicine, University of Nottingham, Nottingham, United Kingdom
[4]Research Delivery Team, Nottinghamshire Healthcare NHS Foundation Trust, Nottingham, United Kingdom
[5]Centre for Longitudinal Studies, University College London, London, United Kingdom
[6]School of Medicine, University of Bristol, Bristol, United Kingdom

Corresponding Author:
Richard Morriss, MD
Institute of Mental Health
University of Nottingham
Jubilee Campus
Triumph Road
Nottingham, NG7 2TU
United Kingdom
Phone: 44 1158230427
Email: richard.morriss@nottingham.ac.uk

Related Article:

This is a corrected version. See correction statement: https://www.jmir.org/2021/7/e31543

Abstract

Background: Effective help for depression and anxiety reaches a small proportion of people who might benefit from it. The scale of the problem suggests the need for effective, safe web-based public health services delivered directly to the public. One model, the Big White Wall (BWW), offers peer support at low cost. As these interventions are delivered digitally, we tested whether a randomized controlled trial (RCT) intervention could also be fully delivered and evaluated digitally.

Objective: This study aims to determine the reach, feasibility, acceptability, baseline costs, and outcomes of a public health campaign for an automated RCT of the BWW, providing digital peer support and information, compared with a standard website used by the National Health Service Moodzone (MZ), to people with probable mild-to-moderate depression and anxiety disorder. The primary outcome was the change in self-rated well-being at 6 weeks, measured using the Warwick-Edinburgh Mental Well-Being Scale.

Methods: An 18-month campaign was conducted across Nottinghamshire, the United Kingdom (target population 914,000) to advertise the trial directly to the public through general marketing, web-based and social media sources, health services, other public services, and third-sector groups. The population reach of this campaign was examined by the number of people accessing the study website and self-registering to the study. A pragmatic, parallel-group, single-blind RCT was then conducted using a fully automated trial website in which eligible participants were randomized to receive either 6 months of access to BWW or signposted to MZ. Those eligible for participation were aged >16 years with probable mild-to-moderate depression or anxiety disorders.

Results: Of 6483 visitors to the study website, 1510 (23.29%) were eligible. Overall, 790 of 1510 (52.32%) visitors participated. Of 790 visitors, 397 (50.3%) were randomized to BWW and 393 (49.7%) to MZ. Their mean age was 38 (SD 13.8) years, 81.0% (640/790) were female, 93.4% (738/790) were White, and 47.4% (271/572) had no contact with health services in the previous 3 months. We estimated 3-month productivity losses of £1001.01 (95% CI 868.75-1133.27; US $1380.79; 95% CI 1198.35-1563.23) per person for those employed. Only 16.6% (131/790) participants completed the primary outcome assessment. There were no differences in the primary or secondary outcomes between the 2 groups.

Conclusions: Most participants reached and those eligible for this trial of digital interventions were White women not in recent contact with health services and whose productivity losses represent a significant annual societal burden. A fully automated RCT recruiting directly from the public failed to recruit and retain sufficient participants to test the clinical effectiveness of this digital intervention, primarily because it did not personally engage participants and explain how these unfamiliar interventions might benefit them.

Trial Registration: International Standard Randomized Controlled Trial Number (ISRCTN) 12673428; https://www.isrctn.com/ISRCTN12673428

International Registered Report Identifier (IRRID): RR2-10.2196/resprot.8061

(J Med Internet Res 2021;23(4):e23487) doi:10.2196/23487

KEYWORDS

peer support; digital mental health; depression; anxiety; population reach; productivity; mobile phone

Introduction

Background

Depression and anxiety are, respectively, the first and sixth leading causes of years lived with disability globally among all health problems [1]. In the United Kingdom in 2014, 15% of the general population of adults aged >18 years had depression or anxiety, but only 1 in 3 sought help for these conditions in the preceding 12-month period [2]. In principle, the internet backed by a public health campaign might be a useful platform for reaching people with depression or anxiety who do not or are unable to access face-to-face health care. However, the use of the internet as a potential therapeutic platform raises a series of important concerns about data safety and privacy, effectiveness, user experience and adherence, exclusion of people without access to the internet, and data integration with care [3,4]. It is important to establish who might be reached by such an approach, whether they are already accessing mental health or general health services, and which recruitment approaches and interventions are most effective [5].

In this study, we explored the reach of a recruitment strategy for internet-based therapy directed at the public in one English county (Nottinghamshire, estimated population size of 914,000; aged >16 years in 2017, with 145,000 people with case-level depression and anxiety [6,7]). We used traditional and internet media and contacts through health, social care, and third-sector organizations for a randomized controlled trial (RCT) of 2 well-established digital services: the Big White Wall (BWW, now known as Togetherall) and Moodzone (MZ). BWW [8] offers (1) web-based assessment to assess common mental health problems and comorbid physical conditions; (2) moderated web-based peer support network: a community of peers, professionally staffed at all times, enabling safe support where individuals can choose if they wish to remain anonymous; and (3) guided support: a range of self-managed and facilitated programs for depression and anxiety based on cognitive behavioral therapy and social support principles. It was founded by a service user because existing services did not provide support when it was needed and was first developed as a collaboration among service users, digital experts, and a National Health Service (NHS) organization providing mental health services in London. Although BWW is widely available in many parts of England, there was little uptake of this intervention in Nottinghamshire before 2017. In England, the NHS provides its own free website with general information and contact details of local and national resources to help people with common mental health problems. It is known as NHS Mental Health and Well-being, previously called NHS Choices Moodzone [9].

To date, BWW and other web-based peer support for depression or anxiety have only been evaluated in RCTs conducted through primary or specialist mental health services rather than directly to the public [10-13]. There is uncertainty about how to conduct fully automated RCTs of digital mental health interventions directly targeted at the public [14]. In principle, fully automated RCTs without human contact are less prone to bias and can better elucidate actual treatment effects attributable to the digital intervention than those conducted with human conduct, where human contact may contribute to part of the treatment effect [15]. They can also be relatively inexpensive to run on a large scale [16]. However, it is unclear if and under what circumstances they would be feasible or acceptable to participants [17]. Automated trials have sometimes failed to engage some populations [18,19], whereas others offering otherwise difficult to obtain structured psychological treatments have been more successful [20-22].

Objectives

This study has two main aims: (1) to investigate the reach, feasibility, and acceptability of a public health recruitment campaign using general media, digital media, health, third sector, and social services for a trial in people with probable mild-to-moderate depression or anxiety and (2) to test the feasibility, acceptability, baseline costs to society, and outcomes of conducting a fully automated RCT of 2 established digital

interventions providing moderated web-based peer support and information (BWW) versus web-based information only (MZ).

Methods

Design

The Lived Experience Advisory Panel (LEAP) of patients and public representatives with personal experience of depression and anxiety that was formed for the study advised us on the best approaches to recruit people with depression and anxiety in the community. They recommended using the terms *low in mood* for depression and *stressed* for anxiety. We ran a public health campaign using these terms and offered the opportunity to take part in a study free of charge by comparing an information-giving website (MZ), which is a standard designed with and used within the NHS in the United Kingdom, with a web-based peer support site (BWW). We used a mix of traditional health research recruitment strategies, such as general practitioner (GP) endorsement, outpatient clinics, and support groups, as well as less traditional advertising, such as on buses and trams and via letter box leafleting. Special efforts were made to reach groups regarded as higher risk and harder to reach, such as the farming community. Reach was defined as the absolute number, proportion, and representativeness of individuals willing to participate in a given initiative [23].

The second stage of the study was a single-blind RCT using a fully automated bespoke study website. A member of the public could participate in the trial as they might do with other web-based applications, including consenting to the trial and all decision making, only seeking technical support if and when they need it. Full details are available in the published protocol [24]. Eligible participants self-referred and were recruited through the study website following a public health campaign. Consenting participants were randomly allocated to receive either 6 months of free access to BWW or signposted to the NHS MZ website.

Ethical approval was granted by the Local Research Ethics Committee (REC 16/EM/0204), and the final approval was received from the UK Health Research Authority.

Public Health Campaign

The research team worked closely with a research delivery and support service (the National Institute of Health Research [NIHR] Clinical Research Network East Midlands), a professional marketing business (The Dairy), the study LEAP (18 people aged 25-65 years, 12 females), and the web developer (Ayup). The aim was to establish a brand for the study that was considered by the LEAP and study team to appeal to people who may be affected by low mood and/or stress in line with marketing materials but that would also instill professionalism and confidence with respect to the research project.

Trifold leaflets (Figure S1 in Multimedia Appendix 1), posters, bus and tram adverts, and business cards with quick response codes were subsequently developed, and a marketing plan for distribution and dissemination across Nottinghamshire was created for the duration of the study. We targeted tram routes across central Nottingham and bus routes that were purposefully targeted at the more deprived regions of the county. We took out newspaper and magazine advertisements and also spoke on local radio programs dedicated to health matters. We used a digital marketing agency (Nativve) to develop and implement a targeted local Facebook campaign and employed other social media, such as Twitter, to disseminate the study through appropriate networks.

In addition to the blanket approach that was adopted for the general public, the recruitment team also targeted particular groups, such as those in areas of greater deprivation, through door drops. Other targeted approaches involved asking organizations and health professionals to hand out leaflets or spread the word to their community to patients in their preferred method. For example, health visitors were asked to hand out leaflets to young parents, and Black and minority ethnic (BAME) groups were targeted through a third-sector organization, Awaaz. Places where the internet was accessed for free, such as self-help, third-sector groups, and local authority-funded libraries, were also asked to host leaflets and posters as well as to add these materials to their self-help sections.

Presentations were made by members of the research team to a range of GPs and primary care staff to raise awareness of the study across the county. These presentations were complemented by posters, cards, and leaflets that were displayed in every primary care waiting room in Nottinghamshire. Community pharmacists were targeted through their health promotion work and asked to promote the study through displaying the leaflets and posters, and when they considered it appropriate, verbally informing people who might be collecting medication for their mental health that a trial was being conducted. We targeted other health and social care workers, including educational establishments, such as universities, further education colleges, and those in contact with socially isolated groups, for example, social workers, health visitors, or those who work with mental health problems such as the Improving Access to Psychological Treatment program providing psychological treatments for depression and anxiety, private counselors and wellness in mind, and an NHS-funded public mental health signposting service.

Sample and Eligibility

Potential participants from the county of Nottinghamshire self-referred to the study, and their eligibility was assessed by an automated digital program on the study website. The study website requested the GP's contact details when the person was ineligible for the study.

Inclusion criteria were as follows: patients (1) aged ≥16 years; (2) reside in the county of Nottinghamshire; (3) scored between 10 and 20 on the 9-item Patient Health Questionnaire (PHQ-9) [25] or 10 or more on the 7-item Generalized Anxiety Disorder questionnaire (GAD-7) [26], indicating probable caseness for depression and anxiety, respectively, but not a definite diagnosis of depression or anxiety disorder; (4) had access to the internet through a computer, tablet, or smartphone (Windows, iOS, or Android) device and email address; and (5) were able and willing to give informed consent (through electronic consent).

Exclusion criteria were as follows: patients (1) scored ≥21 on the PHQ-9 (severe depression); (2) scored 2 or 3 on PHQ-9 item

"thoughts that you would be better off dead or of hurting yourself in some way"; and (3) scored ≤10 on PHQ-9 and GAD-7.

BWW and MZ are only available in the English language. Therefore, the website recommended nonparticipation for those who believed they were insufficiently proficient in the use of the English language. There was no test of proficiency in English or information technology literacy.

Participants were ineligible for the trial because they scored in the severe range on the PHQ-9 or scored 2 or 3 on the suicide item of the PHQ-9. In line with other digital studies [27], a national research committee did not allow us to recruit these potential participants because, in their opinion, research into people with severe depression requires a greater duty of care than could be offered over the internet. These excluded participants were provided with an opportunity to request that the study team inform their GP, mental health care team, or caregiver of their current mood state. If the request was not completed, the study team followed up via email, asking if they would like the team to inform their GP or care team. We followed up the excluded participants on one occasion.

Information on the participants and the associated consent forms were provided electronically within the study website. Participants who wished to discuss the study could email and telephone the study team if they had any further questions before consenting to the study. An email confirming consent was sent to each participant once they had fully enrolled.

Interventions

Randomization: Arm 1—BWW

Participants allocated to receive 6 months of free access to the BWW website [8] were invited to create a user profile using a pseudonym that was linked to the trial identification to which they had been assigned within the study website. They had to create a profile within 14 days of being randomized. Participants were able to access any part of the BWW site (apart from the option of personalized therapy or counseling sessions that have to be prescribed by a clinician, that is, not offered directly to the general public) and interact with other users within the boundaries of the site's house rules. Anonymized records of log-ins, time on site, interactions, and page categories were recorded by BWW on behalf of the study team.

Randomization: Arm 2—Participants Allocated to MZ

Participants were directed to the MZ area of the NHS Choices website [9]. Participants were able to access all available materials on mental health, including depression and anxiety. We did not have records of time on site or use of the site. NHS MZ access was used as the control digital resource, so all participants were offered some help for their problems with depression or anxiety, but this control group did not have access to moderated, anonymized peer social support.

Outcome Measures

Once consented, participants were asked to complete self-rated questionnaires to measure well-being, depression, anxiety, work and social adjustment, receipt of services (for economic analysis), social support, and personality dysfunction at baseline. These were completed on the web (though the study website) for approximately 20 to 30 minutes. All data were stored on the website and downloaded and anonymized by the clinical trial manager.

Participation in the study lasted for 6 months. Participants received electronic follow-up invitations at 3, 6, 12, and 26 weeks after randomization to be completed on the website. Each participant was reminded to log onto the study website and complete follow-up measures by email 24 hours before each follow-up and at the follow-up time point. If follow-up was not completed, they received another reminder 48 hours later. Participants were emailed motivational statements encouraging follow-up as well as the offer of entry into a prize draw at the end of the study if they completed at least the primary outcome measure in all follow-up assessments. There were no other attempts to follow up participants using any form of digital, telephone, or face-to-face contact.

Primary Outcome Measure

The primary outcome measure is change in self-rated well-being from baseline to 6 weeks after baseline using the 14-item Warwick-Edinburgh Mental Well-being Scale (WEMWBS) [28].

Secondary Outcomes Measures

The secondary outcome measures are as follows:

- Well-being was measured at 3, 12, and 26 weeks using the WEMWBS.
- The GAD-7 [25] was completed as part of eligibility at baseline and at 3, 6, 12, and 26 weeks, as a measure of anxiety severity.
- The PHQ-9 [26] was completed as part of eligibility at baseline and at 3, 6, 12, and 26 weeks, as a measure of depression severity.
- Social function on the 8-item Work and Social Adjustment Scale [29], a measure of function, was completed at baseline and at 3, 6, 12, and 26 weeks.

Baseline Measures

At baseline, basic sociodemographic characteristics were collected along with measures of health and social care resource use over the 3 months before study entry [30], social support [31], life events over the previous 3 months [32], and personality dysfunction [33].

Modifications to the Original Protocol Conducted in Real Time Based on Participant Feedback

Feedback left by the first 50 participants suggested that they disliked the intrusiveness and length of some of the measures and assessments at baseline. One participant withdrew from the study for this reason. Therefore, compared with our protocol [24], we omitted the 12-item medical outcomes study short-form health survey version 2.0 (SF-12) [34] at all time points and only carried out the economic resource proforma [31] at baseline. At baseline, the number of questions asked fell from 92 to 80 and at each follow-up time point from 50 to 38.

Sample Size

The sample size calculation and justification is outlined in detail in our protocol paper [24]. A total of 676 patients were needed to detect a 3-point (SD 12), minimal clinically important difference for adults on the 14-item WEMWBS [35] at a significance level of .05 with 90% power. After adjusting for a 50% attrition rate at 6 weeks [36], a total of 1352 participants were required for our RCT.

Randomization and Monitoring

The treatment to which a participant was assigned was determined by a computer-generated pseudorandom code using random permuted blocks of varying sizes by a randomization system embedded within the website. No stratification or minimization was performed. Treatment assignment was relayed by the computer program to the participant and opened to the trial manager (CK) who monitored recruitment, data completion, and technical problems with the website.

Statistical Analysis

Feasibility and acceptability was assessed by recruitment and retention during follow-up using descriptive statistics. All analyses were performed on an intention-to-treat basis by a trial statistician blinded to treatment allocation using STATA 16 (StataCorp LLC). As all outcome scores were repeatedly measured at baseline and at 3, 6, 12, and 26 weeks, multilevel modeling was performed to quantify the treatment effect with participants as a level 2 unit and baseline, treatment arm, follow-up time, and interaction of arm×time as a covariate. Missing outcome values were investigated and imputed for all outcomes under the missing at random assumption with 100 data sets imputed for data analysis. REALCOME and STATA 16 were used to impute missingness. Similar models were conducted on observed values to check the robustness of the treatment effect estimates sensitive to the influence of missingness.

Health Economics

We electronically administered the Client Service Receipt Inventory [30] to participants at baseline, which collected data on NHS service use and other costing variables [37-44]. Owing to the inherent comorbidity of mental and physical health, these questions pertained to all service use and all health-related absenteeism in place of condition-attributable service use, as we wished to capture possible changes in service use and service-seeking behaviors. The withdrawal of the SF-12 instrument meant that we were unable to examine the health state utilities associated with mild-to-moderate depressive episodes and anxiety disorders.

The costs of health-related time taken off work (absenteeism) were estimated using the lost wages approach [45]. We adopted a median team multiplier, as although wages are a suitable estimate of marginal productivity losses to businesses, this measure tends to be an underestimation for individuals working in team environments [46]. We assigned population-level gross weekly salaries to individuals by full-time or part-time employment status and gender. We designated participants as part time if they were employed or self-employed and worked less than 30 hours per week and as full time if they were employed or self-employed and worked for more than 30 hours. We note that this does not attribute any value to health-related presenteeism or those who are unemployed owing to ill health.

The prevalence of mild-to-moderate depression and anxiety was derived using the Adult Psychiatric Morbidity Survey [42] by combining the severity of symptoms of common mental disorders, where a score between 12 and 17 represents a diagnosable mild-to-moderate condition, with common mental disorder in the past week using the Clinical Interview Schedule Revised score. We extrapolated productivity losses using disease prevalence by gender to control for observed self-selection and the number of individuals between 18 and 64 years of age active in the labor market (employed or self-employed) in the United Kingdom. Our estimates of direct-to-NHS costs used a larger population, including those who are inactive in the labor market. The data sources of NHS unit costs and the resources used to estimate productivity losses are displayed in Multimedia Appendix 2 (Table S1).

To control for missingness, we assumed the item response was missing at random. Multiple imputation was run following best practices [47,48] for a total of 50 imputed sets (m50) [49]. The final model specification included variables for gender, WEMWBS, PHQ-9, age, education, and employment status alongside our outcomes of interest. Multiple imputation was inclusive of both trial arms to increase the sample size because neither arm had received treatment at baseline. Models for direct-to-NHS costs and productivity losses were conducted separately, at the item response level, because of the inherent subsampling of productivity losses to only those in employment. Models were run multiple times, and distributions were visually inspected to confirm the robustness and stability of our imputations. The 95% bias-corrected CIs were derived from 1000-iteration bootstraps, and all health economics analyses were conducted in Stata/SE 16.1. Pounds sterling were converted to US dollars using the conversion rate published by the Federal Reserve as of March 26, 2021: £1 to US $1.38.

Data on Barriers and Facilitators

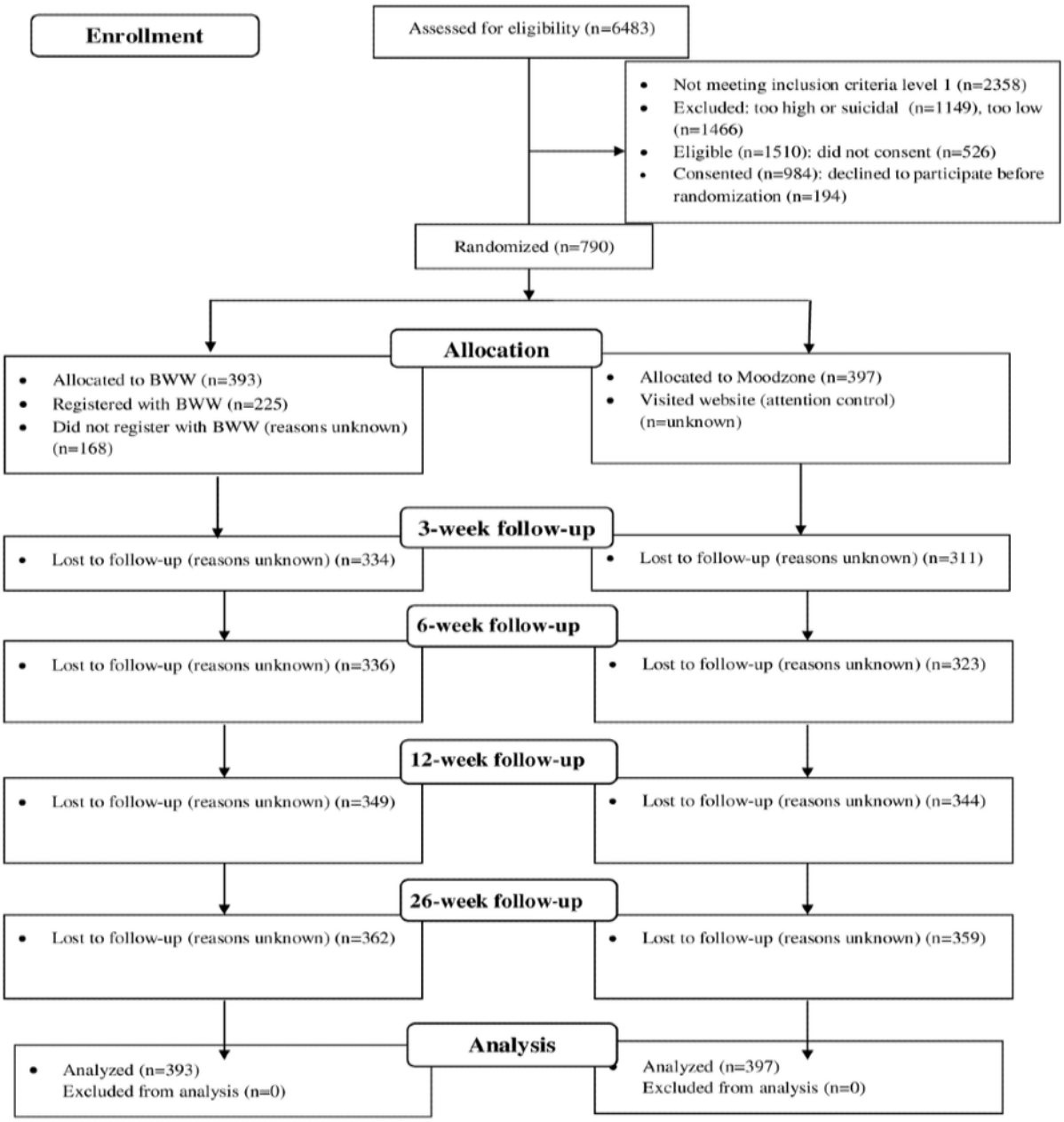
We asked for open-ended survey feedback at baseline and every follow-up and also received email and telephone feedback when participants wished to contact the research team. We then analyzed this feedback thematically.

Results

Overview

Figure 1 shows that there were 6483 visitors to the study website (14 per day) over 18 months of recruitment from September 16, 2016, to May 30, 2018. Of these, 4125 were from Nottinghamshire, aged >16 years, and had continued access to the internet. We excluded a further 1149 participants because their PHQ-9 scores were above 20 or because they scored 2 or 3 on the suicide item of the PHQ-9 and excluded 1466 participants who scored below 10 on both the PHQ-9 and GAD-7. Of the 1510 eligible participants, 790 consented and were randomized, 393 to MZ and 397 to BWW.

Figure 1. Flow of participants through the study.



Methods of Recruitment

Table 1 shows that the 6 most successful ways of recruiting visits to the website were through a university counseling service, direct contact with the research team at presentations, leaflets, pharmacies, and health visitors. However, the 6 most successful enrollment methods to the study (randomization) were through GPs, Facebook, trams, internet and social media, NHS acute hospitals, and leaflet door drop. In relation to primary care, 22,408 registered patients received a text message from their GP endorsing the study, and 180 people were consented and randomized into the study via this route. The efforts directed at raising awareness of this study are listed in Table S2 (Multimedia Appendix 2). Figures S2 and S3 (Multimedia Appendix 2) show recruitment by the geography of Nottinghamshire and in relation to the proximity of tram routes in the county, suggesting that proximity to tram routes in the City of Nottingham was associated with study enrollment.

Table 1. Number of people visiting the study website and those converted to randomized study participants according to their self-reported referral source.

Source	People visiting the study website (N=6483)[a], n (%)	People randomized (percent converted by a referral source; n=790)[a], n (%)	Randomized population by referral source (n=790), n (%)
Awaaz (charity)	2 (0.03)	0 (0.0)	0 (0.0)
Bus advertisement	65 (1.00)	8 (12.3)	8 (1.0)
Direct contact	9 (0.14)	4 (44.4)	4 (0.5)
Door drop	250 (3.87)	58 (23.2)	58 (7.3)
Education	62 (0.96)	13 (20.9)	13 (1.7)
Employer	109 (1.68)	12 (11.0)	12 (1.5)
Facebook	1050 (16.20)	136 (12.9)	136 (17.2)
Family	17 (0.26)	3 (17.7)	3 (0.4)
Friend	213 (3.29)	36 (16.9)	36 (4.6)
General practitioner	977 (15.07)	180 (18.4)	180 (22.8)
Health visitor	15 (0.23)	4 (26.7)	4 (0.5)
Library	30 (0.46)	5 (16.7)	5 (0.6)
Leaflet	48 (0.74)	15 (31.3)	15 (1.9)
Internet	751 (11.58)	63 (8.4)	63 (7.9)
Improving access to psychological therapies program	49 (0.76)	10 (20.4)	10 (1.3)
Magazine	18 (0.28)	1 (5.6)	1 (0.1)
MIND	12 (0.19)	1 (8.3)	1 (0.1)
Newspaper	10 (0.15)	1 (10.0)	1 (0.1)
National Health Service	334 (5.15)	56 (16.8)	56 (7.1)
No recall	1 (0.02)	0 (0.0)	0 (0.0)
Other (unknown)	34 (0.52)	5 (14.7)	5 (0.6)
Pharmacy	135 (2.08)	42 (31.1)	42 (5.3)
Poster	9 (0.14)	2 (22.2)	2 (0.3)
Radio	47 (0.72)	9 (19.2)	9 (1.1)
Social worker	5 (0.08)	0 (0.0)	0 (0.0)
Support service	44 (0.68)	10 (22.7)	10 (1.3)
Tram advertisement	371 (5.72)	67 (18.1)	67 (8.5)
Twitter	20 (0.31)	1 (5.0)	1 (0.1)
University counseling service	17 (0.26)	8 (47.1)	8 (1.0)
Voluntary (third-sector) group	10 (0.15)	2 (20.0)	2 (0.3)
Wellness in mind	5 (0.08)	1 (20.0)	1 (0.1)
No response to question	1764 (27.21)	38 (2.2)	38 (4.8)

[a]Participants were asked how they heard about the study.

Trial Results

Table 2 shows that randomized study participants had a mean age of 38.0 (SD 13.8) years, 81.0% (640/790) were female, 93.4% (738/790) were White, everyone had educational qualifications with 45.0% (354/787) having a university degree, 4.3% (34/790) were retired, and 8.9% (70/790) were unemployed. The mean scores showed moderate levels of depression and anxiety and moderate impairments in function and well-being. Of the 393 participants randomized to BWW, 225 (57.3%) registered to access BWW and 165 (42.5%) accessed it on more than one occasion. No participation data are available for the MZ website.

Table 2. Demographic and clinical characteristics of study participants (n=790).

Characteristics	Randomized to Moodzone (n=397)[a]	Randomized to Big White Wall (n=393)[a]
Age (years), mean (SD)	38.4 (14.3)[b]	37.6 (13.7)[c]
Females, n (%)	316 (79.6)	324 (82.4)
Ethnicity group, n (%)		
White	370 (93.2)	368 (93.6)
South Asian	7 (1.8)	8 (2.0)
Black	5 (1.3)	3 (0.7)
Other	11 (2.8)	14 (3.6)
Missing	4 (1.0)	0 (0.0)
Highest educational attainment, n (%)		
Degree or higher (higher education)	175 (44.1)	179 (45.5)
A levels or Business and Technology Education Council (further education)	126 (31.7)	118 (30.0)
General Certificate of Secondary Education or National Vocational Qualification (basic secondary school)	93 (23.4)	96 (24.4)
No qualifications	0 (0.0)	0 (0.0)
Missing	3 (0.8)	0 (0.0)
Employment, n (%)		
Employed	254 (64.0)	249 (63.0)
Student or training	65 (16.4)	79 (20.1)
Retired	20 (5.0)	14 (3.6)
Unemployed	41 (10.3)	29 (7.4)
Other	17 (4.3)	22 (5.8)
Well-being and mental health assessment scores at study enrollment (self-administered by participant), mean (SD)		
Warwick-Edinburgh Mental Well-Being Scale	34.54 (5.72)[d]	34.54 (5.56)[e]
7-item Generalized Anxiety Disorder Questionnaire	12.99 (3.13)	12.76 (2.91)
9-Item Personal Health Questionnaire	13.66 (2.55)	13.61 (2.54)
Work and Social Adjustment Scale	22.93 (7.07)[f]	23.50 (6.94)[g]

[a]Number of individuals reported only when characteristics were not reported by all participants.
[b]n=309.
[c]n=304.
[d]n=372.
[e]n=359.
[f]n=359.
[g]n=336.

Of the 790 participants enrolled in the study, 572 (72.4%) provided complete case (CC) data on health care service use at baseline. The discussed results correspond to the reported participant service use in the preceding 3 months to baseline (Table 3). Of the 572 patients, no contact with any health service was reported by 271 (47.4%) patients of the CC sample. Out of the 572 patients, only 66 (11.5%) had any contact with mental health services, 228 (39.9%) had contact with their GP and 9 (1.6%) had out-of-hours care. Multimedia Appendix 2 (Tables S1 and S3) shows unit costs, productivity resources, disease prevalence, and CCs of intensity of service use.

Table 3. Baseline self-reported service use in the 3 months before study entry.

Service[a]	Randomized to Moodzone (n=397)		Randomized to Big White Wall (n=393)		Total (N=790)	
	Total, n	Patients, n (%)	Total, n	Patients, n (%)	Total, n	Patients, n (%)
Any service use	293	155 (52.9)	279	146 (52.3)	572	301 (52.6)
Inpatient	269	23 (8.6)	267	15 (5.6)	536	38 (7.1)
General medical ward	257	7 (2.7)	279	8 (2.9)	536	15 (2.8)
Acute psychiatric ward	257	2 (0.8)	279	0 (0.0)	536	2 (0.4)
Outpatient	282	82 (29.1)	277	70 (25.3)	559	152 (27.2)
Emergency room	273	20 (7.3)	286	18 (6.3)	559	38 (6.8)
Radiology	273	17 (6.2)	286	19 (6.6)	559	36 (6.4)
Physiotherapist	273	17 (6.2)	286	13 (4.5)	559	30 (5.4)
Occupational therapist	273	6 (2.0)	286	6 (2.1)	559	12 (4.6)
Psychiatrist	273	15 (5.5)	286	13 (4.5)	559	28 (5.0)
Primary and community	292	115 (39.4)	279	119 (42.7)	571	234 (41.0)
GP[b]	291	113 (38.8)	279	115 (41.2)	570	228 (40.0)
GP home visit	291	2 (0.7)	279	2 (0.7)	570	4 (0.7)
Practice nurse	291	31 (11.0)	279	38 (13.6)	570	69 (12.1)
Psychologist	291	6 (2.1)	279	3 (1.1)	570	9 (1.6)
Psychiatric Nurse	291	4 (1.4)	279	5 (1.8)	570	9 (1.6)
Occupational Therapist	291	2 (0.7)	279	6 (2.2)	570	8 (1.4)
Out-of-hours care	291	4 (1.4)	279	5 (1.8)	570	9 (1.6)
Walk-in center	291	5 (1.7)	279	10 (3.6)	570	15 (2.6)
Social worker	291	1 (0.3)	279	2 (0.7)	570	3 (0.5)
Private counseling or therapy	293	11 (3.8)	276	7 (2.5)	569	18 (3.2)
Other use	293	3 (1.0)	276	9 (3.3)	569	12 (2.1)
No reported information on service use	397	104 (26.2)	393	114 (29.0)	790	218 (27.6)

[a]Other use consisted of other specific National Health Service care, that is, phlebotomist or private health care, that is, physiotherapy and podiatry. Designations of the community psychiatrist were included as outpatient psychiatrist for the 4 participants who double counted by listing psychiatrists within other service use. Binary variables for inpatient or outpatient or primary and community aggregate service use were amended to 1 or 0 if missing and participants specified individual service contact or no individual service use, respectively. If binary variables declared no aggregate service use, individual service use binary variables were set to 0 if missing, and no other individual service use was observed.

[b]GP: general practitioner.

We report the direct-to-NHS costs and productivity losses in Table 4. Participants in employment took a mean 10.93 (95% CI 9.51-12.36) days of health-related time off work during the 3 months, which corresponded to a productivity loss of £1001.01 (95% CI 868.75-1133.27; US $1380.79, 95% CI 1198.35-1563.23), compared with £156.46 (95% CI 114.08-198.84; US $215.82, 95% CI 157.36-274.28) in direct-to-NHS costs per participant. The small variation between the CC and multiply imputed costs suggests that the observed characteristics are poor predictors of response missingness.

Table 4. Direct-to-National Health Service costs and productivity losses in the 3 months before study entry.

Variable	Population, n (%)	Mean (95% CI)	Total annual burden, £ (US $)
Complete case			
Absentee days[a] (n=494)	257 (52)	11.01 (8.62-13.80)	N/A[b]
Productivity losses (n=494)	257 (52)	£995.74 (775.27-1251.21; US $1373.52 [1069.41-1725.92])	2,336,113,572 (3,222,435,061)
Direct-to-NHS[c] costs (n=790)	483 (61)	£158.75 (119.23-209.13; US $218.98 [164.47-288.47])	505,479,259 (697,258,089)
Multiple imputation			
Absentee days (n=494)	494 (100)	10.93 (9.51-12.36)	N/A
Productivity losses (n=494)	494 (100)	£1001.01 (868.75-1133.27; US $1380.79 [1198.35-1563.23])	2,348,477,561 (3,239,489,947)
Direct-to-NHS costs (n=790)	790 (100)	£156.46 (114.08-198.84; US$ 215.82 [157.36-274.28])	498,187,621 (687,200,004)

[a]Days taken off work because of ill health of those employed or self-employed.

[b]N/A: not applicable.

[c]NHS: National Health Service.

Table 5 shows that follow-up assessment rates were very low at each time point. At baseline, 93.5% (739/790) participants completed the WEMWBS, but only 18.4% (145/790) completed it at 3 weeks, 16.6% (131/790) completed it at 6 weeks (the primary outcome), 12.3% (97/790) completed it at 12 weeks, and 8.7% (69/790) completed it at 26 weeks. Proportions of participants completing the primary outcome measure were 14.5% (57/393) in the BWW intervention group and 18.6% (74/397) in the MZ control group. There were no statistically significant differences in the primary outcome WEMWBS at 6 weeks between the 2 treatment groups nor were there any differences at other time points or in the PHQ-9, GAD-7, or Work and Social Adjustment Scale in the imputed or observed results.

Table 5. Modeled mean changes in well-being and mental health scores (95% CIs) at each study follow-up time.

Outcome and follow-up time	Randomized to Moodzone (n=397)		Randomized to Big White Wall (n=393)		Difference	
	Mean (95% CI)	Participant, n (%)[a]	Mean (95% CI)	Participant, n (%)[a]	Mean (95% CI)	P value
7-item Generalized Anxiety Disorder scale						
3 weeks	−2.37 (−3.29 to −1.45)	87 (21.9)	−1.91 (−3.08 to −0.74)	57 (14.3)	0.46 (−1.04 to 1.96)	.55
6 weeks	−2.53 (−3.53 to −1.52)	74 (18.6)	−2.13 (−3.21 to −1.05)	55 (14.0)	0.40 (−1.24 to 2.03)	.63
12 weeks	−2.51 (−3.67 to −1.35)	53 (13.4)	−3.02 (−4.23 to −1.81)	44 (11.2)	−0.51 (−2.26 to 1.24)	.57
26 weeks	−3.26 (−4.63 to −1.89)	38 (9.6)	−3.62 (−5.12 to −2.12)	31 (7.9)	−0.36 (−2.32 to 1.60)	.72
9-item Patient Health Questionnaire						
3 weeks	−1.83 (−2.85 to −0.81)	84 (21.1)	−1.39 (−2.69 to −0.09)	58 (14.8)	0.44 (−1.11 to 1.98)	.58
6 weeks	−1.92 (−2.97 to −0.87)	74 (18.6)	−1.19 (−2.55 to 0.17)	56 (14.2)	0.73 (−0.88 to 2.35)	.37
12 weeks	−2.07 (−3.29 to −0.86)	53 (13.4)	−2.25 (−3.65 to −0.84)	44 (11.2)	−0.17 (−1.92 to 1.58)	.84
26 weeks	−3.36 (−4.62 to −2.10)	38 (9.6)	−2.96 (−4.62 to −1.30)	30 (7.6)	0.40 (−1.69 to 2.49)	.70
Work and Social Adjustment Scale						
3 weeks	−1.31 (−2.97 to 0.35)	79 (19.9)	−2.85 (−5.07 to −0.62)	54 (13.7)	−1.54 (−4.47 to 1.40)	.30
6 weeks	−1.06 (−2.94 to 0.82)	71 (17.9)	−2.68 (−4.83 to −0.52)	51 (13.0)	−1.62 (−4.58 to 1.34)	.28
12 weeks	−1.66 (−3.58 to 0.26)	52 (13.1)	−3.24 (−5.74 to −0.73)	39 (9.9)	−1.57 (−5.07 to 1.92)	.37
26 weeks	−4.21 (−6.39 to −2.02)	37 (9.3)	−5.95 (−8.57 to −3.33)	29 (7.4)	−1.74 (−5.08 to 1.59)	.30
Warwick-Edinburgh Mental Well-being Scale						
3 weeks	2.69 (1.08 to 4.29)	86 (21.7)	2.56 (0.80 to 4.33)	59 (5.0)	−0.12 (−2.55 to 2.30)	.92
6 weeks[b]	3.15 (1.42 to 4.89)	74 (18.6)	1.96 (−0.06 to 3.99)	57 (14.0)	−1.19 (−3.71 to 1.33)	.35
12 weeks	3.92 (1.86 to 5.98)	53 (13.4)	4.29 (2.08 to 6.49)	44 (11.2)	0.37 (−2.64 to 3.38)	.81
26 weeks	5.35 (3.47 to 7.23)	38 (9.6)	6.63 (4.25 to 9.02)	31 (7.4)	1.29 (−1.64 to 4.21)	.39

[a]Participants with available scores at each time point.

[b]Primary outcome of randomized controlled trial.

Barriers to Participation and Retention in the Study

We collected the following feedback organized into 3 themes on barriers to participation and retention in the study from randomized participants, those who had considered participating, and those who refused to participate but left comments for the research team by email, text, or survey.

Lack of Personal Interaction With the Research Team

The study was set up to be automated so that a member of the public could participate in the trial, as they might do with other web-based applications, only seeking technical support if and when they needed it. However, this meant a lack of personal connection and engagement with the research team. Participants described this as contributing to a lack of obligation to complete the study measures and participate in follow-up time points. They viewed the interventions as similar information-giving interventions that were impersonal, with many participants not understanding how the interventions might be tailored to their needs.

Turning Away People With Severe Depression

Potential participants with more severe depression who were trying to take part but were turned away by the automated eligibility criteria on the REBOOT website expressed disappointment, frustration, and a sense of exclusion made apparent through a number of complaints to the study's email account.

Lack of Technical Support

Although people contacted the research team over technical problems, contact with the research team usually led to greater engagement by those participants during follow-up. The study experienced technical issues, such as website downtime, problems with progression through the site, and errors within the measures, which may have deterred the completion of some measures and retention in the study.

Discussion

Principal Findings

Overall, we found that running a fully automated, web-based intervention trial was challenging. Exclusion criteria were exclusive web-based enrollment and measurement, restricted

full enrollment, and retention during follow-up. We recruited and randomized only 790 of 1510 (52.3%) of those who expressed interest and were eligible, despite a considerable amount of effort by the research team using traditional advertising, internet and social media, health services, other public services, and third-sector contacts. Only 16.6% (131/790) of the patients completed the primary outcome at 6 weeks. As a result, the trial lacked the power to demonstrate any differences in outcome between BWW and MZ. The primary reasons for the lack of recruitment and poor retention in this automated RCT are discussed.

There was a demand for web-based information and peer support for the sort offered by the study intervention from people with severe depression and those who were actively suicidal. Far more of these visitors to the website were willing to participate in the study (1149/4125, 27.85%) than those with mild-to-moderate depression and anxiety who were enrolled (790/4125, 19.15%). We were not allowed to recruit these participants by a research ethics committee because of the opinion that research on people with severe depression requires a greater duty of care than could be offered over the internet. However, such patients access digital services on an everyday basis outside the research environment. A restrictive approach to research ethics may have an undesirable effect in that research into digital interventions is not carried out in the most vulnerable manner, where there is arguably more room for beneficial effects and greater safety concerns [50,51].

Participants were most successfully enrolled through GPs (directly and through texting from the practice), pharmacies, internet resources such as Facebook and social media, public transport advertising, and door-drop–posted information in more deprived communities. We recruited a largely female, White, and educated sample who were mostly currently in work or education. Approximately half of those enrolled were not in contact with any health service in the preceding 3 months. A core aim of providing digital mental health approaches to reach people who are not in contact with health services was achieved. Younger females are a part of the population with increasing rates of depression and anxiety in the United Kingdom [2], suggesting that such people might be reached through digital direct-to-public services. However, we failed to recruit enough males, older participants, people from BAME backgrounds, people without any educational qualifications, and people in more rural areas, each of whom may require a combination of different strategies for enrollment.

Compared with the UK average of 4.1 sickness absence days of the general working population, our employed participants took a mean of 43.72 (95% CI 38.04-49.44) days off work per annum, with many individuals undergoing prolonged absences of illness from work. Individuals with mild-to-moderate depression or anxiety may represent a conservative annual burden of approximately £498 million (US $687 million) to the NHS and an additional £1.42 billion (US $1.96 billion) in productivity losses. Our extrapolation to a 1-year time horizon may underestimate or overestimate this burden. There exists a significant value for treating these individuals outside of the standard measures of health gain.

From feedback to the study, the decision to take part in the study was made very quickly by participants, many of whom were not necessarily committed to completing it. Participants found all contact with the study remote and therefore not engaging. Our research group has published a systematic review and meta-synthesis of qualitative data on enrollment and retention in the study or treatment from 24 trials of digital interventions with varying degrees of human engagement in enrollment, follow-up, or treatment of people with depression and anxiety. It identified that enrollment and retention to studies were determined by the participants' initial beliefs about digital health interventions, the offer of personal support, and the enablement of personalization of care [52]. Taking our results together with this meta-synthesis, the public health campaign and automated enrollment only provided a superficial understanding of what was being offered by the trial. The opportunities provided by BWW and MZ to use novel approaches to making choices about obtaining personal support and personalized care, particularly through BWW, which is designed to enable such choices to be made and provided almost immediately, were not made explicit and might have been explained better through initial human or virtual human contact. Those trials that have been automated and yet recruited and retained a high proportion of participants offered well-established treatments for depression and anxiety (cognitive behavioral therapy) to populations that could not otherwise access such help [20-22]. The concept of building a program of tailored personal support and formal psychoeducation through interaction with peers and trained guides that BWW offers was not well-known in the county of Nottinghamshire at this time. Better recruitment and retention rates were obtained in another trial of BWW in mental health service users involving more human support [10]. Therefore, our data suggest that a process of interaction with a human or possibly a virtual human is required to ensure a full understanding of what the trial offered, fully informed consent, and commitment to the study, particularly when the digital interventions are not already well understood in the population being studied.

Participants commented on how burdensome they had found the length of the questionnaires, and longer assessments in digital studies may deter participation [53]. Therefore, during the recruitment of the first 50 participants, we decided to stop collecting questionnaires, namely the SF-12. However, this change did not improve either recruitment or retention in the study.

In this study, participants were routed to 2 other websites, BWW and MZ, which each had the option of completing questionnaires examining their mental state similar to our follow-up measures. These may have confused or deterred participants from completing follow-up measures in the study. We only used email reminders of follow-up and did not use reminders through social media, telephone, or arranged face-to-face contact. More persistent follow-up using a variety of different methods might have improved follow-up rates and provided greater clarity that mental state questionnaires on BWW were independent of follow-up questionnaires in the study. We employed motivational statements and the opportunity to take part in a raffle to win store vouchers at the end of the study. Although

these potential intrinsic and extrinsic rewards may sometimes improve retention and follow-up assessment rates in RCTs [54], they are insufficient.

Conclusions

This study shows that the offer directed to the public of a trial of peer support and information-providing interventions for people with probable mild-to-moderate depression and/or anxiety backed by a public campaign was successful in reaching some parts of the population not in contact with primary care or secondary care mental health services. Most of these were White, educated women who were working and studying below the age of 50 years, who were costly in terms of loss of productivity. However, the public health campaign was not successful in enrolling a large number of high-risk groups for depression or anxiety: men, BAME communities, older people, poorly educated people, and people living in rural communities with poor access to traditional services. This fully automated trial was not successful in engaging or retaining participants because it did not recruit people with severe depression who most wanted these interventions and did not adequately explain how the digital interventions could provide personalized care and support.

Acknowledgments

This study was funded by the NIHR Collaborations for Leadership in Applied Health Research and Care East Midlands and acknowledges the support of the NIHR Clinical Research Network Nottinghamshire Healthcare NHS Foundation Trust Research Delivery Team and the REBOOT LEAP. RM acknowledges support from the NIHR Applied Research Collaboration East Midlands Nottingham NIHR Biomedical Research Centre and NIHR MindTech MedTech and In Vitro Collaborative. The views expressed in this publication are those of the authors and not necessarily those of the NIHR, the Department of Health and Social Care, or the BWW.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Trifold leaflet.
[PDF File (Adobe PDF File), 915 KB - jmir_v23i4e23487_app1.pdf]

Multimedia Appendix 2
Further recruitment and economic data.
[PDF File (Adobe PDF File), 136 KB - jmir_v23i4e23487_app2.pdf]

Multimedia Appendix 3
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 2021 KB - jmir_v23i4e23487_app3.pdf]

References

1. World Health Organization. Depression and Other Common Mental Disorders: Global Health Estimates. Geneva: World Health Organization; 2017. URL: https://www.who.int/mental_health/management/depression/prevalence_global_health_estimates/en/ [accessed 2021-04-21]
2. Demyttenaere K, Bruffaerts R, Posada-Villa J, Gasquet I, Kovess V, Lepine JP, WHO World Mental Health Survey Consortium. Prevalence, severity, and unmet need for treatment of mental disorders in the World Health Organization World Mental Health Surveys. JAMA 2004 Jun 02;291(21):2581-2590. [doi: 10.1001/jama.291.21.2581] [Medline: 15173149]
3. Torous J, Andersson G, Bertagnoli A, Christensen H, Cuijpers P, Firth J, et al. Towards a consensus around standards for smartphone apps and digital mental health. World Psychiatry 2019 Feb;18(1):97-98. [doi: 10.1002/wps.20592] [Medline: 30600619]
4. Digital Exclusion Research Report, January 2019. London: ComRes. URL: http://www.broadbanduk.org/wp-content/uploads/2019/02/ComRes_Broadband-Stakeholder-Group_digital-exclusion-research_final-report_February-2019.pdf [accessed 2021-04-13]
5. Wong QJ, Werner-Seidler A, Torok M, van Spijker B, Calear AL, Christensen H. Service Use History of Individuals Enrolling in a Web-Based Suicidal Ideation Treatment Trial: Analysis of Baseline Data. JMIR Ment Health 2019 Apr 02;6(4):e11521 [FREE Full text] [doi: 10.2196/11521] [Medline: 30938686]
6. Public Health Profiles. Public Health England. URL: https://fingertips.phe.org.uk/search/gp%20survey [accessed 2021-04-13]
7. Population Estimates, Borough and Ward. Office for National Statistics. URL: https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates [accessed 2021-04-13]

8. Togetherall. URL: https://togetherall.com/en-gb/about-us/ [accessed 2021-04-13]
9. Mental Health. National Health Services. URL: https://www.nhs.uk/conditions/stress-anxiety-depression/ [accessed 2021-04-13]
10. Hensel JM, Shaw J, Ivers NM, Desveaux L, Vigod SN, Cohen A, et al. A Web-Based Mental Health Platform for Individuals Seeking Specialized Mental Health Care Services: Multicenter Pragmatic Randomized Controlled Trial. J Med Internet Res 2019 Jun 04;21(6):e10838 [FREE Full text] [doi: 10.2196/10838] [Medline: 31165710]
11. Griffiths KM, Mackinnon AJ, Crisp DA, Christensen H, Bennett K, Farrer L. The effectiveness of an online support group for members of the community with depression: a randomised controlled trial. PLoS One 2012;7(12):e53244 [FREE Full text] [doi: 10.1371/journal.pone.0053244] [Medline: 23285271]
12. Crisp D, Griffiths K, Mackinnon A, Bennett K, Christensen H. An online intervention for reducing depressive symptoms: secondary benefits for self-esteem, empowerment and quality of life. Psychiatry Res 2014 Apr 30;216(1):60-66. [doi: 10.1016/j.psychres.2014.01.041] [Medline: 24534125]
13. Tomasino KN, Lattie EG, Ho J, Palac HL, Kaiser SM, Mohr DC. Harnessing Peer Support in an Online Intervention for Older Adults with Depression. Am J Geriatr Psychiatry 2017 Oct;25(10):1109-1119. [doi: 10.1016/j.jagp.2017.04.015] [Medline: 28571785]
14. Bailey JV, Pavlou M, Copas A, McCarthy O, Carswell K, Rait G, et al. The Sexunzipped trial: optimizing the design of online randomized controlled trials. J Med Internet Res 2013;15(12):e278 [FREE Full text] [doi: 10.2196/jmir.2668] [Medline: 24334216]
15. Baumel A, Edan S, Kane JM. Is there a trial bias impacting user engagement with unguided e-mental health interventions? A systematic comparison of published reports and real-world usage of the same programs. Transl Behav Med 2019 Nov 25;9(6):1020-1033. [doi: 10.1093/tbm/ibz147] [Medline: 31689344]
16. Murray E, Khadjesari Z, White IR, Kalaitzaki E, Godfrey C, McCambridge J, et al. Methodological challenges in online trials. J Med Internet Res 2009;11(2):e9 [FREE Full text] [doi: 10.2196/jmir.1052] [Medline: 19403465]
17. Massoudi B, Holvast F, Bockting CLH, Burger H, Blanker MH. The effectiveness and cost-effectiveness of e-health interventions for depression and anxiety in primary care: A systematic review and meta-analysis. J Affect Disord 2019 Feb 15;245:728-743. [doi: 10.1016/j.jad.2018.11.050] [Medline: 30447572]
18. Anguera JA, Jordan JT, Castaneda D, Gazzaley A, Areán PA. Conducting a fully mobile and randomised clinical trial for depression: access, engagement and expense. BMJ Innov 2016 Jan;2(1):14-21 [FREE Full text] [doi: 10.1136/bmjinnov-2015-000098] [Medline: 27019745]
19. Pratap A, Renn BN, Volponi J, Mooney SD, Gazzaley A, Arean PA, et al. Using Mobile Apps to Assess and Treat Depression in Hispanic and Latino Populations: Fully Remote Randomized Clinical Trial. J Med Internet Res 2018 Aug 09;20(8):e10130 [FREE Full text] [doi: 10.2196/10130] [Medline: 30093372]
20. Ly KH, Ly A, Andersson G. A fully automated conversational agent for promoting mental well-being: A pilot RCT using mixed methods. Internet Interv 2017 Dec;10:39-46 [FREE Full text] [doi: 10.1016/j.invent.2017.10.002] [Medline: 30135751]
21. Schure MB, Lindow JC, Greist JH, Nakonezny PA, Bailey SJ, Bryan WL, et al. Use of a Fully Automated Internet-Based Cognitive Behavior Therapy Intervention in a Community Population of Adults With Depression Symptoms: Randomized Controlled Trial. J Med Internet Res 2019 Nov 18;21(11):e14754 [FREE Full text] [doi: 10.2196/14754] [Medline: 31738173]
22. Clarke J, Sanatkar S, Baldwin PA, Fletcher S, Gunn J, Wilhelm K, et al. A Web-Based Cognitive Behavior Therapy Intervention to Improve Social and Occupational Functioning in Adults With Type 2 Diabetes (The SpringboarD Trial): Randomized Controlled Trial. J Med Internet Res 2019 May 21;21(5):e12246 [FREE Full text] [doi: 10.2196/12246] [Medline: 31115345]
23. RE-AIM. URL: http://www.re-aim.org [accessed 2021-04-13]
24. Kaylor-Hughes CJ, Rawsthorne M, Coulson NS, Simpson S, Simons L, Guo B, REBOOT Notts Lived Experience Advisory Panel, et al. Direct to Public Peer Support and e-Therapy Program Versus Information to Aid Self-Management of Depression and Anxiety: Protocol for a Randomized Controlled Trial. JMIR Res Protoc 2017 Dec 18;6(12):e231 [FREE Full text] [doi: 10.2196/resprot.8061] [Medline: 29254909]
25. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [Medline: 11556941]
26. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
27. Morris ME, Aguilera A. Mobile, Social, and Wearable Computing and the Evolution of Psychological Practice. Prof Psychol Res Pr 2012 Dec;43(6):622-626 [FREE Full text] [doi: 10.1037/a0029041] [Medline: 25587207]
28. Stewart-Brown S, Tennant A, Tennant R, Platt S, Parkinson J, Weich S. Internal construct validity of the Warwick-Edinburgh Mental Well-being Scale (WEMWBS): a Rasch analysis using data from the Scottish Health Education Population Survey. Health Qual Life Outcomes 2009;7:15 [FREE Full text] [doi: 10.1186/1477-7525-7-15] [Medline: 19228398]
29. Mundt JC, Marks IM, Shear MK, Greist JH. The Work and Social Adjustment Scale: a simple measure of impairment in functioning. Br J Psychiatry 2002 May;180:461-464 [FREE Full text] [Medline: 11983645]
30. Beecham J, Knapp M. Costing psychiatric interventions. In: Thornicroft G, Brewin C, Wing JK. editors. Measuring Mental Health Needs. London: Gaskell; 2001.

31. Sherbourne CD, Stewart AL. The MOS social support survey. Soc Sci Med 1991;32(6):705-714. [Medline: 2035047]
32. Brugha T, Bebbington P, Tennant C, Hurry J. The List of Threatening Experiences: a subset of 12 life event categories with considerable long-term contextual threat. Psychol Med 1985 Feb;15(1):189-194. [Medline: 3991833]
33. Moran P, Leese M, Lee T, Walters P, Thornicroft G, Mann A. Standardised Assessment of Personality - Abbreviated Scale (SAPAS): preliminary validation of a brief screen for personality disorder. Br J Psychiatry 2003 Sep;183:228-232 [FREE Full text] [Medline: 12948996]
34. Maruish M. User's manual for the SF-12v2 Health Survey, 3rd ed. Lincoln: RI: QualityMetric Incorporated; 2015.
35. Maheswaran H, Weich S, Powell J, Stewart-Brown S. Evaluating the responsiveness of the Warwick Edinburgh Mental Well-Being Scale (WEMWBS): group and individual level analysis. Health Qual Life Outcomes 2012 Dec 27;10:156 [FREE Full text] [doi: 10.1186/1477-7525-10-156] [Medline: 23270465]
36. Powell J, Hamborg T, Stallard N, Burls A, McSorley J, Bennett K, et al. Effectiveness of a web-based cognitive-behavioral tool to improve mental well-being in the general population: randomized controlled trial. J Med Internet Res 2013;15(1):e2 [FREE Full text] [doi: 10.2196/jmir.2240] [Medline: 23302475]
37. National Health Service. 2017/18 Reference Costs. NHS. URL: https://webarchive.nationalarchives.gov.uk/20200501111106/https://improvement.nhs.uk/resources/reference-costs/ [accessed 2021-04-13]
38. Curtis L, Burns A. Unit Costs of Health and Social Care 2018. Canterbury, UK: Personal Social Services Research Unit, University of Kent; 2018.
39. National Audit Office. Department of Health and NHS England: Out-of-hours GP services in England. London, England: National Audit Office; 2014.
40. Nicholson S, Pauly MV, Polsky D, Sharda C, Szrek H, Berger ML. Measuring the effects of work loss on productivity with team production. Health Econ 2006 Feb;15(2):111-123. [doi: 10.1002/hec.1052] [Medline: 16200550]
41. Office for National Statistics. Dataset: Earnings and hours worked, all employees: ASHE Table 1.: Office for National Statistics; 2018. URL: https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/earningsandworkinghours/datasets/allemployeesashetable1 [accessed 2021-04-21]
42. National Health Services. Survey of Mental Health and Wellbeing, England, 2014. Adult Psychiatric Morbidity Survey.: NHS; 2016. URL: https://digital.nhs.uk/data-and-information/publications/statistical/adult-psychiatric-morbidity-survey/adult-psychiatric-morbidity-survey-survey-of-mental-health-and-wellbeing-england-2014 [accessed 2021-04-13]
43. Office for National Statistics. Dataset: Labour market statistics time series.: ONS; 2019. URL: https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/datasets/labourmarketstatistics [accessed 2021-04-21]
44. Office for National Statistics. Estimates of the population for the UK. England and Wales, Scotland and Northern Ireland: ONS; 2019.
45. Mattke S, Balakrishnan A, Bergamo G, Newberry SJ. A review of methods to measure health-related productivity loss. Am J Manag Care 2007 Apr;13(4):211-217 [FREE Full text] [Medline: 17408341]
46. Zhang W, Sun H, Woodcock S, Anis AH. Valuing productivity loss due to absenteeism: firm-level evidence from a Canadian linked employer-employee survey. Health Econ Rev 2017 Dec;7(1):3 [FREE Full text] [doi: 10.1186/s13561-016-0138-y] [Medline: 28105591]
47. White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice. Stat Med 2011 Feb 20;30(4):377-399. [doi: 10.1002/sim.4067] [Medline: 21225900]
48. Royston P, White I. Multiple Imputation by Chained Equations (MICE): Implementation in. J. Stat. Soft 2011;45(4):1-20. [doi: 10.18637/jss.v045.i04]
49. Graham JW, Olchowski AE, Gilreath TD. How many imputations are really needed? Some practical clarifications of multiple imputation theory. Prev Sci 2007 Sep;8(3):206-213. [doi: 10.1007/s11121-007-0070-9] [Medline: 17549635]
50. McCall HC, Hadjistavropoulos HD, Loutzenhiser L. Reconsidering the ethics of exclusion criteria in research on digital mental health interventions. Ethics & Behavior 2019 Oct 31:1-10. [doi: 10.1080/10508422.2019.1684295]
51. Weisel KK, Lehr D, Heber E, Zarski A, Berking M, Riper H, et al. Severely Burdened Individuals Do Not Need to Be Excluded From Internet-Based and Mobile-Based Stress Management: Effect Modifiers of Treatment Outcomes From Three Randomized Controlled Trials. J Med Internet Res 2018 Jun 19;20(6):e211 [FREE Full text] [doi: 10.2196/jmir.9387] [Medline: 29921562]
52. Patel S, Akhtar A, Malins S, Wright N, Rowley E, Young E, et al. The Acceptability and Usability of Digital Health Interventions for Adults With Depression, Anxiety, and Somatoform Disorders: Qualitative Systematic Review and Meta-Synthesis. J Med Internet Res 2020 Jul 06;22(7):e16228 [FREE Full text] [doi: 10.2196/16228] [Medline: 32628116]
53. Brueton VC, Tierney JF, Stenning S, Meredith S, Harding S, Nazareth I, et al. Strategies to improve retention in randomised trials: a Cochrane systematic review and meta-analysis. BMJ Open 2014 Feb 04;4(2):e003821-e003821 [FREE Full text] [doi: 10.1136/bmjopen-2013-003821] [Medline: 24496696]
54. Brueton VC, Tierney JF, Stenning S, Meredith S, Harding S, Nazareth I, et al. Strategies to improve retention in randomised trials: a Cochrane systematic review and meta-analysis. BMJ Open 2014 Feb 04;4(2):e003821-e003821 [FREE Full text] [doi: 10.1136/bmjopen-2013-003821] [Medline: 24496696]

Abbreviations

BAME: Black and minority ethnic
BWW: Big White Wall
CC: complete case
GAD-7: 7-item Generalized Anxiety Disorder
GP: general practitioner
LEAP: Lived Experience Advisory Panel
MZ: Moodzone
NHS: National Health Service
NIHR: National Institute of Health Research
PHQ-9: 9-item Patient Health Questionnaire
RCT: randomized controlled trial
SF-12: 12-item medical outcomes study short-form health survey
WEMWBS: Warwick-Edinburgh Mental Well-being Scale

Edited by G Eysenbach; submitted 13.08.20; peer-reviewed by S Byrne, S Byrne; comments to author 31.08.20; revised version received 19.10.20; accepted 28.10.20; published 23.04.21.

Please cite as:
Morriss R, Kaylor-Hughes C, Rawsthorne M, Coulson N, Simpson S, Guo B, James M, Lathe J, Moran P, Tata LJ, Williams L
A Direct-to-Public Peer Support Program (Big White Wall) Versus Web-Based Information to Aid the Self-management of Depression and Anxiety: Results and Challenges of an Automated Randomized Controlled Trial
J Med Internet Res 2021;23(4):e23487
URL: https://www.jmir.org/2021/4/e23487
doi:10.2196/23487
PMID:33890858

Original Paper

Feasibility of a Web-Based Psychoeducation Course and Experiences of Caregivers Living With a Person With Schizophrenia Spectrum Disorder: Mixed Methods Study

Anna Laine[1], RN, MNSc; Minna Anttila[1], PhD; Heli Hirvonen[1], RN, MNSc; Maritta Välimäki[1,2], PhD

[1]Department of Nursing Science, University of Turku, Turku, Finland

[2]Xiangya School of Nursing, Central South University, Changsha, Hunan, China

Corresponding Author:
Maritta Välimäki, PhD
Department of Nursing Science
University of Turku
20014 University of Turku
Turku,
Finland
Phone: 358 406890546
Email: mava@utu.fi

Abstract

Background: Schizophrenia is a severe mental illness that burdens both patients and caregivers.

Objective: The aim of this study is to examine the feasibility of a web-based psychoeducation course targeted at caregivers of persons with schizophrenia spectrum disorders (SSDs) and to describe their experiences of living with a person with SSD based on the material caregivers produced during the web-based course.

Methods: A convergent, parallel, mixed methods study design was used. First, caregivers' engagement in the course was evaluated quantitatively. Second, the overview of the course feedback was evaluated using quantitative and qualitative methods. Third, the experiences of being a caregiver to a person with SSD were analyzed qualitatively with the thematic analysis of the writings caregivers produced during the web-based course.

Results: A total of 30 caregivers participated in the study and a web-based psychoeducation course. Less than two-thirds (18/30, 60%) completed the course. Content was most often logged for the first module, *Orientation* (3465 log-ins), and the lowest number of log-ins was recorded for the *Daily life* module (1061 log-ins). Feedback on the course varied; over half (10/17, 59%) of the caregivers considered the content to be very good or good, about half (9/17, 53%) considered the website layout to be good, only 6% (1/17) felt that the usability of the website was poor, and no one felt that it was very poor. From the reported experiences of being a caregiver to a person with SSD, 3 themes were formed: the caregiver's own well-being, relationship with the person with SSD, and experience of health care services.

Conclusions: The web-based psychoeducation course for caregivers living with a person with SSD seems to be especially suitable for those who have little experience as a caregiver. In the future, more planning and the consideration of aspects related to the needs of specific target groups, course content, practical arrangements, and scheduling should be taken into account. In addition, although caregivers can improve their own well-being in different ways, they need regular support and cooperation from health care professionals.

(J Med Internet Res 2021;23(4):e25480) doi:10.2196/25480

KEYWORDS

caregiver; informal caregiver; internet; mental health; psychoeducation; schizophrenia; mobile phone

Introduction

Background

Schizophrenia spectrum disorders (SSDs) [1] are severe mental health disorders that affect informal caregivers (later caregivers) [2-5]. Over half of the caregivers of patients with schizophrenia experience burden [2,3], often conceptualized as an increased level of stress [6], financial concerns [7], heavy responsibilities [8], and stigmatized experiences [6]. Caregivers also tend to have health problems such as psychological distress [9,10], anxiety [11], or depression [11,12]. In contrast, caring for a person with schizophrenia can also offer positive experiences for family members [13]. Shiraishi and Reilly [13] found in their systematic review of qualitative studies that caregivers may experience family solidarity, admiration, self-confidence, personal growth, and appreciation in their role. These positive impacts on caregivers are individual and related to each caregiver's coping methods [14]. Mackey and Pakenham [15] also found in their study that coping resources were the most important predictor to better adjustment among caregivers of persons with mental disorders.

Psychoeducation has been found to be effective in reducing relapses and rehospitalization, shortening hospital stays, and increasing adherence in patients with SSD [16]. On the basis of a Cochrane review by Xia et al [16], the results of psychoeducation were better when caregivers were involved. Psychoeducation is recommended as part of treatment in several care guidelines for patients with SSD [17,18]. It can be realized in groups or individually with different kinds of materials, such as written leaflets, videos [16], or web-based materials [19]. During psychoeducation, individuals gain knowledge about the disorder with which they were diagnosed. This increases their understanding of the disorder and helps them to cope with it [16]. In their review of 32 randomized controlled trials with 2858 caregivers, Sin et al [20] found that psychoeducation for caregivers of patients with psychosis reduces caregivers' global morbidities, negative caregiving experiences, and negative expressed emotions. However, time restraints create an obstacle in face-to-face psychoeducation, as health care professionals need to integrate psychoeducation with other duties [21].

The increased use of digital technologies in health care can strengthen the accessibility of health services [22], and digital technologies offer new opportunities for web-based psychoeducation. A recent review of 46 studies [23] found good acceptance of web-based interventions targeted at caregivers of persons with mental health disorders, neurological disorders, or brain injuries. Web-based psychoeducation interventions have been developed for various caregiver groups, including those with a family member with SSD [24-28]. Web-based psychoeducation has several advantages over face-to-face interventions. Access to web-based psychoeducation is easier because participants can join from home, for example. Web-based psychoeducation also has lower costs [29,30]. The possibility of returning to web-based information when needed can support learning [30], and anonymous participation may reduce stigma [29]. Although access to web-based psychoeducation is easy, it requires an internet connection and a digital device, such as a computer or smartphone. In web-based psychoeducation, participants cannot receive the same level of support from health care professionals as they do with face-to-face interventions, and they must take more direct responsibility for their work. Security issues, such as personal information being shared on the internet, can raise ethical concerns [29].

In previous literature, a variety of web-based psychoeducation programs have been found to have a positive impact on caregivers. Rotondi et al [27] targeted their programs at caregivers and patients with SSD. The program included reading materials with information related to schizophrenia, 3 therapy groups (1 for patients, 1 for caregivers, and 1 for patients and caregivers), the possibility of asking questions from professionals, a library of previously asked questions with their answers, and news and activities in the community. Patients and caregivers used the program independently; however, the sessions were run by a therapist and moderated in therapy forums. During a 14-month period, caregivers used the program for an average of 14 hours, whereas patients used it for an average of 46 hours. Furthermore, another study by Rotondi et al [26] found that the use of the program decreased significantly after the first month. Chan et al [24] described a program that included information related to psychosis in reading materials and videos, information about local mental health services, local and international news, and an interactive discussion forum. In their study, caregivers were satisfied with the program; they visited the website 2-5 times for 5-30 minutes each time. Caregivers reported that the website increased their knowledge about psychosis (85%) and their understanding of local resources (78%). In addition, 84% of caregivers agreed or strongly agreed that they would recommend the website to others. Sin et al [25] studied an intervention for caregivers that included 12 sections. One section was a home page with an introduction and instructions for navigation. Information related to psychosis and caring was divided into 8 sections. Of the 8 sections, 2 sections were for web-based forums, 1 was for experts, and 1 was for peer discussion. In their study, 13 out of 14 caregivers would definitely (72%) or probably (21%) recommend the program to others.

Objectives

Despite the wide range of existing literature on psychoeducation courses, more research is needed to identify optimal ways of using web-based interventions to support caregivers of people with mental illness [31] and to find solutions to enable their psychoeducation. To our knowledge, research combining a web-based psychoeducation course and the use of its content to evaluate caregivers' experiences, especially those with a family member with SSD, is sparse [20]. Therefore, in this study, we examined the feasibility of the web-based course targeted at caregivers and described their experiences of being a caregiver for a person with SSD, based on the web-based material caregivers produced during the course. First, we assessed caregivers' engagement in the web-based course. Second, we collected caregivers' feedback of the web-based course. Third, we investigated the experiences of being a caregiver to a person with SSD, portrayed in writings produced during the web-based course.

In this study, we use the term *SSD* to refer to *schizophrenia spectrum disorder* based on the 10th version of International Classification of Disorders (ICD-10) categorization F20-29 [1]. The term *caregiver* refers to a patient's informal caregiver, such as a family member, close friend, or another person close to the patient who takes care of the patient in their daily life [32].

Methods

Design

A convergent, parallel mixed methods study design [33] was used to evaluate the use of a web-based psychoeducation course. Both quantitative and qualitative data collection and analysis methods were used to provide a broad view of the topic [34]. First, to evaluate engagement, quantitative methods were used to collect and analyze caregivers' log records from the web-based psychoeducation course. Second, a combination of quantitative and qualitative methods was used to collect and evaluate caregivers' numeral and written feedback about the course. Third, qualitative methods were used to collect data portrayed in writings and explore experiences of caregivers caring for a person with SSD. The convergent, parallel design [33] was appropriate, as data were collected at different points in time throughout the course. The study was reported according to Good Reporting of a Mixed Methods Study [35].

Setting

The study was conducted within FinFami (Finnish Central Association of Families With People With Mental Illness), which is a central organization of 18 local associations based throughout Finland. The target group of the organization is caregivers of people who are recovering from mental illness. The mission is to promote well-being by providing information, support, and hope. With its local associations, FinFami organizes group activities, courses, and events for caregivers [36]. For convenience purposes, 5 local family associations in Southern Finland were invited to participate in the study.

Eligibility Criteria

Participants were eligible to participate in the study if they met the following inclusion criteria: a caregiver of a person with SSD (ICD-10, F20-F29 [1]), were closely related (eg, parent, spouse, or child) to a person with SSD, were aged 18 years or more, could read and write in Finnish, and were willing to participate in the study. Caregivers were excluded if they were not able to read and write in Finnish, were unable to give informed consent, or the mental illness of their relative was something other than SSD.

Ethical Issues

This study was approved by the Ethics Committee of the Hospital District of Southwest Finland (ETMK 56/2015). Research permission for data collection was granted by each local family association participating in this study. Consent was not requested from persons with SSD as recruitment occurred in caregiver associations.

During the course, tutors of the course contacted caregivers via phone if there was any concern over the health of the caregiver. Phone calls were also made if topics were too confidential to talk about in the discussion forum or too sensitive to address via email. Tutors were ethically responsible [37] for identifying if a caregiver was in need of professional help, and if they were, the tutors would validate the caregiver's situation, support them in getting help, and provide more information on the services available for caregivers.

Sampling and Recruitment

The consecutive sampling method was used to minimize sampling bias by reaching all potential participants during the recruitment phase [38]. First, the researchers approached a volunteer contact person in each local family association. Second, an information session was held for the contact persons to introduce the study and the practical arrangements of the data collection. Third, contact persons were informed about the study in each local association, and caregivers were tentatively invited to participate if they met the inclusion criteria. A total of 3 information meetings for potentially interested participants were arranged for caregivers, during which a short introduction of the study was conducted. The caregivers had an opportunity to introduce themselves and their role as a caregiver. Caregivers were informed about the study aim, methods, possible advantages and disadvantages, ethical issues (voluntariness, privacy, and confidentiality), and practical arrangements [39] of the web-based psychoeducation course and the use of its content to evaluate caregivers' experiences. Caregivers were also informed in oral and written formats about their right to refuse or withdraw at any phase of the study [39,40].

Caregivers who decided to participate in the study signed an informed consent form after an information session. If they needed more time to think about their participation, they took the consent material home and returned it via post. If a caregiver was willing to participate in the study but was not able to join the information session, written information about the study and a consent form for them to sign was sent to them by post, along with a stamped envelope. Caregivers were able to register for the course after providing consent.

Web-Based Psychoeducation Course

The web-based psychoeducation course was designed to offer information and peer support to caregivers of persons with SSD. The web-based method was selected because the internet is already used for seeking help and information related to health problems [41], and there is a need to increase the use of web-based methods in health care [22]. The course consisted of papers and written tasks on the Moodle learning platform and an information package on MentalNet, which is a psychoeducation website providing information related to SSD (eg, etiology, symptoms, treatment, daily living, and patients' rights; ICD-10 codes F20-F29 [1]). The content of MentalNet is based on a literature review; it was designed in cooperation with health care professionals, patients, and caregivers and originally aimed to support patients' self-management [42]. Nurses have recommended MentalNet as useful for family members, making the website a suitable structure for the course [43]. Anonymity was secured with personal user accounts and passwords on the learning platform. More detailed information about MentalNet is available in other publications [42,43].

The web-based psychoeducation course was hosted by the University of Turku. The length of the course was 8 weeks, and it comprised 6 modules consisting of 6 themes: (1) orientation, (2) mental illness, (3) treatment, (4) daily life, (5) well-being, and (6) patient and caregiver rights. Each module lasted 1 week, except *Orientation*, which lasted 2 weeks. Finally, there were concluding remarks and the opportunity to provide feedback. The deadlines of tasks assigned to caregivers were flexible in case the caregivers needed more time to complete the tasks (Table 1).

Table 1. Description of the structure and content of the web-based psychoeducation course.

Module and length	Content	Task of the week
Module 1: Orientation; 2 weeks	• Welcome • Description of the course • Schedule of the course • Introduction of course leaders and tutors • Service promise • News • Information and instructions about the learning platform • Security and net etiquette • Discussion forum	Please read the orientation material and write your expectations from the course. Introduce yourself shortly and anonymously. Describe the well-being of the person you care for and of yourself by answering the following questions: (1) How does the person you care for act, what do they think, and what do they "look like" when they are in poor condition? (2) What does the person you care for look like when they are doing well? (3) What should the person you care for avoid to maintain a good state of mind and a sense of hope? (4) How can you help make this happen? (5) What do you do to maintain your positive state of mind and a sense of hope? (6) What things should you avoid to maintain a positive state of mind and a sense of hope? Voluntary written task: Draw a line on the paper to mark the most important turning points in your life (joys and sorrows). Describe and write what you marked on the paper.
Module 2: Mental illness; 1 week	• Information about the origin of SSDs[a] and its symptoms • Additional information about how caregivers can help themselves and their relatives with the illness in different stages of the disorder	Please read the theme "Mental illness" and answer the following questions in writing: (1) What kind of difficult situations have you experienced related to your relative's disorder? (2) What things have not worked well in getting through difficult situations? (3) What things have worked well in difficult situations that you would like to share with other caregivers? (4) Please comment on other caregivers' writings in this module.
Module 3: Treatment of the illness; 1 week	• Information about the content of care and rehabilitation for SSD • Additional information about research concerning the care of SSD and different types of care and restrictive treatment methods	Please read the theme "Treatment" and answer the following questions in writing: (1) What is the meaning of a written treatment plan for your relative? (2) What does it mean for your relative if they are treated in inpatient or outpatient care? (3) What does it mean for you if your relative is treated in inpatient or outpatient care?
Module 4: Daily life; 1 week	• Information about support from caregivers • Information about different types of support and benefits that might help in managing daily life	Please read the theme "Daily life" and answer the following questions in writing: (1) What kind of things support your relative's daily life? (2) What kind of things support your own daily life? (3) In what kind of situations is there an increased risk that your relative would not take their medication? (4) What kind of things support your relative in taking their medicine properly?
Module 5: Well-being; 1 week	• Information about mental and physical well-being in relationships	Please read the theme "Well-being," do a writing, and answer and share the following questions in writing with other caregivers: (1) How can you affect your own well-being? (2) How can you decrease the burden caused by the care of your relative?
Module 6: Patient and caregiver rights; 1 week	• Information about patients' rights for sufficient and necessary treatment, self-determination, and their rights to be informed about their treatment and to keep their information confidential • Information about compulsory treatment • Information about caregivers' rights in a relative's treatment	Please read the theme "Patient and caregiver rights" and answer the following questions in writing: (1) How have you acted in situations where a relative with the disorder has not allowed you to take part in their treatment even if you are worried and want to take part? (2) What is good or bad about dealing with health care in your opinion? (3) How would you change practices related to compulsory treatment?

[a]SSD: schizophrenia spectrum disorder.

The caregivers participated in the course at home on their own computers using pseudonyms. For psychoeducation to be truly empowering, it should provide versatile information covering a wide variety of topics (eg, financial help and ethical aspects) in addition to subjects related to the disorder itself [44]. The themes and tasks in the course were structured to provide caregivers with a comprehensive program that would help them identify any problems in their daily lives and explore solutions to the problems. The layout of the course was predesigned [42,43]. However, each participant could tailor the specific content of the 6 modules based on their own needs. Each participant reflected on their own situation and sought answers to their questions using the web-based material available on the course website. The courses were designed to support caregivers' self-management and well-being and help them recognize their own resources. In addition, caregivers were encouraged to share their thoughts in a discussion forum using a peer support approach. The specific content of the course

material in each module was based on a variety of evidence-based sources on the MentalNet website [45,46].

A total of 3 trained professional tutors specialized in mental health provided feedback after each completed module and answered any emerging questions. Of the 3 trained professional tutors, 2 were psychiatric nurses—one was a registered nurse, a family psychotherapist, and a master's student in nursing science and the other was a doctoral student (a registered nurse with extensive experience in psychiatric nursing). The third tutor had a PhD. Via emails or text messages, the tutors encouraged caregivers to continue in the course, share their worries, and use the psychoeducation website for support. Moreover, a course coordinator answered practical questions related to the study or course and informed the caregivers about the phases of the course.

Data Collection

Summary of Data Collection

Data were collected between November 2015 and January 2016. Sociodemographic information was obtained with a paper questionnaire, either filled out at the end of the information session or returned by post. The quantitative data related to caregivers' engagement in the web-based course, which was based on the course's access log, were collected during the course. Quantitative and qualitative feedback of the course was collected after the course (by January 3, 2016) using a web-based questionnaire as part of the course tasks. Finally, caregivers' experiences were gathered qualitatively from the learning platform after the course. The researcher copied all caregivers' writings into one Microsoft Word document, which comprised 156 pages (font: Arial 11; spacing: 1.5).

Data Collection Methods

Feasibility of the Web-Based Psychoeducation Course

Engagement in the Course

Caregivers' engagement in the course was assessed by calculating their activity on the learning platform. Caregivers' log-ins for each module were automatically recorded on the learning platform. The number of caregivers visiting every module and the number of finalized module tasks (writings) were calculated manually.

Feedback on the Course

A questionnaire designed by the research group was used for feedback. The questionnaire included three 5-point Likert scale questions (eg, "My feedback about the content of the psychoeducation course"; 1=very poor; 2=poor; 3=neither good nor poor; 4=good; 5=very good) with items about content, layout, and usability of the course [47,48]. The participants were also invited to provide written feedback about the course and the website.

Caregivers' Experiences of Caring for a Person With SSD

The qualitative data concerning experiences of being a caregiver for a person with SSD were portrayed in writings [49] designed primarily for the psychoeducation course. Caregivers were instructed to read specific course material assigned for each week, complete writing tasks based on their experiences and guided with open-ended questions related to each theme, and return their tasks via the web-based platform (Table 1).

Sociodemographic Information

Caregivers were asked to provide background information on a form regarding their age, gender, marital status, educational level, employment status, housing situation, and relationship with the person with SSD. They were also asked to provide some information about the person with SSD (the duration of mental illness and duration of using mental health services). In addition, caregivers were asked about any possible chronic illnesses they might have and their skills and attitudes regarding computers and the internet.

Data Analysis

An evaluation of the use of the course was based on an analysis of data collected at different time points. Caregivers' engagement in the course was evaluated using descriptive analysis methods. The frequencies of the log-ins to each module were calculated. The number of caregivers visiting each module and completing the tasks over the duration of the course was calculated. Participants' feedback on the course was then evaluated using descriptive statistics (frequencies and percentages).

For written feedback, data were categorized and analyzed using qualitative content analysis [50]. The written feedback was collected into one document and read through so that the researchers could familiarize themselves with the data. Similarities were identified and classified into categories.

To investigate caregivers' experiences of being a caregiver to a person with SSD, the writings produced during the course were analyzed using thematic analysis. Thematic analysis was chosen because it can be used to identify, analyze, organize, describe, and report themes found in the qualitative data [51]. First, the first author (AL) read all writings several times to become familiar with the data and to get an idea of the experiences of these caregivers. Second, initial codes such as *individual hobby* and *social relations* were generated. During coding, sections of the text were marked, indicating the code or subject. Third, potential subthemes were generated inductively from the initial codes. The marked sections were copied and collected into separate files. Fourth, the coded data and potential subthemes were reviewed and refined by collapsing and separating them to ensure that they were in a coherent pattern. An example of a coding tree is presented in Table 2. The data were reread to confirm that the themes were related to the data set and that nothing had been overlooked. Fifth, the themes were defined and named after identifying what each theme contained and considering an appropriate name [51]. One researcher (AL) conducted the coding process. Two researchers (AL and MA) defined the subthemes and themes.

Table 2. Example from a coding tree of caregivers' well-being.

Phrase in the text	Initial code	Subtheme	Theme
Reading a book	Individual hobby	Hobbies	Caregivers' experiences of factors supporting well-being
Jogging	Individual hobby	Hobbies	Caregivers' experiences of factors supporting well-being
Friends	Social relations	Support from other people	Caregivers' experiences of factors supporting well-being
Family	Social relations	Support from other people	Caregivers' experiences of factors supporting well-being

Results

Participants

The participant flow and the sociodemographic information about participants are presented in Figure 1 and Table 3, respectively.

Figure 1. Flow diagram of the participants.

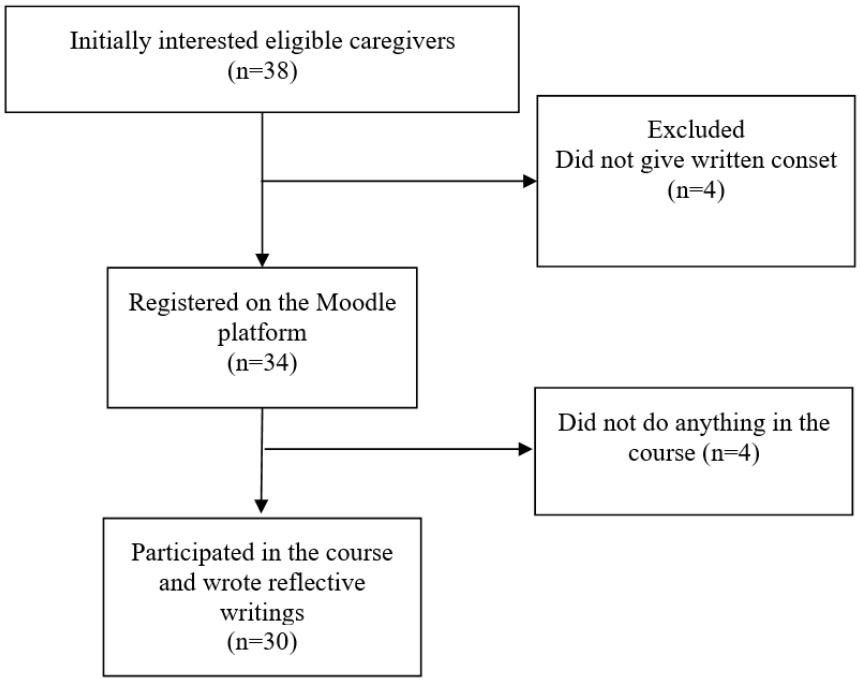


Table 3. Sociodemographic information of participating caregivers (N=30)[a].

Characteristic	Value
Age (years; n=29), mean (SD)	59.7 (9.9)
Gender (n=28), n (%)	
Female	23 (82)
Male	5 (18)
Marital status (n=28), n (%)	
Unmarried	3 (11)
In common-law marriage or married	17 (61)
Divorced	6 (21)
Widow	2 (7)
Educational level (n=27), n (%)	
Primary school	5 (19)
Vocational school or vocational courses	10 (37)
Bachelor's degree	2 (7)
Master's degree	10 (37)
Employment status (n=29), n (%)	
Unemployed	2 (7)
Student	1 (3)
Retired	16 (55)
Employed or on sick leave	10 (34)
Housing situation (n=29), n (%)	
Living with a spouse or family	20 (69)
Living alone	9 (31)
Relationship with the person with SSD[b] (n=29), n (%)	
Spouse	5 (17)
Child	6 (21)
Parent	15 (52)
Other	3 (10)
Duration of the illness of the person with SSD (years; n=28), n (%)	
0-10	14 (50)
11-20	11 (39)
21-30	3 (11)
Duration of using mental health services (years; n=27), n (%)	
0-10	10 (37)
11-20	14 (52)
21-30	3 (11)
Chronic illness of a caregiver (n=29), n (%)	
No	18 (62)
Yes	11 (38)
Computer or internet skills (n=29), n (%)	
Very good	7 (24)
Good	13 (45)

Characteristic	Value
Neutral	8 (28)
Quite poor	1 (3)
Poor	0 (0)
Attitudes toward computers or internet (n=29), n (%)	
Very positive	12 (41)
Positive	15 (52)
Neutral	2 (7)
Negative	0 (0)
Very negative	0 (0)

[a]The number of responses varied depending on how many participants answered each question.

[b]SSD: schizophrenia spectrum disorder.

Feasibility of the Web-Based Course

Regarding engagement in the web-based psychoeducation course, the highest number of log-ins was recorded for the *Orientation* module (3465 log-ins), and the lowest number of log-ins was recorded for module 6, *Daily life* (1061 log-ins; Multimedia Appendix 1). A total of 30 (100%) caregivers visited the modules, 25 (83%) caregivers completed at least one of the main modules besides *Orientation*, and 18 (60%) caregivers completed all the 6 modules. The numbers of caregivers who visited and completed the modules are presented in more detail in Multimedia Appendix 2.

Feedback on the Web-Based Psychoeducation Course

Feedback Characteristics

More than half (10/17, 59%) of the caregivers evaluated the content of the psychoeducation as very good or good, about half (9/17, 53%) evaluated the website layout to be good, only 6% (1/17) considered the usability of the website to be poor, and no one felt it was very poor (Table 4).

Table 4. Feedback about content, layout, and usability (n=17).

Dimension of the feedback	Caregiver, n (%)
Content of the website	
Very good	2 (12)
Good	8 (47)
Neither good nor poor	7 (41)
Poor	0 (0)
Very poor	0 (0)
Layout of the website	
Very good	0 (0)
Good	9 (53)
Neither good nor poor	6 (35)
Poor	2 (12)
Very poor	0 (0)
Usability of the website in psychoeducation	
Very good	2 (12)
Good	5 (29)
Neither good nor poor	9 (53)
Poor	1 (6)
Very poor	0 (0)

Caregivers' written feedback about the course was divided into 2 categories: (1) content of the course and (2) course arrangements.

Content of the Course

The caregivers found the content of the course to be interesting. The theoretical information was considered good, and caregivers

used it in conversations with professionals. In addition, the tasks were found to be good. Caregivers found the writings of other participants valuable because they could compare their own situation with others' situation and read how others had survived. However, they were of the opinion that the content of the course did not provide new information for long-time caregivers. They did not find peer support when there were no similar situations or there was only limited conversation between the caregivers. In addition, they found it depressing to read others' stories.

Course Arrangements

The caregivers found that the course was well constructed. Caregivers learned how to use the learning platform during the course. There were, however, some technical problems, especially at the beginning of the course. The caregivers felt that the end of the year was not a good time to take part in the course. They found that there was too little time allocated for each module. In addition, the information meeting at the beginning could have been a web-based conversation instead of a face-to-face meeting.

Experiences of Caregivers Caring for Persons With SSD

Themes and Subthemes

An overview of the themes and subthemes is presented in Textbox 1.

Textbox 1. Themes and subthemes.

Influence on caregivers' lives

- Experiences of the impact of illness on the caregivers themselves
  - Physical well-being
  - Mental well-being
- Factors supporting the caregivers themselves
  - Positive mood
  - Hobbies
  - Everyday routines
  - Support from other people and the health care system

Relationship with the person with schizophrenia spectrum disorder

- Caregivers' experiences of a good relationship
  - Help in everyday life
  - Mental support
  - Help related to the illness
- Caregivers' experiences of challenges in a relationship
  - Physical distance
  - Challenges related to schedules
  - Difficulties in communication
  - Challenges related to stage of the illness

Experience of health care services

- Caregivers' positive experiences in the treatment of the illness
  - A sense of security created by hospital care
  - Benefits of outpatient care
  - Good cooperation with health care
- Challenges faced by caregivers in the treatment of the illness
  - Inadequacy of outpatient care
  - Challenges of multiple diagnosis
  - Challenges in cooperation between caregivers and health care
  - Challenges related to commitment in care
  - Challenges related to legislation

Influence on Caregivers' Lives

The influence on caregivers' lives was divided into 2 themes: (1) experiences of the impact of illness on the caregivers themselves and (2) factors supporting the caregivers themselves. Living with a person with SSD has an impact on caregivers themselves. Caregivers expressed signs of depression with feelings of sadness, tiredness, and dejection. Burdening feelings appeared when caregivers were worried about the management of the person with SSD and other family members. They also described a link between the well-being of the caregiver and that of a person with SSD. For example, caregivers found that their own well-being worsened when the state of illness of the person they cared for deteriorated:

> *After all, the camel's back was broken because I was and lived alone. In addition to physical disorders, depression hit as it hits many other caregivers.*
>
> *During the illness, the fear of mine has been related to coping of other family members.*
>
> *I notice that my own mood depends on my son's situation.*

Caregivers described events or occasions that supported them. These included positive mood, everyday routines, hobbies, and

support from other people. Caregivers tried to keep themselves positive by using humor to get through hard moments and tried to avoid energy-consuming people and events. Maintaining regular routines, such as maintaining proper nutrition and regular rest, helped to manage daily activities. Hobbies, including exercise, cultural activities, societal activities, voluntary work, and traveling, gave meaning to their lives. Caregivers received support from other people, such as family members and friends, who helped in difficult times. Colleagues also showed support, as they helped caregivers focus on work. Support from the official social and health service sector included discussions with professionals when the relative was in hospital. Support from the service sector consisted of peer support and family association meetings. Caregivers understood that the passage of time supported their well-being and that they needed to take care of themselves to support others and live their own life:

> *Focusing on "normal things," sense of humour and friends are what carries me if it otherwise goes against the walls.*
>
> *I try to do things I like: I meet people, I have hobbies like studying and exercise, theatre and travelling.*
>
> *When I feel particularly difficult and hopeless, I talk to a friend. I go and talk with a professional about my situation.*

Relationship With the Person With SSD

Caregivers described good relationships with the persons with SSD they cared for. They provided practical daily support for the persons with SSD, such as assistance in shopping or household chores. They encouraged people with SSD to keep up their routines and stay away from substance abuse. For example, they participated in public events together without being ashamed. Caregivers discussed general topics that the person with SSD was interested in, introduced easy-to-discuss topics, and tried to get the person to think more positively. They also paid attention to whether they were showing symptoms or not. Caregivers supported the person with the illness in seeking help, even if it was a hard decision. They were involved in planning and monitoring care and attended meetings organized with mental health professionals:

> *I visit the city about once a week with my mother. At that time, we go banking, shops and other places.*
>
> *I try to encourage her and talk about her good sides and what she is good at. The illness is not usually seen but can sometimes be, so I have decided not to be ashamed of her anywhere.*
>
> *We go to public events and we have done trips together abroad.*
>
> *Fortunately, at least for now I have had an opportunity to join the care meetings and my wife has wanted me to be there.*

Caregivers described challenging relationships with the persons with SSD. These challenges were related to physical distance, timetables, the stage of illness, and challenges in communication. Sometimes, the physical distance between the caregiver and the person with SSD was too great for them to meet each other often. Schedules were sometimes difficult to coordinate, or there was only a short amount of time to see each other. Sometimes, in a bad stage of the illness, the person with SSD might not want to be in touch or want the caregiver to be involved if they were in hospital. The person with SSD could also be threatening toward the caregiver. Communication with the person with SSD was challenging when that person lacked insight, and caregivers had to be careful about what topics could be discussed without triggering anxiety or anger:

> *She lives with her partner in another city, so we are mainly in contact by phone, in the summertime we meet more.*
>
> *I try to visit her weekly, but we often have scheduling conflicts, so our meetings do not always succeed. Many times, I have needed to withdraw, change the subject and let it be, because arguing with a man with lack of insight is somewhat hopeless.*
>
> *Once he picked up a scythe. I looked at the exits and didn't stay long that time.*

Experience of Health Care Services

Positive experiences related to care included the feeling of safety created by hospital care, good cooperation with health care services and support from professionals, and advantages of outpatient care. Caregivers were relieved when the person with SSD was safe in a hospital when they needed it. The cooperation with health care services was good when the person with SSD had had a long relationship with the same professionals, and caregivers were familiar with them and were able to contact them easily. Caregivers received concrete support from professionals, such as nurses and police, when they needed help getting the person with SSD into hospital care. Outpatient care supported the person, enabled them to be out of hospital, and enabled caregivers to see the person and how they were coping;

> *I was able to sleep a bit longer at night when she was in the hospital, and during the day I didn't have to worry about what was waiting at home when I came back from work.*
>
> *Outpatient care is also the best option for me. I can keep my husband at home and I see him every day. Likewise, I know when he goes to see his therapist and in what condition he comes home from there.*
>
> *As the departments and staff became more familiar, cooperation began to work better in every way.*

On the contrary, the challenges faced by caregivers included insufficient care, problems in cooperation between the caregiver and the health care service, and legislative issues. Caregivers described how physical symptoms and side effects of medicine were not taken seriously or were not appropriately dealt with, and there were challenges with substance abuse problems. Sometimes, it took long to get help for the person with SSD, and outpatient care seemed powerless. Caregivers had witnessed inappropriate behavior, such as verbal mocking and excessive use of force by the police when they had taken the person to a hospital. The person with SSD sometimes denied caregivers access to information, and the opinions or experiences of caregivers might have been ignored when professionals had contradictory views about the need for inpatient care.

Furthermore, caregivers and health care professionals did not understand each other, especially if a doctor was not a native Finn. Legislation related to self-determination was a barrier to care if the patient wanted to stop their medication or treatment. Examples of the phrases are provided below:

> *Of outpatient care, I often stated that it was not treatment at all, but indeed neglect. If the patient does not come to the appointment once a month, no one from healthcare services goes home to see what the situation is.*
>
> *The first thing that comes to mind is the physical illnesses of a mental patient. It seems to me that one doesn't believe they exist.*
>
> *They may have decades of experience about the person with illness. One would expect that cooperation would be more beneficial than detrimental.*
>
> *For a long time, I tried to keep my husband to stay even in outpatient care, go to doctor's appointments and take his medicine.*
>
> *Even if the denial of information is standing, one should listen to relatives.*

Discussion

Principal Findings

The goal of this study is 2-fold. We first evaluated the feasibility of the web-based course with caregivers who participated in a psychoeducation course by analyzing engagement and feedback on the course. Second, participants' experiences of being a caregiver to a person with SSD were portrayed in writings produced during the web-based course. These 2 goals complemented each other by providing valuable information to both caregivers and professionals. It has already been shown that psychoeducational interventions for family caregivers of people with psychosis are effective [20]. It is therefore important to consider how web-based methods can be used to support family members, especially those who lack daily support [52]. At the same time, web-based courses could provide insights into caregivers' daily concerns and how support from professionals could more efficiently meet the individual needs of caregivers. It has been shown that web-based methods are sometimes preferable to face-to-face methods, as some caregivers are less likely to share their inner thoughts in face-to-face counseling sessions [53,54].

Out of 34 caregivers providing informed consent, 30 caregivers started the web-based course. Engagement with the modules decreased as the course proceeded. The number of log-ins similarly decreased in a study by Rotondi et al [26]. This may be because caregivers lack the strength to continue [2,3]. Some web-based programs with different target groups have shown a similar decrease in log-ins and an increase in withdrawals during the program [55]. In our study, of the 30 caregivers, 60% (18/30) completed all the 6 modules, which is a fairly good result compared with similar studies [28].

The question of how feasible web-based psychoeducation courses are for caregivers still remains if 53% (9/17) of the participants are unsure of its usability. Therefore, we need to discuss what might have been the reason for the lower ratings in our study. First, it is possible that, considering participants' extensive experience in caregiving, the course content did not offer new or valuable information for them, although the content of the course seemed to be interesting to the participants. Therefore, this kind of course might be more topical for people new to their role as a family caregiver. It has already been recommended that low-intensity interventions could be used as a first step of service for illnesses of milder severity before stepping up to treatments with higher intensities if needed [56].

Second, we offered opportunities for peer support via a web-based course platform. However, only a limited number of conversations between caregivers can be identified. Caregivers in our study may have been less active users of web-based discussion forums, which could have limited their willingness to have discussions via the internet. Furthermore, some participants may have been worried about the privacy issues of web-based services [57]. Third, our professional tutors offered individual support to each participant and actively contacted them if any worries could be recognized based on their writings. Caregivers might not have been used to remote support if they were familiar with face-to-face contact with professionals. On the other hand, acceptance of web-based interventions has been found to be high even if contact with professionals is limited [58]. In general, support provided by professionals seems to increase attrition [59,60]. Therefore, in practice, the preferences for psychoeducation courses and the presence of tutors or other professionals should be clearly indicated by the caregivers before the intervention. During these exceptional times in health services, we need to be able to provide web-based services effectively to those who are not able to travel to seek help or are not motivated to use web-based services.

Fourth, we may also ask whether these web-based courses should be targeted only to participants who are fully engaged and facilitated to use web-based methods or if more individual support should be offered to participants with a lack of engagement and skills. On the basis of a previous study, available technical support is important when participants lack technical skills [61]. The caregivers in our study had the possibility of asking questions via the internet [62]. Chat functions, video meetings, and video-recorded instructions could also be added to the course platform to help caregivers with any technical problems. Although these solutions may help technically oriented participants, they may not be useful to those who are less technically literate.

Fifth, the findings of this study indicate that a successful web-based psychoeducation intervention for caregivers demands that participants are supported during the course so that they will likely be engaged in the entire course. Ensuring that participants are motivated, have the necessary technology skills, and get help when needed is crucial. Engagement in the course could be supported by adding professional guidance and discussing the timing of the course and how much time should be allotted for each task of the course. On the other hand, keeping in mind refusals and dropout rates, we need to consider whom web-based psychoeducation courses should

target—should only those who are familiar with web-based methods be the focus, or should everyone, even those with fewer skills in digital technology, be targeted? To answer this fundamental question, more empirical studies are still needed. In any case, researchers should be clearly aware of the challenges of tailoring the web-based method to the needs and requirements of the target group.

Caregivers reflected on their own experiences of being a family member of a person with SSD. Their writings show that caregivers have problems with their own physical and mental well-being, which is similar to earlier findings [9-11,63]. Compared with the results of the study by Pollio et al [5], who identified 355 problems caregivers faced in managing mental illness in their relatives, our study focused on positive methods and how to manage daily life. We found that, in addition to burdening experiences, caregivers had versatile methods for supporting their own well-being and they received support from people around them. This is contrary to previous studies, in which caregivers of persons with schizophrenia lacked social support [64]. This is a positive result, as social support is associated with a lower level of burden [65]. Caregivers also described their relationship with the person with SSD they cared for, including both good and challenging experiences. Caregivers were able to modify their communication, social relationships, and daily activities based on the patient's mood and mental status. This flexible communication style would also be welcome in interactions with professionals.

Indeed, cooperation with professionals was seen as challenging, especially if they had different views of the patient's needs in inpatient care. As found in earlier studies [66,67], caregivers had contradictory feelings when a person with SSD was admitted to a hospital. They were relieved that the person was getting help; however, they also felt that they did not receive information and that they were ignored as a resource in care. The legislation in Finland [68] determines that information contained in patient documents is confidential, and professionals are not allowed to give out information to anyone about a patient's care without written consent from the patient. In general, confidentiality is highly respected in health care services. However, it has been found to be problematic in psychiatric services [66]. Even so, it is still possible for health care professionals to listen to the worries of caregivers, give them general information about treatment, and show empathy without breaking any confidentiality rules in patient care. Web-based psychoeducation programs could serve as a tool for understanding caregivers' situations and offering support and knowledge in a neutral way.

Limitations and Strengths

Our study has a number of limitations. First, because of the small sample size, the results cannot be generalized to a wider context. Second, most participants in our study were women aged around 60 years [24,25,27,28] who had been caregivers for quite a long time [28]. This means that most of them were *experienced* caregivers with long histories of being support persons for someone close to them with SSD. It may be that our target group would have benefited more from the involvement of a therapist in the support process, and therefore, our intervention might not have been properly suited to our target audience.

Third, we used the convenience sampling method, and potential participants were invited to participate in the study in the local associations. As the recruitment of this study was realized via the local family associations, caregivers in our study might have had more social support than caregivers generally have, which might have biased the results. Fourth, not all caregivers gave their feedback about the web-based course, although the option to do so was offered in quantitative and qualitative formats. It is also possible that the most critical caregivers dropped out in the early study phase, and therefore, the feedback was biased. We could have also used validated scales to collect feedback data [69], and the results could have been used to improve the course and its content in the future. Fifth, the length of each module could have been longer to allow more time to complete all the tasks. This might have decreased the dropout rate during the course. Sixth, the individual writing tasks were guided with questions, which may have focused on the content of the writings. Finally, some caregivers had technical problems at the beginning of the course, which may have discouraged their participation later in the course.

One strength of this study is that the data collection method was innovative. We used web-based methods for data collection, which have been found to be particularly useful in conducting a qualitative study with vulnerable study groups [70] and with stigmatized topics [4,71]. Typically, caregivers' experiences of dealing with mental health disorders have often been studied using quantitative methods and questionnaires (eg, Zarit Burden Interview [72] and Camberwell Family Interview [73]). Qualitative methods using individual interviews [74], focus groups [75], and a combination of these [76] have also been used. With caregivers' writings, we were given access to caregivers' uninhibited thoughts and feelings based on their stories, experiences, and insight, instead of extra pressure caused by face-to-face interviews with researchers.

Although not conclusive, our descriptive, small-scale study can contribute to improving the feasibility of web-based psychosocial interventions. In addition, having regular meetings with the local family associations allowed us to develop good relationships with them and deepen our collaboration with caregivers. Flexible tailoring and support, good monitoring, and good topics should allow us to continue our studies with larger sample sizes in the future.

Conclusions

Our study relates to the feasibility of a web-based psychoeducation course and the exploration of experiences of being a family caregiver. These focuses complement each other in that the study offers valuable information about usability and insights into the lives of caregivers. On the basis of the results, we can conclude that web-based psychoeducation may not be an optimal method for all caregivers of persons with SSD. However, it might be suitable for those who are quite new in their situation and those familiar with technological devices and social media platforms. To be successful, the use of web-based methods requires detailed planning related to content, practical arrangements, and the scheduling of the intervention. Carefully

matching the target group and the web-based methods is especially key in meeting the participants' needs and expectations. Effort should also be put into course participant selection. In addition, we can confirm that web-based courses are a usable method for obtaining rich, informative, qualitative data from writings that can increase professionals' understanding of caregivers' experiences and current needs. The next step is to run a larger study with sufficient power to obtain more rigorous information on the topic. We can conclude that although caregivers have versatile ways to improve their own well-being, they still need regular support and cooperation with health care professionals.

Acknowledgments

This study was funded by the University of Turku. The authors thank FinFami for their cooperation and the caregivers for their participation in the study. The authors would like to thank the course coordinator, VP, from the University of Turku for her dedicated work in running the web-based course.

Authors' Contributions

MV initiated the study. MV, MA, and HH designed the study. HH and MA were responsible for acquiring the data. AL was responsible for analyzing the data and drafting the manuscript. AL, MV, and MA wrote the manuscript, and HH participated in manuscript revision. All authors approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Number of log-ins among caregivers (N=30) for each module.
[PNG File , 24 KB - jmir_v23i4e25480_app1.png]

Multimedia Appendix 2
Number of caregivers who visited and completed each module.
[PNG File , 42 KB - jmir_v23i4e25480_app2.png]

References

1. International Statistical Classification of Diseases and Related Health Problems, Tenth Revision. World Health Organization. URL: https://icd.who.int/browse10/2016/en [accessed 2021-03-30]
2. Yu Y, Liu Z, Tang B, Zhao M, Liu X, Xiao S. Reported family burden of schizophrenia patients in rural China. PLoS One 2017;12(6):e0179425 [FREE Full text] [doi: 10.1371/journal.pone.0179425] [Medline: 28628657]
3. Shamsaei F, Cheraghi F, Bashirian S. Burden on Family Caregivers Caring for Patients with Schizophrenia. Iran J Psychiatry 2015 Sep;10(4):239-245 [FREE Full text] [Medline: 27006669]
4. Velligan DI, Brain C, Bouérat Duvold L, Agid O. Caregiver Burdens Associated With Treatment-Resistant Schizophrenia: A Quantitative Caregiver Survey of Experiences, Attitudes, and Perceptions. Front Psychiatry 2019;10:584 [FREE Full text] [doi: 10.3389/fpsyt.2019.00584] [Medline: 31551821]
5. Pollio DE, North CS, Foster DA. Content and curriculum in psychoeducation groups for families of persons with severe mental illness. Psychiatr Serv 1998 Jul;49(6):816-822. [doi: 10.1176/ps.49.6.816] [Medline: 9634164]
6. Thunyadee C, Sitthimongkol Y, Sangon S, Chai-Aroon T, Hegadoren KM. Predictors of depressive symptoms and physical health in caregivers of individuals with schizophrenia. Nurs Health Sci 2015 Dec;17(4):412-419. [doi: 10.1111/nhs.12205] [Medline: 26081195]
7. Opoku-Boateng YN, Kretchy IA, Aryeetey GC, Dwomoh D, Decker S, Agyemang SA, et al. Economic cost and quality of life of family caregivers of schizophrenic patients attending psychiatric hospitals in Ghana. BMC Health Serv Res 2017 Dec 04;17(Suppl 2):697 [FREE Full text] [doi: 10.1186/s12913-017-2642-0] [Medline: 29219074]
8. Awad AG, Voruganti LNP. The burden of schizophrenia on caregivers: a review. Pharmacoeconomics 2008;26(2):149-162. [Medline: 18198934]
9. Lerner D, Chang H, Rogers WH, Benson C, Lyson MC, Dixon LB. Psychological Distress Among Caregivers of Individuals With a Diagnosis of Schizophrenia or Schizoaffective Disorder. Psychiatr Serv 2018 Feb 01;69(2):169-178. [doi: 10.1176/appi.ps.201600422] [Medline: 28967321]
10. Hegde A, Chakrabarti S, Grover S. Caregiver distress in schizophrenia and mood disorders: the role of illness-related stressors and caregiver-related factors. Nord J Psychiatry 2019 Jan;73(1):64-72. [doi: 10.1080/08039488.2018.1561945] [Medline: 30638102]

11. Sharma R, Sharma SC, Pradhan SN. Assessing Caregiver Burden in Caregivers of Patients with Schizophrenia and Bipolar Affective Disorder in Kathmandu Medical College. J Nepal Health Res Counc 2018 Jan 01;15(3):258-263. [doi: 10.3126/jnhrc.v15i3.18851] [Medline: 29353899]
12. Souza ALR, Guimarães RA, de Araújo Vilela D, de Assis RM, de Almeida Cavalcante Oliveira LM, Souza MR, et al. Factors associated with the burden of family caregivers of patients with mental disorders: a cross-sectional study. BMC Psychiatry 2017 Oct 25;17(1):353 [FREE Full text] [doi: 10.1186/s12888-017-1501-1] [Medline: 29070012]
13. Shiraishi N, Reilly J. Positive and negative impacts of schizophrenia on family caregivers: a systematic review and qualitative meta-summary. Soc Psychiatry Psychiatr Epidemiol 2019 Mar;54(3):277-290. [doi: 10.1007/s00127-018-1617-8] [Medline: 30349961]
14. Doval N, Sharma E, Agarwal M, Tripathi A, Nischal A. Experience of Caregiving and Coping in Caregivers of Schizophrenia. Clin Schizophr Relat Psychoses 2018;12(3):113-120B. [doi: 10.3371/csrp.DOSH.123015] [Medline: 26780599]
15. Mackay C, Pakenham KI. A Stress and Coping Model of Adjustment to Caring for an Adult with Mental Illness. Community Ment Health J 2011 Jun 30;48(4):450-462. [doi: 10.1007/s10597-011-9435-4]
16. Xia J, Merinder LB, Belgamwar MR. Psychoeducation for schizophrenia. Cochrane Database Syst Rev 2011 Jun 15(6):CD002831 [FREE Full text] [doi: 10.1002/14651858.CD002831.pub2] [Medline: 21678337]
17. The National Institute for HealthCare Excellence. Psychosis and Schizophrenia in Adults: Prevention and Management. 2014. URL: https://www.nice.org.uk/guidance/cg178/ [accessed 2021-03-30]
18. Working group appointed by the Finnish Medical Society Duodecim, the Finnish Psychiatric Association. Schizophrenia. Current Care Guidelines. 2020. URL: https://www.kaypahoito.fi/en/ccs00038 [accessed 2021-03-30]
19. Välimäki M, Hätönen H, Lahti M, Kuosmanen L, Adams CE. Information and communication technology in patient education and support for people with schizophrenia. Cochrane Database Syst Rev 2012 Oct;10:CD007198. [doi: 10.1002/14651858.CD007198.pub2] [Medline: 23076932]
20. Sin J, Gillard S, Spain D, Cornelius V, Chen T, Henderson C. Effectiveness of psychoeducational interventions for family carers of people with psychosis: A systematic review and meta-analysis. Clin Psychol Rev 2017 Aug;56:13-24 [FREE Full text] [doi: 10.1016/j.cpr.2017.05.002] [Medline: 28578249]
21. Fiorillo A, Del Vecchio V, Luciano M, Sampogna G, Sbordone D, Catapano F, et al. Feasibility of a psychoeducational family intervention for people with bipolar I disorder and their relatives: Results from an Italian real-world multicentre study. J Affect Disord 2016 Jan 15;190:657-662. [doi: 10.1016/j.jad.2015.10.060] [Medline: 26590513]
22. World Health Organization. WHO Guideline: recommendations on digital interventions for health system strengthening. 2019. URL: https://www.who.int/reproductivehealth/publications/digital-interventions-health-system-strengthening/en/ [accessed 2021-03-30]
23. Spencer L, Potterton R, Allen K, Musiat P, Schmidt U. Internet-Based Interventions for Carers of Individuals With Psychiatric Disorders, Neurological Disorders, or Brain Injuries: Systematic Review. J Med Internet Res 2019 Jul 09;21(7):e10876 [FREE Full text] [doi: 10.2196/10876] [Medline: 31290399]
24. Chan SKW, Tse S, Sit HLT, Hui CLM, Lee EHM, Chang WC, et al. Web-Based Psychoeducation Program for Caregivers of First-Episode of Psychosis: An Experience of Chinese Population in Hong Kong. Front Psychol 2016;7:2006 [FREE Full text] [doi: 10.3389/fpsyg.2016.02006] [Medline: 28082935]
25. Sin J, Woodham LA, Henderson C, Williams E, Sesé Hernández A, Gillard S. Usability evaluation of an eHealth intervention for family carers of individuals affected by psychosis: A mixed-method study. Digit Health 2019;5:2055207619871148 [FREE Full text] [doi: 10.1177/2055207619871148] [Medline: 31489205]
26. Rotondi AJ, Haas GL, Anderson CM, Newhill CE, Spring MB, Ganguli R, et al. A Clinical Trial to Test the Feasibility of a Telehealth Psychoeducational Intervention for Persons With Schizophrenia and Their Families: Intervention and 3-Month Findings. Rehabil Psychol 2005 Nov;50(4):325-336 [FREE Full text] [doi: 10.1037/0090-5550.50.4.325] [Medline: 26321774]
27. Rotondi AJ, Anderson CM, Haas GL, Eack SM, Spring MB, Ganguli R, et al. Web-based psychoeducational intervention for persons with schizophrenia and their supporters: one-year outcomes. Psychiatr Serv 2010 Nov;61(11):1099-1105. [doi: 10.1176/appi.ps.61.11.1099] [Medline: 21041348]
28. Glynn SM, Randolph ET, Garrick T, Lui A. A proof of concept trial of an online psychoeducational program for relatives of both veterans and civilians living with schizophrenia. Psychiatr Rehabil J 2010;33(4):278-287. [doi: 10.2975/33.4.2010.278.287] [Medline: 20374986]
29. MacDonell KW, Prinz RJ. A Review of Technology-Based Youth and Family-Focused Interventions. Clin Child Fam Psychol Rev 2017 Dec;20(2):185-200 [FREE Full text] [doi: 10.1007/s10567-016-0218-x] [Medline: 27787701]
30. Andersson G, Titov N. Advantages and limitations of Internet-based interventions for common mental disorders. World Psychiatry 2014 Feb;13(1):4-11 [FREE Full text] [doi: 10.1002/wps.20083] [Medline: 24497236]
31. Stjernswärd S, Hansson L. A web-based supportive intervention for families living with depression: content analysis and formative evaluation. JMIR Res Protoc 2014 Feb 14;3(1):e8 [FREE Full text] [doi: 10.2196/resprot.3051] [Medline: 24550185]

32. Family Caregiver Alliance. Report from A National Consensus Development Conference. National Consensus Development ConferenceCaregiver Assessment: Principles, Guidelines and Strategies for Change. Volume 1. 2006. URL: https://www.caregiver.org/uploads/legacy/pdfs/v1_consensus.pdf [accessed 2021-03-30]
33. Creswell J. Research Design: Qualitative, Quantitative and Mixed Methods Approaches. Thousand Oaks, California: SAGE Publications; 2014.
34. Kettles AM, Creswell JW, Zhang W. Mixed methods research in mental health nursing. J Psychiatr Ment Health Nurs 2011 Aug;18(6):535-542. [doi: 10.1111/j.1365-2850.2011.01701.x] [Medline: 21749560]
35. O'Cathain A, Murphy E, Nicholl J. The quality of mixed methods studies in health services research. J Health Serv Res Policy 2008 Apr;13(2):92-98. [doi: 10.1258/jhsrp.2007.007074] [Medline: 18416914]
36. Finnish Central Association of Families of People with mental illness. URL: https://finfami.fi/in-english [accessed 2021-03-30]
37. Finnish National Board on Research Integrity. The ethical principles of research with human participants and ethical review in the human sciences in Finland. 2019. URL: https://tenk.fi/sites/tenk.fi/files/Ihmistieteiden_eettisen_ennakkoarvioinnin_ohje_2019.pdf [accessed 2021-03-30]
38. Polit D, Beck C. Essentials of nursing research: Appraising evidence for nursing practice. 8th edition. Philadelphia: Wolters Kluwer Health /Lippincott Williams & Wilkins; 2014.
39. Ministry of Social Affairs and Health. Medical Research Act 488/1999. 1999. URL: https://www.finlex.fi/fi/laki/kaannokset/1999/en19990488.pdf [accessed 2021-03-30]
40. Council for International Organizations of Medical Sciences. International Ethical Guidelines for Health-related Research Involving Humans. 2016. URL: https://cioms.ch/wp-content/uploads/2017/01/WEB-CIOMS-EthicalGuidelines.pdf [accessed 2021-03-30]
41. Wong DK, Cheung M. Online Health Information Seeking and eHealth Literacy Among Patients Attending a Primary Care Clinic in Hong Kong: A Cross-Sectional Survey. J Med Internet Res 2019 Mar 27;21(3):e10831 [FREE Full text] [doi: 10.2196/10831] [Medline: 30916666]
42. Valimaki M, Anttila M, Hatonen H, Koivunen M, Jakobsson T, Pitkanen A, et al. Design and development process of patient-centered computer-based support system for patients with schizophrenia spectrum psychosis. Inform Health Soc Care 2008 Jun;33(2):113-123. [doi: 10.1080/17538150802127207] [Medline: 18604755]
43. Laine A, Anttila M, Välimäki M. Modification of an internet-based patient education program for adults with schizophrenia spectrum disorder to suit adolescents with psychosis. Inform Health Soc Care 2015 Aug 21:1-17. [doi: 10.3109/17538157.2015.1008487] [Medline: 25710517]
44. Johansson K, Leino-Kilpi H, Salanterä S, Lehtikunnas T, Ahonen P, Elomaa L, et al. Need for change in patient education: a Finnish survey from the patient's perspective. Patient Educ Couns 2003 Nov;51(3):239-245. [doi: 10.1016/s0738-3991(02)00223-9] [Medline: 14630380]
45. Valimaki M, Anttila M, Hatonen H, Koivunen M, Jakobsson T, Pitkanen A, et al. Design and development process of patient-centered computer-based support system for patients with schizophrenia spectrum psychosis. Inform Health Soc Care 2008 Jun;33(2):113-123. [doi: 10.1080/17538150802127207] [Medline: 18604755]
46. Laine A, Anttila M, Välimäki M. Modification of an internet-based patient education program for adults with schizophrenia spectrum disorder to suit adolescents with psychosis. Inform Health Soc Care 2015 Aug 21:1-17. [doi: 10.3109/17538157.2015.1008487] [Medline: 25710517]
47. Laine A, Välimäki M, Pekurinen V, Löyttyniemi E, Marttunen M, Anttila M. Feasibility, Acceptability, and Preliminary Impacts of Web-Based Patient Education on Patients With Schizophrenia Spectrum Disorder: Quasi-Experimental Cluster Study. J Med Internet Res 2019 Oct 17;21(10):e13073 [FREE Full text] [doi: 10.2196/13073] [Medline: 31625952]
48. Goldberg L, Lide B, Lowry S, Massett HA, O'Connell T, Preece J, et al. Usability and accessibility in consumer health informatics current trends and future challenges. Am J Prev Med 2011 May;40(5 Suppl 2):S187-S197. [doi: 10.1016/j.amepre.2011.01.009] [Medline: 21521594]
49. Sullivan GM, Sargeant J. Qualities of qualitative research: part I. J Grad Med Educ 2011 Dec;3(4):449-452 [FREE Full text] [doi: 10.4300/JGME-D-11-00221.1] [Medline: 23205188]
50. Graneheim UH, Lundman B. Qualitative content analysis in nursing research: concepts, procedures and measures to achieve trustworthiness. Nurse Educ Today 2004 Feb;24(2):105-112. [doi: 10.1016/j.nedt.2003.10.001] [Medline: 14769454]
51. Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Research in Psychology 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
52. Yesufu-Udechuku A, Harrison B, Mayo-Wilson E, Young N, Woodhams P, Shiers D, et al. Interventions to improve the experience of caring for people with severe mental illness: systematic review and meta-analysis. Br J Psychiatry 2015 Apr;206(4):268-274 [FREE Full text] [doi: 10.1192/bjp.bp.114.147561] [Medline: 25833867]
53. Wong KP, Bonn G, Tam CL, Wong CP. Preferences for Online and/or Face-to-Face Counseling among University Students in Malaysia. Front Psychol 2018;9:64 [FREE Full text] [doi: 10.3389/fpsyg.2018.00064] [Medline: 29445352]
54. Renn BN, Hoeft TJ, Lee HS, Bauer AM, Areán PA. Preference for in-person psychotherapy versus digital psychotherapy options for depression: survey of adults in the U.S. NPJ Digit Med 2019;2:6 [FREE Full text] [doi: 10.1038/s41746-019-0077-1] [Medline: 31304356]

55. Välimäki M, Anttila K, Anttila M, Lahti M. Web-Based Interventions Supporting Adolescents and Young People With Depressive Symptoms: Systematic Review and Meta-Analysis. JMIR Mhealth Uhealth 2017 Dec 08;5(12):e180 [FREE Full text] [doi: 10.2196/mhealth.8624] [Medline: 29222079]
56. van SA, Hill J, Richards DA, Cuijpers P. Stepped care treatment delivery for depression: a systematic review and meta-analysis. Psychol Med 2014 Mar 26:1-16. [doi: 10.1017/S0033291714000701] [Medline: 25065653]
57. Gulliver A, Bennett K, Bennett A, Farrer LM, Reynolds J, Griffiths KM. Privacy Issues in the Development of a Virtual Mental Health Clinic for University Students: A Qualitative Study. JMIR Ment Health 2015;2(1):e9 [FREE Full text] [doi: 10.2196/mental.4294] [Medline: 26543915]
58. Andrews G, Cuijpers P, Craske MG, McEvoy P, Titov N. Computer therapy for the anxiety and depressive disorders is effective, acceptable and practical health care: a meta-analysis. PLoS One 2010;5(10):e13196 [FREE Full text] [doi: 10.1371/journal.pone.0013196] [Medline: 20967242]
59. Gershkovich M, Herbert JD, Forman EM, Schumacher LM, Fischer LE. Internet-Delivered Acceptance-Based Cognitive-Behavioral Intervention for Social Anxiety Disorder With and Without Therapist Support: A Randomized Trial. Behav Modif 2017 Sep;41(5):583-608. [doi: 10.1177/0145445517694457] [Medline: 28776431]
60. Lawn S, Zhi X, Morello A. An integrative review of e-learning in the delivery of self-management support training for health professionals. BMC Med Educ 2017 Oct 10;17(1):183 [FREE Full text] [doi: 10.1186/s12909-017-1022-0] [Medline: 29017521]
61. Roddy C, Amiet DL, Chung J, Holt C, Shaw L, McKenzie S, et al. Applying Best Practice Online Learning, Teaching, and Support to Intensive Online Environments: An Integrative Review. Front. Educ 2017 Nov 21;2. [doi: 10.3389/feduc.2017.00059]
62. Sin J. Focus group study of siblings of individuals with psychosis: views on designing an online psychoeducational resource. J Psychosoc Nurs Ment Health Serv 2013 Jun;51(6):28-36. [doi: 10.3928/02793695-20130404-02] [Medline: 23586363]
63. Wan K, Wong MMC. Stress and burden faced by family caregivers of people with schizophrenia and early psychosis in Hong Kong. Intern Med J 2019 Mar;49 Suppl 1:9-15. [doi: 10.1111/imj.14166] [Medline: 30815980]
64. Chen L, Zhao Y, Tang J, Jin G, Liu Y, Zhao X, et al. The burden, support and needs of primary family caregivers of people experiencing schizophrenia in Beijing communities: a qualitative study. BMC Psychiatry 2019 Feb 20;19(1):75 [FREE Full text] [doi: 10.1186/s12888-019-2052-4] [Medline: 30786852]
65. Kate N, Grover S, Kulhara P, Nehra R. Relationship of caregiver burden with coping strategies, social support, psychological morbidity, and quality of life in the caregivers of schizophrenia. Asian J Psychiatr 2013 Oct;6(5):380-388. [doi: 10.1016/j.ajp.2013.03.014] [Medline: 24011684]
66. Jankovic J, Yeeles K, Katsakou C, Amos T, Morriss R, Rose D, et al. Family caregivers' experiences of involuntary psychiatric hospital admissions of their relatives--a qualitative study. PLoS One 2011;6(10):e25425 [FREE Full text] [doi: 10.1371/journal.pone.0025425] [Medline: 22022393]
67. Stuart R, Akther SF, Machin K, Persaud K, Simpson A, Johnson S, et al. Carers' experiences of involuntary admission under mental health legislation: systematic review and qualitative meta-synthesis. BJPsych Open 2020 Feb 11;6(2):e19 [FREE Full text] [doi: 10.1192/bjo.2019.101] [Medline: 32043435]
68. Ministry of Social Affairs and Health. Act on the Status and Rights of Patients 785/1992. 1992. URL: https://www.finlex.fi/en/laki/kaannokset/1992/en19920785_20120690.pdf [accessed 2021-03-30]
69. Hubley S, Lynch SB, Schneck C, Thomas M, Shore J. Review of key telepsychiatry outcomes. World J Psychiatry 2016 Jun 22;6(2):269-282 [FREE Full text] [doi: 10.5498/wjp.v6.i2.269] [Medline: 27354970]
70. Neville S, Adams J, Cook C. Using internet-based approaches to collect qualitative data from vulnerable groups: reflections from the field. Contemp Nurse 2016 Dec;52(6):657-668. [doi: 10.1080/10376178.2015.1095056] [Medline: 26394073]
71. Rössler W. The stigma of mental disorders: A millennia-long history of social exclusion and prejudices. EMBO Rep 2016 Sep;17(9):1250-1253 [FREE Full text] [doi: 10.15252/embr.201643041] [Medline: 27470237]
72. Zarit SH, Reever KE, Bach-Peterson J. Relatives of the impaired elderly: correlates of feelings of burden. Gerontologist 1980 Dec;20(6):649-655. [Medline: 7203086]
73. Vaughn CE, Leff JP. The influence of family and social factors on the course of psychiatric illness. A comparison of schizophrenic and depressed neurotic patients. Br J Psychiatry 1976 Aug;129:125-137. [doi: 10.1192/bjp.129.2.125] [Medline: 963348]
74. McCann TV, Bamberg J, McCann F. Family carers' experience of caring for an older parent with severe and persistent mental illness. Int J Ment Health Nurs 2015 Jun;24(3):203-212 [FREE Full text] [doi: 10.1111/inm.12135] [Medline: 25963281]
75. Pusey-Murray A, Miller P. 'I need help': caregivers' experiences of caring for their relatives with mental illness in Jamaica. Ment Health Fam Med 2013 Jun;10(2):113-121. [Medline: 24427177]
76. Ae-Ngibise KA, Doku VCK, Asante KP, Owusu-Agyei S. The experience of caregivers of people living with serious mental disorders: a study from rural Ghana. Glob Health Action 2015;8:26957 [FREE Full text] [doi: 10.3402/gha.v8.26957] [Medline: 25967587]

Abbreviations

FinFami: Finnish Central Association of Families With People With Mental Illness
ICD-10: 10th version of International Classification of Disorders
SSD: schizophrenia spectrum disorder

Edited by R Kukafka; submitted 03.11.20; peer-reviewed by K Machin, S Chakrabarti; comments to author 09.12.20; revised version received 23.02.21; accepted 16.03.21; published 23.04.21.

Please cite as:
Laine A, Anttila M, Hirvonen H, Välimäki M
Feasibility of a Web-Based Psychoeducation Course and Experiences of Caregivers Living With a Person With Schizophrenia Spectrum Disorder: Mixed Methods Study
J Med Internet Res 2021;23(4):e25480
URL: https://www.jmir.org/2021/4/e25480
doi:10.2196/25480
PMID:33890862

Original Paper

An Internet-Based Intervention for Cardiovascular Disease Management Integrated With Primary Care Electronic Health Records: Mixed Methods Evaluation of Implementation Fidelity and User Engagement

Genevieve Coorey[1,2], BAppSc, MA, MSc, PhD; David Peiris[1,3], MBBS(Hons), MIPH, PhD; Anish Scaria[1], MSc; John Mulley[1], BHlthSc; Lis Neubeck[4,5], BA(Hons), PhD; Nashid Hafiz[2], MBBS, MIPH; Julie Redfern[2], BSc, BAppSc(Physio Hons 1), PhD

[1]The George Institute for Global Health, Sydney, Australia, Sydney, Australia

[2]School of Health Sciences, Faculty of Medicine and Health, University of Sydney, Sydney, Australia

[3]Faculty of Medicine, The University of New South Wales, Sydney, Australia

[4]School of Health and Social Care, Edinburgh Napier University, Edinburgh, United Kingdom

[5]Susan Wakil School for Nursing and Midwifery, The University of Sydney, Sydney, Australia

Corresponding Author:
Genevieve Coorey, BAppSc, MA, MSc, PhD
The George Institute for Global Health, Sydney, Australia
1 King St, Newtown
Sydney, 2042
Australia
Phone: 61 80524644
Email: gcoorey@georgeinstitute.org.au

Abstract

Background: Growing evidence supports the benefits of eHealth interventions to increase patient engagement and improve outcomes for a range of conditions. However, ineffective program delivery and usage attrition limit exposure to these interventions and may reduce their effectiveness.

Objective: This study aims to evaluate the delivery fidelity of an eHealth intervention, describe use patterns, compare outcomes between low and high users, and identify mediating factors on intervention delivery and receipt.

Methods: This is a mixed methods study of an internet-based intervention being evaluated for effectiveness in a randomized controlled trial (RCT). The intervention comprised medication and cardiovascular disease (CVD) risk data uploaded from the primary care electronic health record (EHR); interactive, personalized CVD risk score estimation; goal setting and self-monitoring; an interactive social forum; and optional receipt of heart health messages. Fidelity was assessed over 12 months. Trial outcomes were compared between low and high users. Data sources included program delivery records, web log data, trial data, and thematic analysis of communication records.

Results: Most participants in the intervention group (451/486, 93%) had an initial training session conducted by telephone (413/447, 92.4% of participants trained), with a mean duration of 44 minutes (range 10-90 minutes). Staff conducted 98.45% (1776/1804) of the expected follow-ups, mostly by telephone or email. Of the 451 participants who commenced log-ins, 46.8% (211) were categorized as low users (defined as at least one log-in in 3 or fewer months of follow-up), 40.4% (182) were categorized as high users (at least one log-in in more than 3 months of follow-up), and 12.8% (58) were nonadopters (no log-ins after their training session). The mean log-in frequency was 3-4 per month in ongoing users. There was no significant difference between the groups in the primary trial outcome of adherence to guideline-recommended medications (P=.44). In unadjusted analyses, high users had significantly greater eHealth literacy scores (P=.003) and were more likely to meet recommended weekly targets for fruit (P=.03) and fish (P=.004) servings; however, the adjusted findings were not significant. Interactive screen use was highest for goal tracking and lowest for the chat forum. Screens with EHR-derived data held only an early interest for most users. Fidelity measures (reach, content, dose delivered, and dose received) were influenced by the facilitation strategies used by staff, *invisible* qualities of staff-participant communication, and participants' responsiveness to intervention attributes.

Conclusions: A multifeature internet-based intervention was delivered with high fidelity to the RCT protocol and was regularly used by 40.4% (182/451) of users over 12 months. Higher log-in frequency as an indicator of greater intervention exposure was not associated with statistically significant improvements in eHealth literacy scores, lifestyle changes, or clinical outcomes. Attributes of the intervention and individualized support influenced initial and ongoing use.

(J Med Internet Res 2021;23(4):e25333) doi:10.2196/25333

KEYWORDS

eHealth; electronic health record; web-based intervention; implementation fidelity; user engagement; mixed methods; cardiovascular disease; primary health care; mobile phone

Introduction

Background

Cardiovascular disease (CVD) accounts for approximately one-third of deaths globally [1], and effective prevention remains the cornerstone of efforts to reduce disease morbidity and mortality. Emphasis is on improving modifiable lifestyle factors, such as tobacco smoking, obesity, unhealthy diet, low physical activity, and alcohol consumption, and controlling high-risk conditions such as hyperglycemia, hypertension, and hyperlipidemia [2]. Technology-based approaches have the potential to support self-care, self-monitoring, and adoption of CVD prevention recommendations and more active interactions between patients and health care providers for managing chronic conditions [3-5]. Evidence for the effectiveness of internet-based strategies for CVD risk factor control is promising [6,7] but considered inconclusive [8].

Studies of internet-based interventions to support lifestyle behavior modification identify attrition of web use among the key reasons for suboptimal exposure or participation, which in turn influences uptake, adherence, and potential effectiveness [9-12]. The reasons are likely multifactorial: interventions targeting behavior change often consist of many interactive components and functions for voluntary use by recipients in home-based or nonclinical settings—attributes of intervention complexity [13]. Engagement with such interventions is characterized not only by subjective user qualities, such as interest, motivation, and sensory and intellectual satisfaction [14], but also by how these intersect with website use or behavior [15]. However, self-reported use of internet-based health resources generally does not objectively describe actual website interactions over time. This presents a gap in the fuller understanding of the activities undertaken at such websites [16]. Information about user activities has important implications for design and functionality decisions that can influence initial adoption and ongoing use [17].

Within a process evaluation, it is therefore useful to examine the fidelity of intervention delivery and receipt in terms of actual use behavior and a fuller understanding of the content that most influences user interaction. Intervention fidelity encapsulates both the implementation and delivery by providers and the treatment receipt and interaction with the intervention by the intended target audience [18]. Both dimensions can influence future program adaptation and scale-up [19]. Other proposed benefits of measuring or evaluating fidelity include ensuring internal program validity [20] and identifying which intervention elements affected participant responsiveness or reaction [21]. It is further recommended that these analyses precede trial outcome assessments to avoid bias in their interpretation [13,22].

Objectives

Using data from a cohort of participants at moderate to high risk for CVD within a randomized controlled trial (RCT), this study aims to (1) evaluate the delivery fidelity of a consumer-focused web application with integration of data from primary health care electronic health records (EHRs), (2) describe objectively measured patterns of website usage and compare RCT outcomes between low and high users, and (3) identify mediating factors on delivery and receipt of the intervention that may explain the engagement and interaction patterns observed.

Methods

Design

A mixed methods study was nested within a process evaluation of the Consumer Navigation of Electronic Cardiovascular Tools (CONNECT) RCT (Australian New Zealand Clinical Trials Registry ID: 12613000715774). Protocols detailing the RCT and process evaluation have been previously reported [23,24]. The process evaluation is multifactorial. The focus of this study was delivery fidelity; other aspects of the overall evaluation that have been previously reported [25,26] are not part of this study. Ethics approval was obtained from the University of Sydney (reference 2013/716) and the New South Wales Aboriginal Health and Medical Research Council (reference 959/13).

Participants

Adults with or at increased risk for CVD who had access to an internet-enabled device at least once per month and who could provide written informed consent were eligible to participate in the RCT. It was not a condition of enrollment that participants met any specific threshold of digital literacy or skill. Interest and willingness to use their device for study participation were required, but the extent of dependency on others to use their device for study purposes was not formally assessed at enrollment. Therefore, the digital support needs of individual participants were initially unknown to the staff who delivered the intervention. eHealth literacy scores were among the baseline data obtained from all those recruited but were not analyzed until completion of the RCT. Recruitment was conducted from 24 primary health care services in Sydney, Australia. Of the 934 participants who enrolled, 52.0% (486) were allocated to the intervention group and 48.0% (448) were allocated to the

control group. The control group received usual care from their primary health care provider; the intervention group received their usual care plus access to the eHealth intervention. This evaluation addresses the fidelity of intervention delivery in the latter group of participants.

Trial Outcomes

Briefly, the RCT tested the effectiveness of an eHealth intervention integrated with the primary care EHR for improving CVD risk factor control through better adherence to prescribed medications and lifestyle recommendations. The main trial results have been published elsewhere [27]. In summary, there was no significant difference between the control and intervention groups in the primary outcome of adherence to guideline-recommended medications, as defined by the proportion of days covered by guideline-recommended medications using pharmacy dispensing data (29.9% vs 32.8%; *P*=.48). There was little heterogeneity in the outcomes observed for prespecified subgroups, such as those with and without established CVD. However, the intervention was associated with improvements in 2 secondary outcomes: higher eHealth literacy scores (*P*<.001) and increased self-reported physical activity (87.0% vs 79.7%; *P*=.02).

Intervention

Participants in the intervention group received access to a purpose-built, multi-feature web application securely linked to data within their primary care EHR. The extensive, systematic process by which the application was co-designed and evaluated for usability in patients with CVD is detailed elsewhere [28]. The logic model linking program inputs to expected user uptake activities within the CONNECT RCT has been described previously [24]. Briefly, software to enable EHR data transfer (RecordConnect, Extensia Pty Ltd) was installed into the clinical software systems of participating primary care providers. After each medical encounter, changes or additions made within selected data fields of the EHR were uploaded to the consumer-facing website by the provider. Concurrently, participants could enter and track measurements in charts displaying trends, targets, and current results for personal biometric and pathology data related to vascular risk factors. Other personalized interactive features included absolute CVD risk score and heart age estimation, goal setting and tracking for healthy eating, physical activity, smoking cessation and emotional well-being, and updateable display of diagnoses and prescription medications with accompanying consumer information resources from national peak health bodies. Participants could also read and contribute to a closed chat forum page and could opt to receive semipersonalized heart health messages via email and/or text message format. Participants could receive an alert when their portal was updated from the EHR and could select their preferences for the heart health messages content from topics such as healthy lifestyle, goal reminders, and medication knowledge. The text message content was adapted from messages that our colleagues initially developed and tested for use with patients with CVD [29]. In this study of patients with similar demography and diagnoses to the earlier study, the multidisciplinary research team further reviewed and expanded the messages to ensure alignment with current Australian guidelines for primary and secondary prevention of CVD. The intervention was intended to take place in real life; it was home-based and without mandatory task completion, data entry, or counselor supervision. As a self-directed intervention, the frequency and extent of use were entirely controlled by the participants, although the study staff encouraged them to interact with and revisit the information and features.

Staff Support

Project staff consisting of registered nurses, dietitians, and pharmacists provided support for patients throughout the intervention period. Although these staff encouraged participants to interact with and revisit the information and features, they were not expected to formally counsel participants about risk factor management pertaining to their health circumstances.

During the initial one-to-one telephone-based or face-to-face training session, study staff provided participants with their *go-live* log-in information and orientation to the website features and navigation, and they answered initial technical and clinical questions. In terms of the quality of the intervention presented to participants, portals were checked before each participant's go-live session to ensure that data uploads from the EHR to the portal were current and that any identified screen errors were rectified before go-live. Thereafter, follow-up contact occurred at 4 scheduled time points (weeks 2, 6, 12, and 26) to answer questions and encourage website interaction and return visits. All scheduled and ad hoc contacts were documented on a standard form and included comments made by the participants and/or notes made by the staff. Participants who did not perform the initial training session with staff were sent self-directed materials; if they initiated log-ins, staff then made the required scheduled contacts.

Scheduled follow-up contact was by telephone unless the participant indicated a preference for email communication. Wherever possible, the staff member who conducted the go-live training conducted the scheduled follow-up with the participant and responded to their email and telephone communication. At each follow-up time point, the staff made at least two attempts to contact participants by telephone; if unsuccessful, an email was sent. To conduct the training sessions and scheduled follow-ups, standard operating procedures were developed, refined, and adopted throughout the RCT to optimize implementation consistency. A separate group of staff, blinded to the intervention allocation, conducted study outcome assessments and did not participate in the delivery of the intervention.

Data Sources

A total of five core measures of intervention delivery fidelity derived from concepts in digital health engagement [15] and guidance on process evaluation and implementation research, [30] were evaluated before knowledge of the RCT outcomes: reach, content fidelity, dose delivered, frequency and duration, and exposure (Table 1).

Table 1. Fidelity measures.

Fidelity measure	Description [15,30]	Data sources
Reach	Proportion of the intended target audience that participates in all or part of the intervention	Program delivery records
Content fidelity	The intervention content is delivered in the intended manner and quality	Program delivery records
Dose delivered	The amount of the intervention components that were provided to participants	Program delivery records
Frequency and duration	How long the intervention was implemented as intended in the trial design and how often participants made contact with the intervention	Program delivery records
Dose received or exposure	How much of the activities or components was read, viewed, or used for the intended duration	Web log files

Program delivery records regarding go-live training, and scheduled and ad hoc communication with participants, were reviewed. Participant factors affecting intervention delivery and exposure were gathered from textual data within the email and telephone records. These 2 fidelity measures closely relate to the 2 stages of a program's logic model: the intervention inputs and the intervention uptake [20].

Website use was central to the *dose* of intervention the participants received. Web log data were used to quantify use in terms of frequency and intensity over the 12 months of study follow-up. Go-live and subsequent session or log-in numbers and screens viewed per month were available for each participant. Session or page view duration and the number of daily interactions were not available. To further characterize exposure, participants were categorized as nonadopters, low (lower) users, or high (higher) users. The number of months in which a participant logged into the application at least once was counted during the 12-month follow-up period to categorize use. Nonadopters were defined as those who logged into the application only once. Low users were defined as those who logged into the application at least once per month in 3 or fewer months (eg, 12 log-ins in month 4 and 5 log-ins in month 7 of follow-up would be categorized as low use). High users were defined as those who logged into the application at least once per month in more than 3 months (eg, 12 log-ins in month 4, 5 log-ins in month 7, 3 log-ins in month 8, and 1 log-in in month 11 of follow-up would be categorized as high use). Hence, the definitions prioritized returning or ongoing log-in activity, often in nonconsecutive months, over the absolute number of log-ins. The log-in frequency data were skewed. Three months was chosen as the cutoff based on the median number of monthly log-ins over 12 months of follow-up. In the absence of an agreed standard for low versus high use of such interventions, the criteria used in this study resulted from an exploratory analysis of the log-in data at the completion of the RCT rather than from definitions set a priori. Optional email and/or SMS text messaging receipt was compared with website use to quantify the combination of passive and active content exposure.

Data Analysis

Descriptive statistics were used to describe and compare the format, the time required, and the content of all staff-participant contact episodes during the go-live and study follow-up periods, thus indicating resource needs for implementing the intervention content as planned. Textual data in the participant contact records were coded for similarities in ideas, and themes were identified inductively to describe the broader issues surrounding implementation needs and participant responsiveness [31]. The themes identified as influencing delivery and uptake were then merged with the original program logic model to indicate where they affect assumptions about program function and could be important implementation considerations for future similar interventions.

The demographic and clinical characteristics of the 3 user subgroups (nonadopters, low users, and high users) were compared using baseline data from the RCT database. The use patterns and characteristics of different types of user groups may be important for the future design of portal functions to optimize adoption. Subanalyses of RCT outcomes were also performed according to the 3 subgroups. For the subanalyses, the nonadopters and low users were combined into 1 group because of the small numbers in the former group. For the primary outcome subanalysis, adherence to guideline-recommended medication was calculated using the proportion of days covered from national pharmaceutical dispensing data. If a participant had a proportion of days covered of ≥80%, the participant was considered adherent to the guideline-recommended medication. Adherence (yes or no) was analyzed using a logistic regression model with treatment as an independent variable. Relative risk with 95% CIs was estimated by comparing high and low users. The primary subanalysis was adjusted for age, sex, and diabetes status.

In the secondary and tertiary outcome subanalyses, for categorical variables where a difference in proportions between high and low users at 12 months was calculated, a chi-square test of independence was used. If cell counts were too small, Fisher exact test was used. For continuous variables where a difference in mean value between high and low users at 12 months was calculated, an independent samples *t* test was used. Statistical significance was set at $P<.05$. The tertiary outcome subanalysis was adjusted for baseline values and analyzed using the ANCOVA (analysis of covariance) model. Analyses were conducted using Excel (Microsoft Office Professional Plus 2016) and SAS version 9.4.

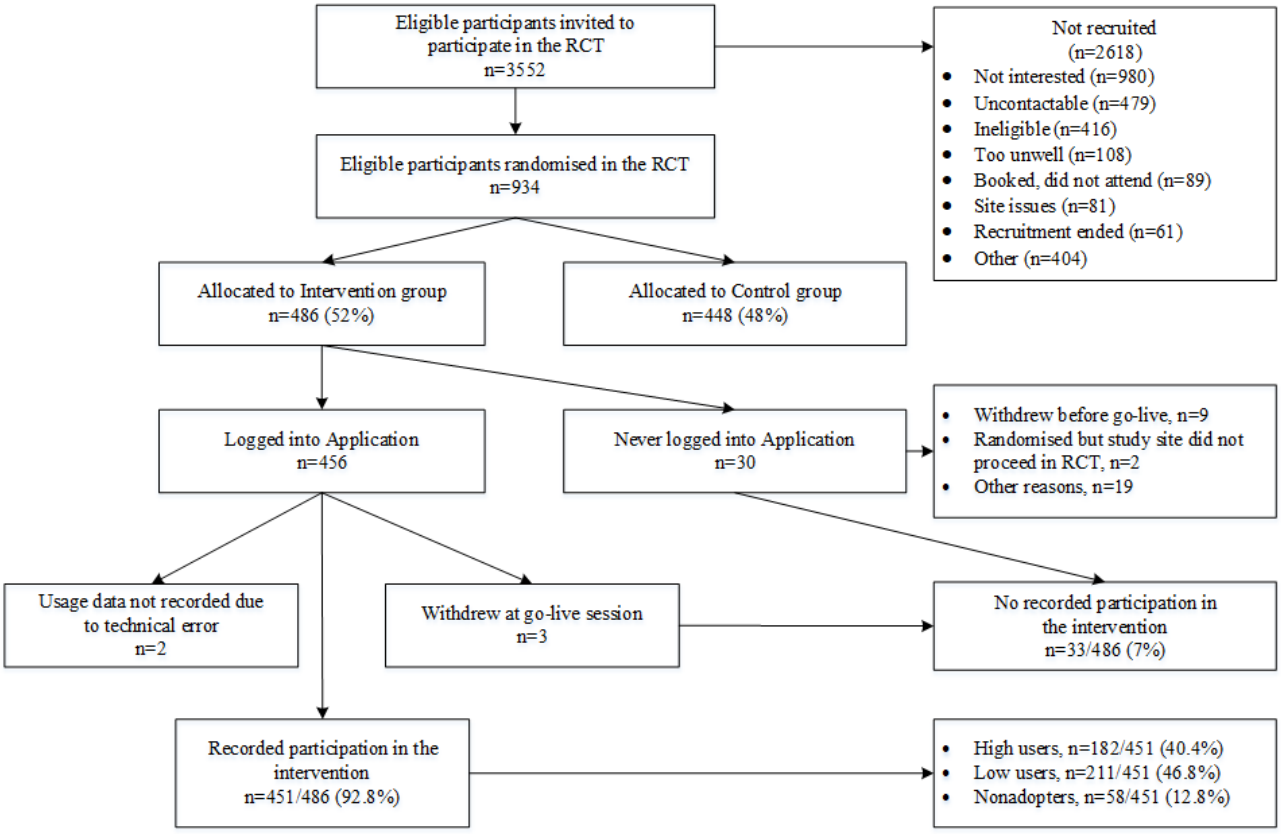
Results

Intervention Reach

Between November 2014 and May 2017, 934 eligible patients were enrolled in the RCT. Of these, 486 (52.0%) were allocated to the intervention group. The flow of the participants in the

intervention group is shown in Figure 1. Participants exposed to all or part of the intervention (451/486, 92.8%) were defined as being registered in the shared record software (RecordPoint) *and* logged into the app *and* used at least one feature or component at least once. Reasons for part-participation were (1) participants opting out of system-generated SMS text messaging or email content (although they could still access the website or rejoin system-generated content at any time), (2) participants stopping log-ins to the website but remaining opted in to system-generated SMS text messaging and/or email content, and (3) participants withdrawing from the study after go-live but before completion of the follow-up period.

Figure 1. Intervention reach. RCT: randomized controlled trial.



Intervention Content Delivery

Of the 486 participants allocated to the RCT intervention group, 99.6% (484/486) had a personalized, secure website populated with data from their EHRs. A study site that discontinued participation left 2 enrolled participants without access to the intended intervention.

In total, 2.7% (13/486) of the cohort withdrew at or before go-live training, leaving 97.3% (473/486) to undergo training. Of these, 94.5% (447/473) received the training session with staff as intended, and 6% (26/473) who could not keep either their telephone or their in-person appointment were mailed the self-directed materials, enabling them to log in and orient themselves to the application and receive staff assistance if they chose to do so. Of the 451 participants with a recorded go-live, most (447, 99.1%) received data uploads from their EHR for a full follow-up period of 12 months. Four participants from a primary health care site that closed within the course of the study did not have EHR-sourced data updated in their portal for their full follow-up period. Most log-in sessions by participants occurred on laptops, desktop computers, and tablets, and a minority of the log-in sessions occurred on smartphones.

An automated system sent messages at a preset, tapering frequency, commencing at the go-live session. Each month, 20% of the message delivery records were randomly selected and reviewed for compliance with the intended delivery schedule. Across both formats, delivery accuracy ranged from 86% to 100%. A severe disruption to weekly volumes of SMS text message delivery was identified by staff in March 2017. Program records indicated the text messages that were queued to be sent but did not indicate the actual message receipt; hence, nonreceipt would go unnoticed unless participants reported the problem. Once notified, programmers traced and rectified the source of the problem within the notification system; however, in the interim, a small number of participants did not receive the heart health content delivered via SMS text messaging.

Scheduled and Ad Hoc Participant Contact

Intervention training occurred over the telephone (413/447, 92.4%) or in-person (34/447, 7.6%). The mean training time for each format was 44 minutes (range 17-120 minutes) and 38 minutes (range 10-90 minutes), respectively. The characteristics of the scheduled participant support after training are summarized in Table 2.

Overall, 1804 follow-up calls were predicted based on 4 calls to each of the 451 participants, and 98.45% (1776/1804) were achieved. Of those, 80.29% (1426/1776) were verbal conversations and 19.71% (350/1776) were communications

via email or voicemail messages. No contact was attempted with participants who, in the time since the previous scheduled contact, requested no further calls from the study staff. Overall, the median time required for follow-up communication was 5 minutes. Technical problems or *how to* questions comprised most of the reported problems, and although they reduced over the 4 contact time points, forgotten passwords and SMS text messaging or email faults or preferences together accounted for approximately half of the technical assistance at week 26.

Unlike the scheduled contact, most of the ad hoc contact was initiated by participants (286/483 episodes, 59.2%) and involved 285 unique participants, representing 63.2% (285/451) of the participants. Ad hoc contact was more frequent by email (363/557, 65.2%) than by telephone (158/557, 28.4%) and required an average of 8.5 (SD 7.5) minutes of staff time (range 1-80 minutes; median 5 minutes). Most of the contact was for resolution of technical problems such as log-in or password problems (244/483, 50.5%), turning message receipt on or off (66/244, 27.0%), or discussing tracking or self-monitoring measurements (53/244, 21.7%). General feedback and miscellaneous administrative issues accounted for 26.3% (127/483). Content data were incomplete in 23.2% (112/483) of the contact logs.

Table 2. Characteristics of scheduled intervention participant support during study follow-up.

Characteristic	Scheduled follow-up after go-live training session				
	Week 2	Week 6	Week 12	Week 26	All weeks
Contact episodes initiated by project staff (n)	451	446	442	437	1776
Communication format (where specified)					
Contact attempts, n (%)	458 (25.46)	450 (25.01)	446 (24.79)	445 (24.74)	1799 (100.00)
Telephone	432 (94.32)	427 (94.89)	425 (95.29)	423 (95.06)	1707 (94.88)
Email	20 (4.37)	21 (4.67)	20 (4.48)	19 (4.27)	80 (4.45)
SMS	5 (1.09)	2 (0.44)	1 (0.22)	3 (0.67)	11 (0.61)
Time taken (min)					
Total	2767	2606	2564	2576	10,513
Mean (SD)	6 (4.8)	6 (4.1)	6 (4.6)	6 (3.8)	6 (4.3)
Minimum, maximum	1, 45	1, 45	1, 60	1, 35	1, 60
Median (Q1[a]; Q3[b])	5 (5, 5)	5 (4, 5)	5 (5, 5)	5 (5, 5)	5 (5, 5)
Problem category (where specified), n (%)					
No problems reported by participant[c]	320 (83.55)	332 (90.22)	309 (90.88)	297 (88.66)	1258 (88.22)
More training (technical steps and/or explaining clinical information)[d]	9 (1.99)	4 (0.90)	4 (0.90)	4 (0.92)	21 (1.18)
Technical problem or *how to* inquiry[c]	57 (12.64)	33 (7.40)	30 (6.79)	23 (5.26)	143 (8.05)
Log-in problem or password reset	18 (31.58)	7 (21.21)	13 (43.33)	12 (52.17)	50 (34.97)
Turning automated heart health message tip receipt on or off; faulty email delivery	19 (33.33)	11 (33.33)	7 (23.33)	11 (47.83)	48 (33.57)
Tracking or self-monitoring measurements	12 (21.05)	10 (30.30)	6 (20.00)	2 (8.70)	30 (20.98)
Request to withdraw or change study participation status	3 (0.67)	2 (0.45)	2 (0.45)	1 (0.23)	8 (0.45)

[a]Q1: first quartile.

[b]Q3: third quartile.

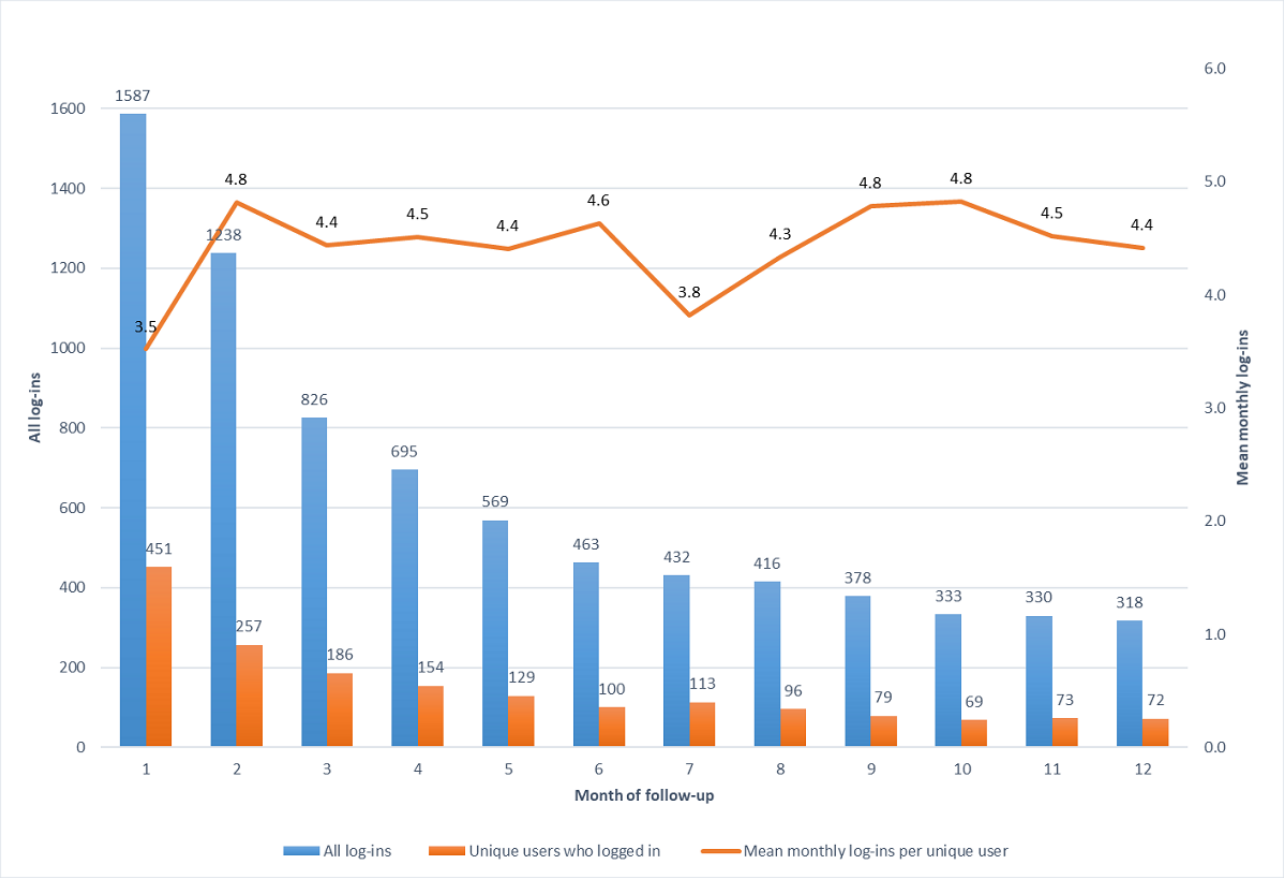
[c]The denominator includes telephone and email communication.

[d]The denominator is contact episodes in which staff directly spoke to the participant.

Intervention Log-In Frequency and Duration

The overall log-in activity pattern was used to indicate the frequency with which participants made contact with the intervention. The highest number of log-ins (n=1587) occurred in month 1 of participation, with subsequent monthly log-ins decreasing steeply by month 6 (n=463) and then remaining relatively stable from months 7 to 12 (Figure 2). Log-in activity by unique users began with all users (n=451) logging in at least to go-live. Unique user log-ins declined markedly in month 2 (257/451, 56.9%) and thereafter were made by up to 50.9% (200/393) of those who made any use of the intervention (after nonadopters were excluded). The mean log-in frequency among ongoing unique users was 3 to 4 times per month over 12 months. The median monthly log-in frequency was 4.5 (Q1, Q3: 4.4, 4.7) among ongoing users.

Figure 2. Characteristics of log-in activity.



Intervention Exposure

Excluding the nonadopters (58/451), monthly unique users dropped below 200 from month 3 among participants who continued to make any use of the intervention (393/451) and below 100 from month 8 (Figure 3). Only goals tracking was accessed by more than 50 unique users per month for 12 months. In terms of goal setting, 86.0% (388/451) of all users set goals for healthier lifestyle behavior, and most goals were for healthier eating and physical activity. More people set goals than returned to track their goal achievements. Overall, the social media or chat forum feature was the least subscribed, being visited by 12.7% (50/393) of unique users per month beyond month 4. Contributions to the forum by participants ranged from 0 to 12 postings each month.

In terms of the intensity of screen visits over time by unique users, visits to track goal progress markedly exceeded visits to all other screens over each of the 12 months, even though fewer unique users were logging in (Figure 4). By month 12, for example, 18.3% (72/393) of unique users made approximately 500 visits to the goal-tracking screen. In comparison, in the same month, there were fewer than 100 visits to each of the other interactive features, suggesting that electronic goal tracking was a valued feature determining returning log-ins. Overall, personal goal setting, risk factor monitoring, and CVD risk score estimation were accessed more than the chat forum and medicine features, with the highest interest in the earlier months.

Figure 3. Number of unique users who log in and access interactive screens per month of follow-up. CVD: cardiovascular disease.

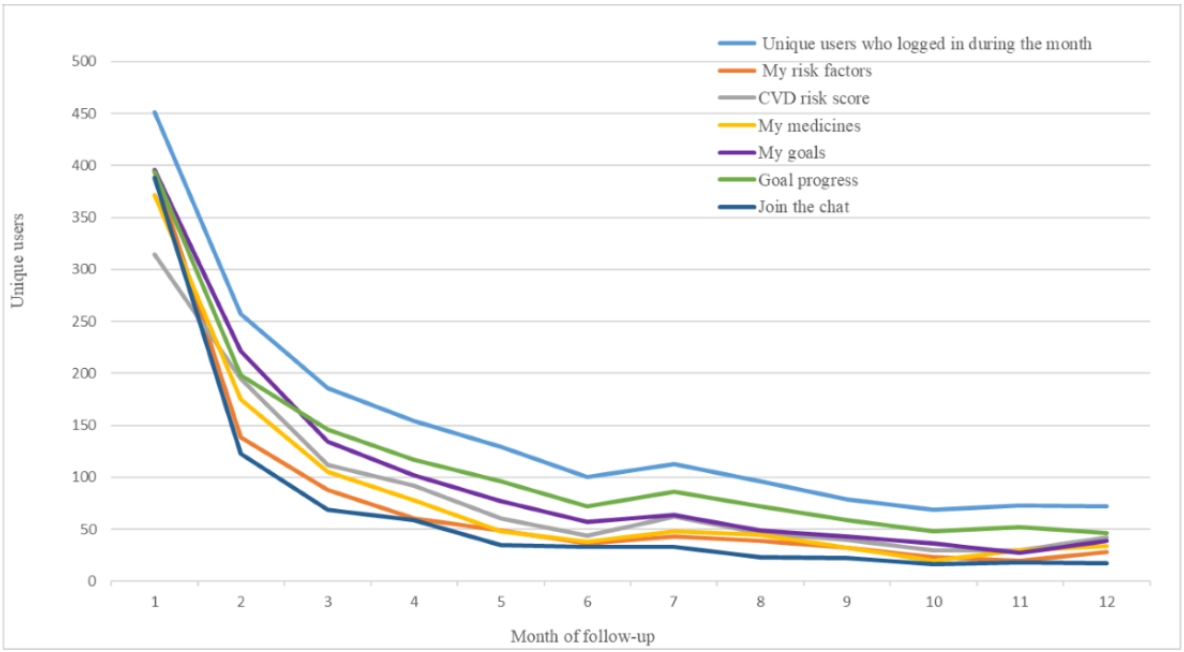
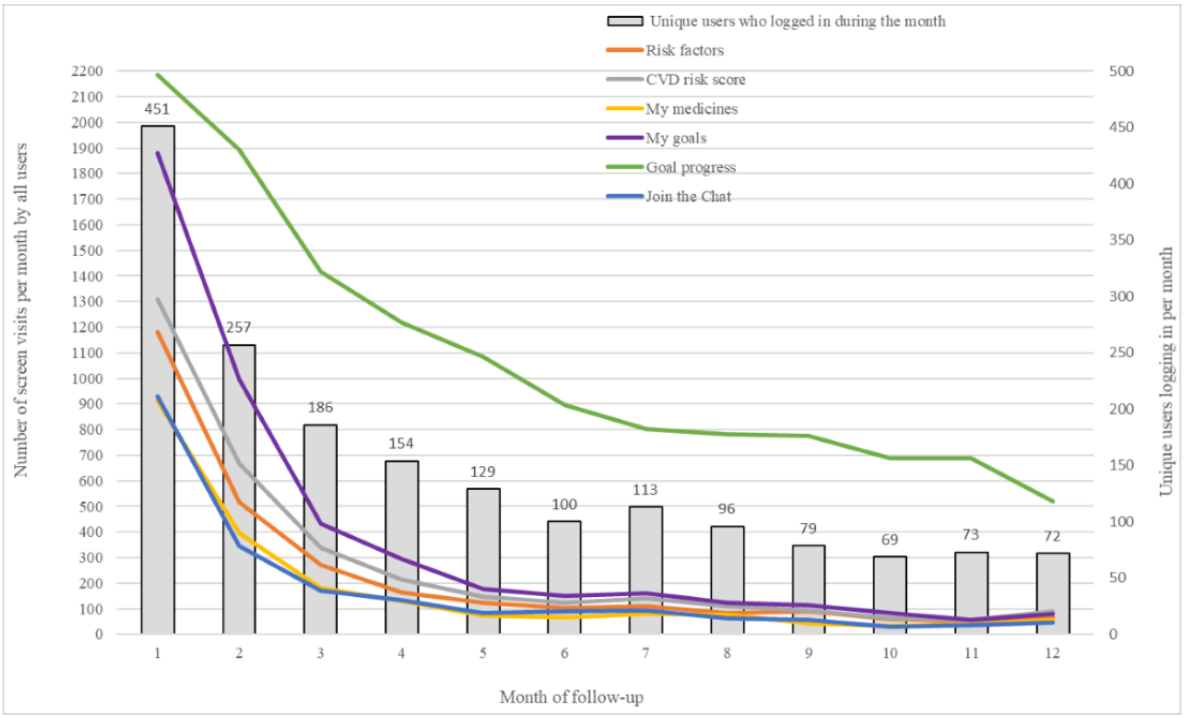


Figure 4. Intensity of screen visits over 12 months of follow-up. CVD: cardiovascular disease.



Subanalysis of High Versus Low Intervention Users

Baseline Characteristics

Of the 451 participants with usage records, 12.8% (58/451) were nonadopters, 46.8% (211/451) were low users, and 40.4% (182/451) were high users (Table 3). There were no major baseline differences between these groups in terms of demographic and biometric data, CVD risk status, or other health conditions. The characteristics of all intervention group participants in this evaluation study mirrored those of the overall RCT cohort [27], although the proportion of Aboriginal and Torres Strait Islander Australians was higher than in the RCT (4.9% compared with 4.0%) and slightly higher than the proportion of Aboriginal and Torres Strait Islander representation in the general Australian population in the most recent census (3.3%) [32]. The mean difference in eHealth literacy scale (eHEALS) scores between low and high users was significant (1.13; 95% CI 0.53-1.72; P<.001). Most of the nonadopters (32/58, 55%) received school education only, whereas postschool education was more frequent in the low (150/211, 71.1%) and high (142/180, 78.9%) users.

Table 3. Baseline characteristics by intervention user subgroups.

Characteristics	Nonadopters[a] (n=58)	Low users[b] (n=211)	High users[c] (n=182)	Total (N=451)
Age (years), mean (SD)	66.8 (8.2)	66.8 (8.2)	67.2 (8.7)	67.0 (8.4)
Male, n (%)	40 (68.9)	158 (74.9)	147 (80.8)	345 (76.5)
CVD[d] risk, n (%)				
Existing	20 (34.5)	91 (43.1)	77 (42.3)	188 (41.7)
High	38 (65.5)	120 (56.9)	105 (57.7)	263 (58.3)
Ethnicity, n (%)				
Indigenous Australian	6 (10.3)	10 (4.7)	6 (3.3)	22 (4.9)
White	46 (79.3)	176 (83.4)	161 (88.5)	383 (84.9)
South Asian	3 (5.2)	3 (1.4)	6 (3.3)	12 (2.7)
Other Asian	1 (1.7)	3 (1.4)	3 (1.6)	7 (1.6)
Other	2 (3.4)	19 (9.0)	6 (3.3)	27 (5.9)
Education level, n (%)				
Secondary school or below	32 (55.2)	61 (28.9)	38 (21.1)	131 (29.2)
Technical or vocational qualification, or above	26 (44.8)	150 (71.1)	142 (78.9)	318 (70.8)
Annual household income, n (%)				
<Aus $104,000 (US $79,366)	34 (58.6)	120 (56.9)	114 (63.0)	268 (59.6)
≥Aus $104,000 (US $79,366)	24 (41.4)	91 (43.1)	67 (37.0)	182 (40.4)
CVD risk factors				
BMI (kg/m^2), mean (SD)	29.4 (6.74)	30.3 (5.45)	29.5 (5.22)	29.8 (5.54)
BMI ≥30 kg/m^2, n (%)	22 (37.9)	101 (47.9)	69 (37.9)	192 (42.6)
Waist circumference (cm), mean (SD)	104.5 (18.1)	107.0 (15.4)	104.9 (13.1)	105.8 (14.9)
SBP[e] (mm Hg), mean (SD)	138.7 (16.7)	136.4 (15.6)	139.1 (16.1)	137.8 (15.9)
DBP[f] (mm Hg), mean (SD)	79.7 (11.4)	78.4 (10.6)	79.4 (10.5)	79.0 (10.7)
Current smoker[g], n (%)	10 (17.2)	22 (10.6)	18 (9.9)	50 (11.2)
LDL[h] cholesterol (mmol/L), mean (SD)	2.6 (1.09)	2.6 (1.09)	2.6 (1.0)	2.6 (1.05)
HDL[i] cholesterol (mmol/L), mean (SD)	1.4 (0.49)	1.3 (0.37)	1.3 (0.36)	1.3 (0.39)
HbA_{1c}[j]				
Participant, n (%)	21 (36.2)	71(33.6)	49(26.9)	141(31.3)
Mean (SD)	6.8 (1.0)	7.0 (1.2)	6.9 (1.4)	6.9 (1.3)
Comorbidities, n (%)				
Previous stroke	6 (10.3)	20 (9.5)	14 (7.7)	40 (8.9)
Coronary heart disease	15 (25.9)	78 (36.9)	65 (35.7)	158 (35.0)
Atrial fibrillation	6 (10.3)	21 (9.9)	17 (9.3)	44 (9.8)
Diabetes mellitus	22 (37.9)	75 (35.5)	50 (27.5)	147 (32.6)
COPD[k] or emphysema	4 (6.9)	17 (8.1)	9 (4.9)	30 (6.7)
Self-reported medication use, n (%)				
Antihypertensive	36 (72)	128 (61.8)	106 (60.6)	270 (62.5)
Lipid-lowering	26 (52)	123 (59.4)	94 (53.7)	243 (56.3)
Antithrombotic	16 (32)	86 (41.5)	71 (40.6)	173 (40.0)
PBS[l] medication use, n (%)				

Characteristics	Nonadopters[a] (n=58)	Low users[b] (n=211)	High users[c] (n=182)	Total (N=451)
Antihypertensive[m]	31 (62)	106 (51.2)	89 (50.9)	226 (52.3)
Lipid-lowering[m]	18 (36)	79 (34.8)	72 (41.1)	169 (39.1)
Antithrombotic[m]	3 (6)	38 (18.4)	17 (9.7)	58 (13.4)
eHEALS[n] mean (SD)	24.6 (7.7)	26.7 (6.3)	28.5 (5.7)	27.1 (6.4)

[a]Nonadopter is defined as a participant who logged into the app only once in total.

[b]Low user is defined as a participant who logged into the app at least once in 3 or fewer months of follow-up.

[c]High user is defined as a participant who logged into the app at least once in more than 3 months of follow-up.

[d]CVD: cardiovascular disease.

[e]SBP: systolic blood pressure.

[f]DBP: diastolic blood pressure.

[g]Eight participants with missing carbon monoxide breath test results.

[h]LDL: low-density lipoprotein.

[i]HDL: high-density lipoprotein.

[j]HbA_{1c}: hemoglobin A_{1c}.

[k]COPD: chronic obstructive pulmonary disease.

[l]PBS: pharmaceutical benefits scheme.

[m]Forty-three participants withdrew consent for the use of their pharmaceutical benefits scheme data.

[n]eHEALS: eHealth literacy scale. Maximum score is 40.

RCT Outcomes

There was no significant difference in the primary outcome of adherence to guideline-recommended medication between the low- and high-user groups (*P*=.44), although the proportion of participants that was adherent increased by 5.7% in the high-user group but only by 3.1% in the low-user group. In the unadjusted analyses, compared with the low-user group, the high-user group had significantly higher eHEALS scores and mean numbers of fruit and fish serves per week; however, these differences were not significant after adjustment for baseline scores. Compared with low users, high users also had nonsignificant higher adherence rates to blood pressure–lowering therapy and statin medications, meeting Australian guideline targets for both blood pressure and low-density lipoprotein cholesterol, and doing at least one thing to control their salt intake.

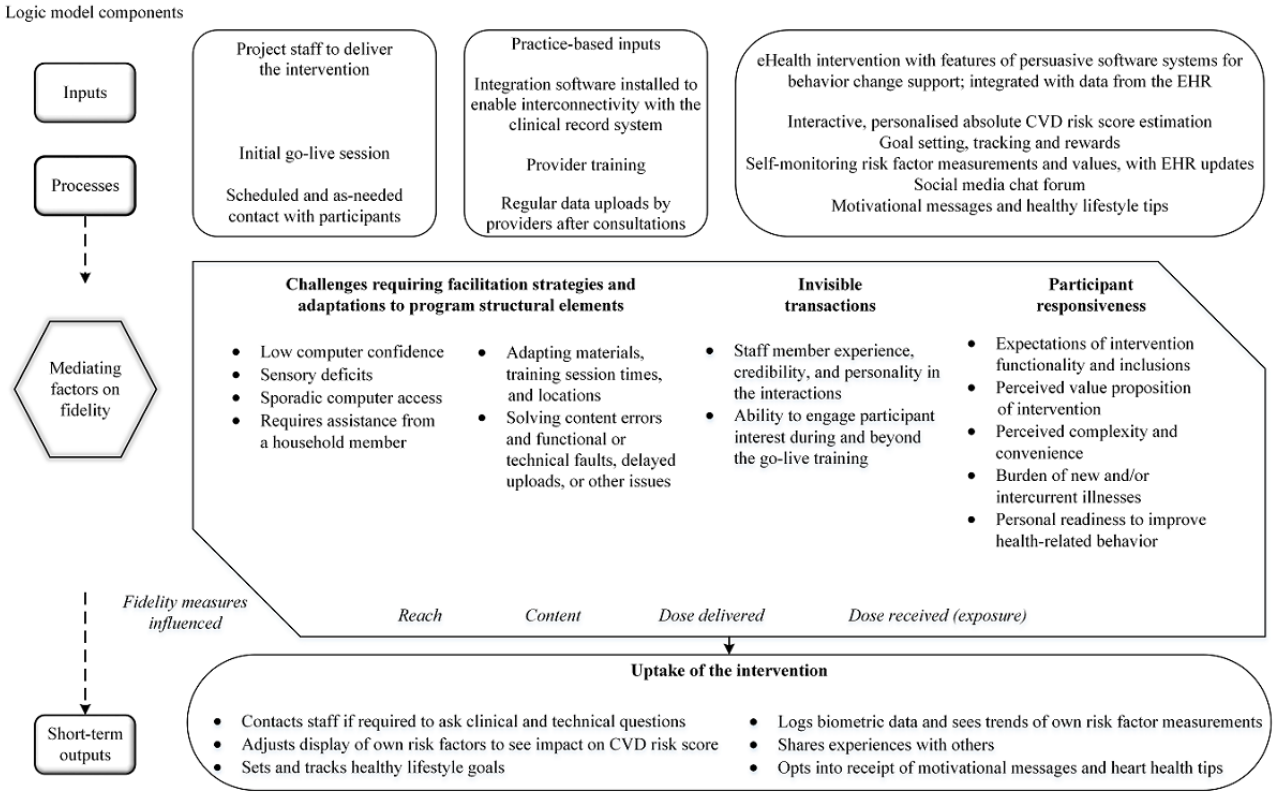
Exposure to Email and SMS Content

Most participants in each of the subgroups opted to receive messages for more than 3 months of follow-up: 96.2% (175/182) of high users, 86.9% (166/191) of low users (data missing for 20/211, 9.5% of low users), and 90% (52/58) of nonadopters.

Factors Mediating Intervention Delivery and Receipt or Early Adoption

At the completion of the RCT, 3 factors were identified from the review of program delivery and participant communication records as having influenced intervention dose delivery and receipt, in relation to the stages of the program logic model (Figure 5).

Figure 5. Modified logic model showing mediating factors on the relationship between the intervention processes and outputs. CVD: cardiovascular disease; EHR: electronic health record.



Facilitation Strategies by Staff and Adaptation of Structural Elements of Program Delivery

Standardized intervention delivery was subject to some variation and adaptation, as different participant needs were identified by staff conducting the go-live training and follow-up communication. After reviewing the communication records with participants, a range of facilitating strategies emerged as having been important to intervention receipt; they are summarized in Textbox 1. The conduct of go-live training was always adapted to enable those with hearing or speech impairments, or less computer confidence, to take part. A paper-based reference guide was provided to those who found website navigation and functions difficult to recall. Telephone-based training with participants with low skills or memory impairment extended the time requirements and underscored the need for clear, simple instructions with less technical jargon; setting up website shortcuts and browser bookmarks to simplify return log-ins; allowing time to practice on interactive screens; and split-session training. Some participants needed a family or household member to be present or required their training appointment outside business hours.

Textbox 1. Facilitation strategies required to support intervention delivery and receipt.

Facilitating strategies identified from communication records:

- Face-to-face go-live training offered to those unwilling to do so by telephone
- Go-live scheduling outside business hours (eg, evenings or Saturday mornings)
- Inclusion of a carer or family member in go-live training or ongoing communication to facilitate use of the intervention
- A self-directed paper guide for participants who
  - required a nonelectronic reference to improve confidence with navigation of the website and its features
  - had hearing or speech impairments that would make telephone-based go-live training unsuitable
  - did not participate in go-live training with staff
- Retraining of website navigation skills at any time in the follow-up period
- Resolution of technical questions that could reasonably be done by the participant without staff assistance (eg, changing preferences or settings within the app)
- Trouble-shooting delayed upload of recent measurements or pathology test results from the electronic health record, which may have revealed a fault with the data integration software at the health service
- Sending a courtesy email to a participant when an error or technical problem within their application had been corrected or if a required correction was delayed
- Removing or correcting erroneous biometric data entry within screens or charts that were not editable by the participant

Invisible Transactions

Staff member experience and personality as a mediating factor on the quality of professional communication and interaction with participants and the ability to engage their interest during and beyond *go-live* training were not formally measured. It is uncertain if, and how much, the staff members' credibility, trustworthiness, skill, and friendliness affected participants' initial and ongoing willingness to engage with the intervention. Regardless, program delivery and communication records revealed that these qualitative influences were likely at play, although they could not be quantified. Staff notes about personal health topics raised by participants in calls and emails and several elements of the study design underscored the value of staff attributes, namely, (1) the personalized nature of EHR-derived data that often required clear, accurate explanations by telephone or email, without complex medical or technical jargon; (2) the remote intervention delivery and support arrangements that increased anonymity between parties and prevented face-to-face communication cues; and (3) the wide variation in participant ages, education, and digital literacy that necessitated more supportive approaches for some. During a call at week 26, a participant stated that she always felt more energized after calls with staff, suggesting that human contact may still be valued even after 6 months using a self-directed resource.

Participant Responsiveness

Four concepts appeared to influence participant responsiveness to the intervention, and hence log-in frequency and exposure, given that all components and content were available.

First, participants' preferences for intervention functionality and inclusion influenced their reactions. The concept and presentation were appealing, but some participants' expectations were not met. Comments included:

> *...I would like to see log space for blood sugar levels measured at home.* [Go-live phone call]
>
> *Simply adding data and reading all the information is not enough* [Week 2 phone call]
>
> *...I wish I had this when I had my heart attack. I am doing more walking and eating more vegetables than I ever have before. It keeps me on track all the time.* [Week 6 phone call]

A second and related expectation was the perceived value offered by the intervention. Some viewed it as minimally useful compared with their existing resources or habits for managing their health, suggesting the need for strong personal relevance of content or functions. Others gained motivation for their healthier lifestyle efforts and were prompted to log in. The descriptive comments included the following:

> *...After my angioplasty...the information was largely what had been covered in the hospital rehab program and did not really provide any additional motivation to me.* [Email from a participant]
>
> *...It's been a big help...tracking goals has become my routine now. Email and SMS tips are reinforcing, it makes me want to go back into my app to update the tracker.* [Week 26 phone call]
>
> *Setting goals gave me an incentive to log on.* [Email from a participant]

Interestingly, the EHR data integration was of less value to participants who felt a close rapport with their care provider or for whom new data uploads would be infrequent:

> *...The tie-in with the [doctor] was good but not useful for me because I don't go that often.* [Week 26 phone call]

> *...A lot of the Program I found was not useful to a person like me who has a good relationship with their GP and practice.* [Email from a participant]

A third issue affecting log-ins was perceived convenience compared with other digital health apps or devices. The intervention was useful for monitoring and assessing risk but not for fitness tracking, for example, which was a priority for some users. For example, participants described the concurrent use of commercial phone-based apps or a wearable device on which they could readily track their physical activity and dietary behaviors. Others were satisfied with the content from the automated messages and, therefore, logged in less often:

> *...I haven't logged onto the app very often; however, I like receiving the tips, they remind me to be good.* [Week 12 phone call]

> *I use other apps to count calories and steps.* [Week 26 phone call]

> *...I am not much on the app; I prefer to access an app quickly on the phone with few clicks to see the essential info.* [Week 6 phone call]

Fourth, timing relative to other personal priorities affected intervention exposure. Participants described infrequent log-ins because they (or their spouse, for example) had a new or existing intercurrent illness with immediate priority. Examples included new diagnoses of cancer or other long-term conditions, unforeseen surgery, the demands of a new treatment or therapy, and frequent clinic appointments.

These responsibilities lowered engagement with the intervention. For others, study enrollment coincided with their readiness to improve their health-related behavior. A participant commented that his participation was "a wake-up call" that helped him to succeed with weight loss goals. Another stated after 6 months that "This is the catalyst that made me get stuck in." Thus, the timing of participation was both a constraint and an enabler of intervention exposure.

Discussion

Principal Findings

This mixed methods evaluation examined the intervention delivery fidelity of a consumer-focused web application with data integration from the primary health care EHR and optional health message receipt. No single or uniform measure defines or quantifies fidelity [20,33], but core evaluation metrics of reach, content fidelity, and dose delivered [19,30] overall were fulfilled as intended. Adaptations to routine implementation, known as structural adaptations [34], were made to overcome individual barriers to intervention reach and receipt. Initial high-user log-ins dropped early in the follow-up period, a trend noted across such interventions more generally [35,36], and the unique user log-in rate tapered to a mean frequency of 3 to 4 per month. Progress tracking had the highest screen visit intensity for the longest duration. Screens with EHR-derived data generally held stronger interest in the early follow-up period. Email and SMS text message receipt augmented active participation by high users and strengthened content exposure in low users and nonadopters. Intervention nonadopters, low users, and high users were similar across a wide range of characteristics, suggesting that the website was amenable to use in general but not which characteristics were important for log-in frequency. The association of patient-level factors and log-in behavior would thus be of further research interest for improving design or excluding content [37]. More frequent intervention use was associated with nonsignificant differences in clinical measures and health-related lifestyle behaviors after 12 months. It is possible that the study was insufficiently powered to fully elucidate the impact of higher intervention exposure; however, in any event, the differences were small in absolute effect size and may not be clinically important. Further research could ascertain if the higher eHEALS scores noted at baseline in the high-user group are an important precursor to using digital health interventions and if more frequent use raises self-reported eHealth skills. Prescribed log-in activity in relation to the adoption of desired offline behaviors may be an area of further inquiry, particularly as website engagement has been shown to benefit physical activity and eating behaviors, even when not significantly associated with biological outcomes [38]. Qualitative data revealed important influencing factors on intervention delivery and receipt, a noted advantage of mixed methods inquiry in process evaluations [13], and a strength of this study. The original program logic model was used as not only a representation of the causal assumptions within the intervention [13] but also as a scaffold on which to show where the identified factors impacted intervention delivery and uptake [20].

Engagement with technology has been defined as a 4-stage cyclical process consisting of an initial point of engagement, the period of engagement, disengagement, and reengagement, with the stages having both shared and exclusive sensory-emotional and spatiotemporal attributes [39]. For example, the system's novelty and esthetic attributes, combined with user emotions of motivation and interest, are especially important to initiation; attributes such as customization, feedback, and control that promote positive affect appear important to ongoing and return visits; and low-level interaction, boredom, or other negative emotions influence disengagement [39]. A recent systematic review found that changes to health status caused disruption or drop out, as did user perception of the technology's compatibility with their routine, their own digital literacy, and relevance to their symptoms or (dis)abilities [40]. In the diffusion of innovations theory, individuals look to reduce their uncertainty about the consequences of adopting an innovation and to perceive a relative advantage if they do so [41]. Factors such as personal convenience, satisfaction, and suitability for needs are proposed to matter more than the intervention's objective advantage. Notably, for this study, reliance on what the participants themselves chose to undertake was an important contextual factor for adoption [42], underscoring the significance of user-perceived relevance and value.

A more recent framework calls the desirability of the technology to the user its value proposition, and more complex interventions may be less likely to reach this threshold for adoption [43]. Personal relevance, program expectations, current health behavior, convenience, and so on, as identified in this study,

concur with engagement factors for digital health interventions previously identified, particularly the themes of personal agency and motivation, personal life and values, and perceived quality of the intervention [44]. Notably, in a multidomain model for engagement with web-based interventions [14], the authors point out the utility of these determinants in any framework of assumptions about intervention use. The relevant framework in this study was the program logic model, in which the determinants derived from the qualitative data analysis were merged to help explain program uptake and what the intervention delivery required in practice (Figure 5). The perception of advantage from using such interventions may have an upstream influence on recruitment to the RCT; of the eligible invited participants who were not recruited, 37.43% (980/2618) gave their reason as *not interested* (Figure 1). Although reasons for disinterest among the RCT nonparticipants were outside the scope of this evaluation, we recommend that further studies be conducted to explore and define a value proposition for similar interventions targeting the primary health care context. Furthermore, the overall generalizability of findings may be improved with greater participation of primary care attendees in studies of digital interventions. The number of participants randomized represented 26.29% (934/3552) of those invited to participate in the RCT. Hence, there may be barriers and enablers of engagement and uptake that would be further understood with higher study participation in this setting.

Patient engagement with portals directly linked to an EHR has inconsistent definitions of both adoption [17] and active use, ranging from at least one use, [45] to more than 1 log-in every 4 months, [37] and at least two log-ins in 12 months [16]. Other studies of web-based lifestyle and disease management interventions have noted that 46% of participants abandon the program after a single log-in [46] and high-frequency use of progress-tracking features by returning users [38]. In general, interventions in which information is tailored hold more user interest and show less attrition than those with generic information design [47], an inflexible or static website [48], or ones that give virtual rather than human feedback [38]. A study in which participants with diabetes and/or CVD accessed a web-based portal to view and track information within their EHR noted that those who used data-tracking functions only comprised between 4% and 11% of users and were among the most frequent and consistent users [16]. Clearly, not all EHR-integrated functions will attract all users, and a range of administrative functions may hold value for many patients over clinically oriented functions.

However, within this study, intervention features with EHR-derived data, such as CVD risk score estimation, medications, and risk factor status, were visited more frequently in the early follow-up period than in the later follow-up period. Overall log-in attrition contrasted with comparatively high average monthly log-ins by ongoing users, suggesting that they derived some personal value. The constant appeal of tailored progress tracking was perhaps because of more immediate visual feedback for everyday healthier lifestyle behavior. The informational nature of CVD risk estimation and risk factor status may benefit an initial call to action but not persistent revisits if few updates occur in the shorter term, or it may have entirely disinterested some users. Furthermore, few outcome measures were associated with greater log-in frequency. Further research could explore whether ongoing exposure impacts important but less quantifiable cognitive and emotional stages and processes of health behavior change, [49] as this information becomes increasingly important to design of technologies with this intent [50]. Direct interaction with the primary care provider was not designed into this intervention, so its value to EHR-linked innovations in the local context requires further inquiry. In provider-linked portals with disease self-management intent, more frequent contact from a clinician may nudge website engagement rates [51] although other user differences may be important, such as their primary care providers' *buy-in*, their baseline self-efficacy, or having a new versus long-term diagnosis [47]. Provider endorsement of a digital application as an extrinsic motivator of engagement is hypothesized but requires investigation [41,52].

High delivery fidelity raises the resource-related question of the optimal support requirements for ongoing website engagement. The type of intervention support appears to benefit different exposure measures. Peer and counselor support facilitate longer session length, for example, when offered in interventions for weight, alcohol, and smoking reduction, whereas updated website content and email and/or telephone contact appear important for site revisits [35]. Other triggers of staff-initiated phone contact could prompt website revisits, for example, a defined number of consecutive days without log-in activity, in addition to scheduled follow-up [44]. Patients are receptive to automated revisit reminders via telephone or email but prefer personal feedback from, for example, a nurse about their disease self-monitoring and health behavior activities [36]. Variations in the intended intervention delivery are generally more likely with complex interventions [21] and with foreseeable differences in provider experience and skill [45]. In this study, the characteristics of study personnel were important contextual and quality factors because the study relied largely on non-face-to-face communication to deliver and support the intervention. In such a context, personality, attitude, and perceived expertise may influence user interest and the reliability of program delivery [21,34]. In addition, contact initiated by participants yielded more app-related problem solving, suggesting that ad hoc support is important to be able to access on the participants' own timeline of website interaction. Flexible rather than fixed staff facilitation may, therefore, be a worthwhile design consideration because different types of users prefer the option of more human involvement for guidance or accountability, whereas others may choose less or require less over time [46].

Limitations

First, although typical fidelity measures were chosen, quantified, and described to suit the intervention, it is acknowledged that fidelity research requires systematic scoring practices to better enable comparison and replication of like interventions [53]. Furthermore, there were no prescribed usage or adherence goals with which to compare participant subgroups, but it is acknowledged that uniform concepts of adherence in eHealth would assist in understanding which elements and in what dose, are associated with the intended outcomes of the intervention.

[54]. Second, the duration of session and screen visits would have enabled further quantification of use intensity and the relative appeal of features and website components. Third, this study intentionally focused on process rather than outcomes and excluded any assessment of participants' offline lifestyle behavior for CVD risk factor control. For this reason, other types of engagement with the overall program content may have been underestimated [52]. Fourth, this study excludes cost-effectiveness measures that are important when balancing user uptake with the required facilitator resources, software application oversight and maintenance, and health service support of the EHR linkage software. Furthermore, this study was a summative evaluation. Although intervention delivery and monitoring were dynamic and adaptable throughout the RCT, it is acknowledged that formal evaluations at interim time points can be useful if it is possible to make regular or ad hoc modifications to the program.

Conclusions

A complex eHealth intervention designed for overall self-directed use can be implemented with high delivery fidelity. Personal progress tracking was consistently used; EHR-derived data features were early but not persistent triggers of revisits. This design intent may reflect how consumers use web-based resources in everyday life but makes usage frequency and thus exposure more unpredictable. Hence, mediating influences and support intensity should be factored into future program planning as drivers of reach and uptake. Despite high delivery fidelity, more frequent intervention use was only associated with small, nonstatistically significant improvements in medication adherence; some clinical measures; and lifestyle behaviors after 12 months. In recognition of multifactorial drivers of engagement, a more explicit personal value proposition should target broad user variables, such as motivation, personal relevance, the context of health care provider use, the timing of program exposure, digital literacy, and preference for log-in adherence accountability. In future research, outcomes related to intervention exposure and reach are important to report so as to expand the evidence about the user and system attributes that promote the uptake of health records with disease self-management functions.

Acknowledgments

The authors would like to thank the patients and general practitioners from the primary health care services that participated in the CONNECT RCT. The authors thank the following project staff for their assistance with the setup, maintenance, delivery, and support of the EHR-integrated intervention: S Parker, C Pitt, J Forbes, and F Bampi. The authors would like to acknowledge Professor Mark Harris, Professor Noel Hayman, Professor Anthony Rodgers, Professor Anushka Patel, Associate Professor Kathryn Panaretto, Professor Clara Chow, Dr Annie Lau, and Dr Emma Heeley for their contributions to the design of the CONNECT RCT and process evaluation protocols. GC was supported by a Health Professional Scholarship (ID 101544) from the National Heart Foundation of Australia. The CONNECT RCT was funded by the Australian National Health and Medical Research Council (ID 1047508). JR is funded by the National Health and Medical Research Council Career Development Fellowship (APP1143538). The funders had no role in the review and approval of the manuscript for publication.

Conflicts of Interest

None declared.

References

1. Joseph P, Leong D, McKee M, Anand SS, Schwalm J, Teo K, et al. Reducing the Global Burden of Cardiovascular Disease, Part 1: The Epidemiology and Risk Factors. Circ Res 2017 Sep 01;121(6):677-694. [doi: 10.1161/CIRCRESAHA.117.308903] [Medline: 28860318]
2. Jennings C, Astin F. A multidisciplinary approach to prevention. Eur J Prev Cardiol 2017 Jun;24(3_suppl):77-87. [doi: 10.1177/2047487317709118] [Medline: 28618913]
3. Riegel B, Moser DK, Buck HG, Dickson VV, Dunbar SB, Lee CS, American Heart Association Council on CardiovascularStroke Nursing; Council on Peripheral Vascular Disease;Council on Quality of CareOutcomes Research. Self-Care for the prevention and management of cardiovascular disease and stroke: a scientific statement for healthcare professionals from the American Heart Association. J Am Heart Assoc 2017 Aug 31;6(9):- [FREE Full text] [doi: 10.1161/JAHA.117.006997] [Medline: 28860232]
4. Barello S, Triberti S, Graffigna G, Libreri C, Serino S, Hibbard J, et al. eHealth for patient engagement: a systematic review. Front Psychol 2015;6:2013 [FREE Full text] [doi: 10.3389/fpsyg.2015.02013] [Medline: 26779108]
5. Fuster V, Kelly BB, Vedanthan R. Promoting global cardiovascular health: moving forward. Circulation 2011 Apr 19;123(15):1671-1678. [doi: 10.1161/CIRCULATIONAHA.110.009522] [Medline: 21502585]
6. Pietrzak E, Cotea C, Pullman S. Primary and secondary prevention of cardiovascular disease: is there a place for Internet-based interventions? J Cardiopulm Rehabil Prev 2014;34(5):303-317. [doi: 10.1097/HCR.0000000000000063] [Medline: 25079147]
7. Storm V, Dörenkämper J, Reinwand DA, Wienert J, De Vries H, Lippke S. Effectiveness of a web-based computer-tailored multiple-lifestyle intervention for people interested in reducing their cardiovascular risk: a randomized controlled trial. J Med Internet Res 2016 Apr 11;18(4):e78 [FREE Full text] [doi: 10.2196/jmir.5147] [Medline: 27068880]

8. Devi R, Singh SJ, Powell J, Fulton EA, Igbinedion E, Rees K. Internet-based interventions for the secondary prevention of coronary heart disease. Cochrane Database Syst Rev 2015(12):CD009386. [doi: 10.1002/14651858.CD009386.pub2] [Medline: 26691216]
9. Wantland DJ, Portillo CJ, Holzemer WL, Slaughter R, McGhee EM. The effectiveness of web-based vs non-web-based interventions: a meta-analysis of behavioral change outcomes. J Med Internet Res 2004 Nov 10;6(4):e40 [FREE Full text] [doi: 10.2196/jmir.6.4.e40] [Medline: 15631964]
10. Lie SS, Karlsen B, Oord ER, Graue M, Oftedal B. Dropout from an eHealth intervention for adults with type 2 diabetes: a qualitative study. J Med Internet Res 2017 May 30;19(5):e187 [FREE Full text] [doi: 10.2196/jmir.7479] [Medline: 28559223]
11. Beleigoli AM, Andrade AQ, Cançado AG, Paulo MN, Diniz MD, Ribeiro AL. Web-based digital health interventions for weight loss and lifestyle habit changes in overweight and obese adults: systematic review and meta-analysis. J Med Internet Res 2019 Jan 08;21(1):e298 [FREE Full text] [doi: 10.2196/jmir.9609] [Medline: 30622090]
12. Donkin L, Christensen H, Naismith SL, Neal B, Hickie IB, Glozier N. A systematic review of the impact of adherence on the effectiveness of e-therapies. J Med Internet Res 2011;13(3):e52 [FREE Full text] [doi: 10.2196/jmir.1772] [Medline: 21821503]
13. Moore GF, Audrey S, Barker M, Bond L, Bonell C, Hardeman W, et al. Process evaluation of complex interventions: medical research council guidance. Br Med J 2015;350:h1258 [FREE Full text] [Medline: 25791983]
14. Short C, Rebar A, Plotnikoff R, Vandelanotte C. Designing engaging online behaviour change interventions: a proposed model of user engagement. Eur Health Psychol 2015;17(1):32-38 [FREE Full text]
15. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Transl Behav Med 2016 Dec 13:-. [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
16. Jones JB, Weiner JP, Shah NR, Stewart WF. The wired patient: patterns of electronic patient portal use among patients with cardiac disease or diabetes. J Med Internet Res 2015 Feb 20;17(2):e42 [FREE Full text] [doi: 10.2196/jmir.3157] [Medline: 25707036]
17. Irizarry T, DeVito DA, Curran CR. Patient portals and patient engagement: a state of the science review. J Med Internet Res 2015;17(6):e148 [FREE Full text] [doi: 10.2196/jmir.4255] [Medline: 26104044]
18. Zvoch K. How does fidelity of implementation matter? Using multilevel models to detect relationships between participant outcomes and the delivery and receipt of treatment. Am J Eval 2012 Sep 26;33(4):547-565. [doi: 10.1177/1098214012452715]
19. Hasson H. Systematic evaluation of implementation fidelity of complex interventions in health and social care. Implement Sci 2010 Sep 03;5:67 [FREE Full text] [doi: 10.1186/1748-5908-5-67] [Medline: 20815872]
20. Holliday LR. Using logic model mapping to evaluate program fidelity. Stud Edu Eval 2014 Sep;42:109-117. [doi: 10.1016/j.stueduc.2014.04.001]
21. Carroll C, Patterson M, Wood S, Booth A, Rick J, Balain S. A conceptual framework for implementation fidelity. Implement Sci 2007 Nov 30;2:40 [FREE Full text] [doi: 10.1186/1748-5908-2-40] [Medline: 18053122]
22. Oakley A, Strange V, Bonell C, Allen E, Stephenson J. Process evaluation in randomised controlled trials of complex interventions. Br Med J 2006 Feb 18;332(7538):413-416 [FREE Full text] [doi: 10.1136/bmj.332.7538.413] [Medline: 16484270]
23. Redfern J, Usherwood T, Harris MF, Rodgers A, Hayman N, Panaretto K, et al. A randomised controlled trial of a consumer-focused e-health strategy for cardiovascular risk management in primary care: the Consumer Navigation of Electronic Cardiovascular Tools (CONNECT) study protocol. BMJ Open 2014 Jan 31;4(2):e004523 [FREE Full text] [doi: 10.1136/bmjopen-2013-004523] [Medline: 24486732]
24. Coorey GM, Neubeck L, Usherwood T, Peiris D, Parker S, Lau AYS, et al. Implementation of a consumer-focused eHealth intervention for people with moderate-to-high cardiovascular disease risk: protocol for a mixed-methods process evaluation. BMJ Open 2017 Dec 11;7(1):e014353 [FREE Full text] [doi: 10.1136/bmjopen-2016-014353] [Medline: 28077414]
25. Coorey G, Peiris D, Usherwood T, Neubeck L, Mulley J, Redfern J. Persuasive design features within a consumer-focused eHealth intervention integrated with the electronic health record: A mixed methods study of effectiveness and acceptability. PLoS One 2019;14(6):e0218447 [FREE Full text] [doi: 10.1371/journal.pone.0218447] [Medline: 31220127]
26. Coorey G, Peiris D, Neubeck L, Redfern J. A realist evaluation approach to explaining the role of context in the impact of a complex eHealth intervention for improving prevention of cardiovascular disease. BMC Health Serv Res 2020 Aug 18;20(1):764 [FREE Full text] [doi: 10.1186/s12913-020-05597-5] [Medline: 32811480]
27. Redfern J, Coorey G, Mulley J, Scaria A, Neubeck L, Hafiz N, et al. A digital health intervention for cardiovascular disease management in primary care (CONNECT) randomized controlled trial. NPJ Digit Med 2020;3:117 [FREE Full text] [doi: 10.1038/s41746-020-00325-z] [Medline: 32964140]
28. Neubeck L, Coorey G, Peiris D, Mulley J, Heeley E, Hersch F, et al. Development of an integrated e-health tool for people with, or at high risk of, cardiovascular disease: The Consumer Navigation of Electronic Cardiovascular Tools (CONNECT) web application. Int J Med Inform 2016 Dec;96:24-37. [doi: 10.1016/j.ijmedinf.2016.01.009] [Medline: 26847070]

29. Redfern J, Thiagalingam A, Jan S, Whittaker R, Hackett ML, Mooney J, et al. Development of a set of mobile phone text messages designed for prevention of recurrent cardiovascular events. Eur J Prev Cardiol 2014 Apr;21(4):492-499. [doi: 10.1177/2047487312449416] [Medline: 22605787]
30. Steckler A, Linnan L. (Eds) Process Evaluation for Public Health Interventions and Research. San Francisco, CA: Jossey-Bass; 2002:0787959766.
31. Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Research in Psychology 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
32. Australian Bureau of Statistics. Estimates of Aboriginal and Torres Strait Islander Australians. URL: https://www.abs.gov.au/statistics/people/aboriginal-and-torres-strait-islander-peoples/estimates-aboriginal-and-torres-strait-islander-australians/latest-release [accessed 2020-12-10]
33. Gearing RE, El-Bassel N, Ghesquiere A, Baldwin S, Gillies J, Ngeow E. Major ingredients of fidelity: a review and scientific guide to improving quality of intervention research implementation. Clin Psychol Rev 2011 Feb;31(1):79-88. [doi: 10.1016/j.cpr.2010.09.007] [Medline: 21130938]
34. Century J, Rudnick M, Freeman C. A framework for measuring fidelity of implementation: a foundation for shared language and accumulation of knowledge. Am J Eval 2010 May 11;31(2):199-218. [doi: 10.1177/1098214010366173]
35. Brouwer W, Kroeze W, Crutzen R, de Nooijer J, de Vries NK, Brug J, et al. Which intervention characteristics are related to more exposure to internet-delivered healthy lifestyle promotion interventions? A systematic review. J Med Internet Res 2011;13(1):e2 [FREE Full text] [doi: 10.2196/jmir.1639] [Medline: 21212045]
36. Nijland N, van Gemert-Pijnen JE, Kelders SM, Brandenburg BJ, Seydel ER. Factors influencing the use of a Web-based application for supporting the self-care of patients with type 2 diabetes: a longitudinal study. J Med Internet Res 2011;13(3):e71 [FREE Full text] [doi: 10.2196/jmir.1603] [Medline: 21959968]
37. Tsai R, Bell EJ, Woo H, Baldwin K, Pfeffer MA. How patients use a patient portal: an institutional case study of demographics and usage patterns. Appl Clin Inform 2019 Jan;10(1):96-102 [FREE Full text] [doi: 10.1055/s-0038-1677528] [Medline: 30727003]
38. Glasgow RE, Christiansen SM, Kurz D, King DK, Woolley T, Faber AJ, et al. Engagement in a diabetes self-management website: usage patterns and generalizability of program use. J Med Internet Res 2011;13(1):e9 [FREE Full text] [doi: 10.2196/jmir.1391] [Medline: 21371992]
39. O'Brien HL, Toms EG. What is user engagement? A conceptual framework for defining user engagement with technology. J Am Soc Inf Sci 2008 Apr;59(6):938-955. [doi: 10.1002/asi.20801]
40. Simblett S, Greer B, Matcham F, Curtis H, Polhemus A, Ferrão J, et al. Barriers to and facilitators of engagement with remote measurement technology for managing health: systematic review and content analysis of findings. J Med Internet Res 2018 Jul 12;20(7):e10480 [FREE Full text] [doi: 10.2196/10480] [Medline: 30001997]
41. Rogers E. Diffusion of Innovations. 4th edition. New York, NY: The Free Press; 1995.
42. Zvoch K. Treatment fidelity in multisite evaluation: a multilevel longitudinal examination of provider adherence status and change. Am J Eval 2009 Mar;30(1):44-61. [doi: 10.1177/1098214008329523]
43. Greenhalgh T, Russell J. Why do evaluations of eHealth programs fail? An alternative set of guiding principles. PLoS Med 2010;7(11):e1000360 [FREE Full text] [doi: 10.1371/journal.pmed.1000360] [Medline: 21072245]
44. O'Connor S, Hanlon P, O'Donnell CA, Garcia S, Glanville J, Mair FS. Understanding factors affecting patient and public engagement and recruitment to digital health interventions: a systematic review of qualitative studies. BMC Med Inform Decis Mak 2016 Sep 15;16(1):120 [FREE Full text] [doi: 10.1186/s12911-016-0359-3] [Medline: 27630020]
45. Fraccaro P, Vigo M, Balatsoukas P, Buchan IE, Peek N, van der Veer SN. The influence of patient portals on users' decision making is insufficiently investigated: a systematic methodological review. Int J Med Inform 2018 Mar;111:100-111 [FREE Full text] [doi: 10.1016/j.ijmedinf.2017.12.028] [Medline: 29425621]
46. van Vugt M, de Wit M, Sieverink F, Roelofsen Y, Hendriks SH, Bilo HJG, et al. Uptake and effects of the e-Vita personal health record with self-management support and coaching, for type 2 diabetes patients treated in primary care. J Diabetes Res 2016;2016:5027356 [FREE Full text] [doi: 10.1155/2016/5027356] [Medline: 26955640]
47. Aalbers T, Baars MA, Rikkert MG. Characteristics of effective internet-mediated interventions to change lifestyle in people aged 50 and older: a systematic review. Ageing Res Rev 2011 Sep;10(4):487-497. [doi: 10.1016/j.arr.2011.05.001] [Medline: 21628005]
48. Ludden GD, van Rompay TJ, Kelders SM, van Gemert-Pijnen JE. How to increase reach and adherence of web-based interventions: a design research viewpoint. J Med Internet Res 2015;17(7):e172 [FREE Full text] [doi: 10.2196/jmir.4201] [Medline: 26163456]
49. Norcross JC, Krebs PM, Prochaska JO. Stages of change. J Clin Psychol 2011 Feb;67(2):143-154. [doi: 10.1002/jclp.20758] [Medline: 21157930]
50. Mohr DC, Schueller SM, Montague E, Burns MN, Rashidi P. The behavioral intervention technology model: an integrated conceptual and technological framework for eHealth and mHealth interventions. J Med Internet Res 2014;16(6):e146 [FREE Full text] [doi: 10.2196/jmir.3077] [Medline: 24905070]

51. Tate DF, Jackvony EH, Wing RR. Effects of internet behavioral counseling on weight loss in adults at risk for type 2 diabetes: a randomized trial. J Am Med Assoc 2003 Apr 9;289(14):1833-1836. [doi: 10.1001/jama.289.14.1833] [Medline: 12684363]
52. Yardley L, Spring BJ, Riper H, Morrison LG, Crane DH, Curtis K, et al. Understanding and promoting effective engagement with digital behavior change interventions. Am J Prev Med 2016 Nov;51(5):833-842. [doi: 10.1016/j.amepre.2016.06.015] [Medline: 27745683]
53. Slaughter SE, Hill JN, Snelgrove-Clarke E. What is the extent and quality of documentation and reporting of fidelity to implementation strategies: a scoping review. Implement Sci 2015 Sep 07;10:129 [FREE Full text] [doi: 10.1186/s13012-015-0320-3] [Medline: 26345357]
54. Sieverink F, Kelders SM, van Gemert-Pijnen E. Clarifying the concept of adherence to eHealth technology: systematic review on when usage becomes adherence. J Med Internet Res 2017 Dec 06;19(12):e402 [FREE Full text] [doi: 10.2196/jmir.8578] [Medline: 29212630]

Abbreviations

CONNECT: Consumer Navigation of Electronic Cardiovascular Tools
CVD: cardiovascular disease
eHEALS: eHealth literacy scale
EHR: electronic health record
RCT: randomized controlled trial

Edited by R Kukafka; submitted 29.10.20; peer-reviewed by CY Ooi, A Bermon; comments to author 30.11.20; revised version received 21.12.20; accepted 16.03.21; published 26.04.21.

Please cite as:
Coorey G, Peiris D, Scaria A, Mulley J, Neubeck L, Hafiz N, Redfern J
An Internet-Based Intervention for Cardiovascular Disease Management Integrated With Primary Care Electronic Health Records: Mixed Methods Evaluation of Implementation Fidelity and User Engagement
J Med Internet Res 2021;23(4):e25333
URL: https://www.jmir.org/2021/4/e25333
doi:10.2196/25333
PMID:33900204

Original Paper

A Web-Based and In-Person Risk Reframing Intervention to Influence Mothers' Tolerance for, and Parenting Practices Associated With, Children's Outdoor Risky Play: Randomized Controlled Trial

Mariana Brussoni[1,2,3], PhD; Christina S Han[1], MA; Yingyi Lin[4], MSc; John Jacob[1], MSc, MBA; Ian Pike[1,3], PhD; Anita Bundy[5,6], PhD; Guy Faulkner[7], PhD; Jennifer Gardy[8], PhD; Brian Fisher[9], PhD; Louise Mâsse[2,3], PhD

[1]Department of Pediatrics, Faculty of Medicine, University of British Columbia, Vancouver, BC, Canada
[2]School of Population & Public Health, Faculty of Medicine, University of British Columbia, Vancouver, BC, Canada
[3]British Columbia Injury Research & Prevention Unit, British Columbia Children's Hospital Research Institute, Vancouver, BC, Canada
[4]Spatial Sciences Institute, University of Southern California, Los Angeles, CA, United States
[5]Department of Occupational Therapy, Colorado State University, Fort Collins, CO, United States
[6]Occupational Theory, Faculty of Health Sciences, University of Sydney, Sydney, Australia
[7]School of Kinesiology, Faculty of Medicine, University of British Columbia, Vancouver, BC, Canada
[8]Bill & Melinda Gates Foundation, Seattle, WA, United States
[9]School of Interactive Arts & Technology, Simon Fraser University, Surrey, BC, Canada

Corresponding Author:
Mariana Brussoni, PhD
Department of Pediatrics
Faculty of Medicine
University of British Columbia
F508 – 4480 Oak Street
Vancouver, BC, V6H 3V4
Canada
Phone: 1 6048753712
Email: mbrussoni@bcchr.ubc.ca

Abstract

Background: Outdoor risky play, such as climbing, racing, and independent exploration, is an important part of childhood and is associated with various positive physical, mental, and developmental outcomes for children. Parental attitudes and fears, particularly mothers', are a major deterrent to children's opportunities for outdoor risky play.

Objective: The aim of this study was to evaluate the efficacy of 2 versions of an intervention to reframe mothers' perceptions of risk and change parenting behaviors: a web-based intervention or an in-person workshop, compared with the control condition.

Methods: The Go Play Outside! randomized controlled trial was conducted in Canada from 2017 to 2018. Participants were recruited through social media, snowball sampling, and community notices. Mothers of children aged 6-12 years were self-assessed through eligibility questions, and those eligible and consented to participate in the study were randomized into a fully automated web-based intervention, the in-person workshop, or the control condition. The intervention was underpinned by social cognitive theory, incorporating behavior change techniques. Participants progressed through a series of self-reflection exercises and developed a goal for change. Control participants received the Position Statement on Active Outdoor Play. The primary outcome was increase in tolerance of risky play and the secondary outcome was goal attainment. Data were collected online via REDCap at baseline, 1 week, and 3 months after the intervention. Randomization was conducted using sealed envelope. Allocations were concealed to researchers at assignment and data analysis. We conducted mediation analyses to examine whether the intervention influenced elements of social cognitive theory, as hypothesized.

Results: A total of 451 mothers were randomized and completed baseline sociodemographic assessments: 150 in the web-based intervention, 153 in the in-person workshop, and 148 in the control condition. Among these, a total of 351 mothers completed the intervention. At 1 week after the intervention, 113, 85, and 135 mothers completed assessments for each condition, respectively, and at 3 months after the intervention, 105, 84, and 123 completed the assessments, respectively. Compared with mothers in the

control condition, mothers in the web-based intervention had significantly higher tolerance of risky play at 1 week (*P*=.004) and 3 months after the intervention (*P*=.007); and mothers in the in-person workshop had significantly higher tolerance of risky play at 1 week after the intervention (*P*=.02). No other significant outcomes were found. None of the potential mediators were found to significantly mediate the outcomes.

Conclusions: The trial demonstrates that the web-based intervention was effective in increasing mothers' tolerance for risk in play.

Trial Registration: ClinicalTrials.gov NCT03374683; https://clinicaltrials.gov/ct2/show/NCT03374683

International Registered Report Identifier (IRRID): RR2-10.1186/s13063-018-2552-4

(J Med Internet Res 2021;23(4):e24861) doi:10.2196/24861

KEYWORDS

outdoor play; mothering; independent mobility; physical activity; risk perception; risky play; risk reframing

Introduction

Evidence is growing regarding the importance of outdoor play for children's health and well-being, as are efforts to improve children's access to these opportunities [1]. Outdoor play includes inherent risks, including those explicitly sought out by children as they explore their bodies and environments. Risky play, such as climbing trees, building dens, or even walking home from school without an adult, is an inherent and important part of outdoor play [2], and has been associated with increases in physical activity, decreases in sedentary behavior, and positive influences on physical, social, and cognitive development [3,4]. Parental attitudes and fears, particularly mothers', are a primary barrier to children's outdoor risky play opportunities [5-7]. They include fear of serious injury, traffic, abduction, or even the belief that time spent in outdoor play has little value in contrast to academic or other pursuits. Efforts to shift parent attitudes are a frequent focus of practitioners wanting to promote high-quality play opportunities [8], and policy makers wanting to improve rates of children walking to and from school [9,10].

Previous attempts to influence parent attitudes and practices toward risky play have been limited to in-person workshops and have not been rigorously evaluated [11]. We sought to develop and evaluate a risk reframing intervention for parents, particularly mothers as they may be typically the more limiting parent [12,13], that would be accessible through either an in-person workshop or on a web-based platform. A web-based intervention that was freely available would facilitate parents' access, as well as the ease with which practitioners could incorporate the intervention into their existing practice by, for example, encouraging the parents in their network to complete the intervention prior to children's enrollment in activities.

This paper reports the results of a randomized controlled trial (RCT) evaluating the efficacy of a risk reframing intervention to increase mothers' tolerance for risky play and attain a behavior change goal related to providing risky play opportunities for their 6-12-year-old children. The intervention development, content, and theoretical framework were previously published [14]. We tested 2 versions of the intervention: a web-based and an in-person workshop. We hypothesized that participants in either intervention condition would have significantly greater increase of tolerance for risky play than those in the control condition at 1 week and 3 months after the intervention. We also hypothesized that a greater proportion of participants in either intervention condition would attain their behavior change goal than those in the control condition.

We further examined whether social cognitive theory (SCT), the behavior change model that underpinned the development of the intervention, produced the hypothesized effect on the outcome variables. Specifically, we hypothesized that self-efficacy, outcome expectations, and knowledge of risky play would mediate the relationship between the risk reframing intervention and the outcomes: tolerance for risky play and goal attainment.

Methods

Study Design

A description of the protocol for this study has been published [14]. Briefly, the study was a single-blind (researchers and outcome assessors), 3-parallel condition RCT. The trial was conducted between December 2017 and September 2018 in the Metro Vancouver area of British Columbia, Canada. Measures were collected at baseline, 1 week, and 3 months after the intervention. The primary outcome was increase in tolerance of risk in play at either follow-up time point. The secondary outcome was mothers' goal attainment at either follow-up time point.

The trial was registered with the United States National Institutes of Health's Protocol Registration and Results System (NCT03374683) and approved by the University of British Columbia/Children's and Women's Health Centre of British Columbia Research Ethics Board (H15-03271). No change was made to methods after trial commencement. We followed the CONSORT-EHEALTH guidelines in reporting this study [15]. The completed checklist is presented in Multimedia Appendix 1.

Participant Recruitment and Eligibility Criteria

Participants were recruited through advertising on online forums and social media, distributing notices through our networks, snowball sampling, and posting notices in community centers. Interested participants visited the study home page in the REDCap electronic data capture tool hosted at the British

Columbia Children's Hospital Research Institute [16]. There they were provided with a description of study procedures and information that completing the survey questions indicated consent. Potential participants were self-assessed through eligibility questions establishing that they were a mother with primary custody of a child/children aged 6-12 years; residing in the Metro Vancouver Regional District; and able to speak, read, and understand English. Computer/internet access and literacy were implicit eligibility criteria because accessing the study home page and completing the eligibility questionnaire would be otherwise impossible. Enrolled participants were emailed a unique link to the baseline questionnaire package in REDCap.

Randomization and Blinding

Enrolled participants were automatically assigned to 1 of the 3 conditions by REDCap: control, web-based, or in-person workshop. Participants had equal likelihood of being assigned to each condition (33%, ie, nearly 170). The randomization schedule was generated beforehand via sealed envelope [17] using randomized permuted blocks of size 3, 6, and 9. The list was then transferred to REDCap. Unbeknownst to participants, randomized allocation occurred in the background before participants completed the baseline questionnaire. This had to be done because REDCap limited capabilities for real-time, streamlined randomized allocation. The nature of the intervention did not permit participant blinding, but they were informed of their allocated treatment after completing the baseline questionnaires. The in-person workshop facilitator could not be blinded to allocation as the other 2 arms did not have a facilitator. Likewise, research staff who coordinated in-person workshop schedules could not be blinded to allocation of the in-person workshop. In-person workshop participant information (name and email address) was only used to coordinate the workshop. This information was encrypted, password protected, and stored in a password-protected folder in a secured network at the British Columbia Children's Hospital Research. However, allocations were concealed to the researchers at participant assignment and data analysis.

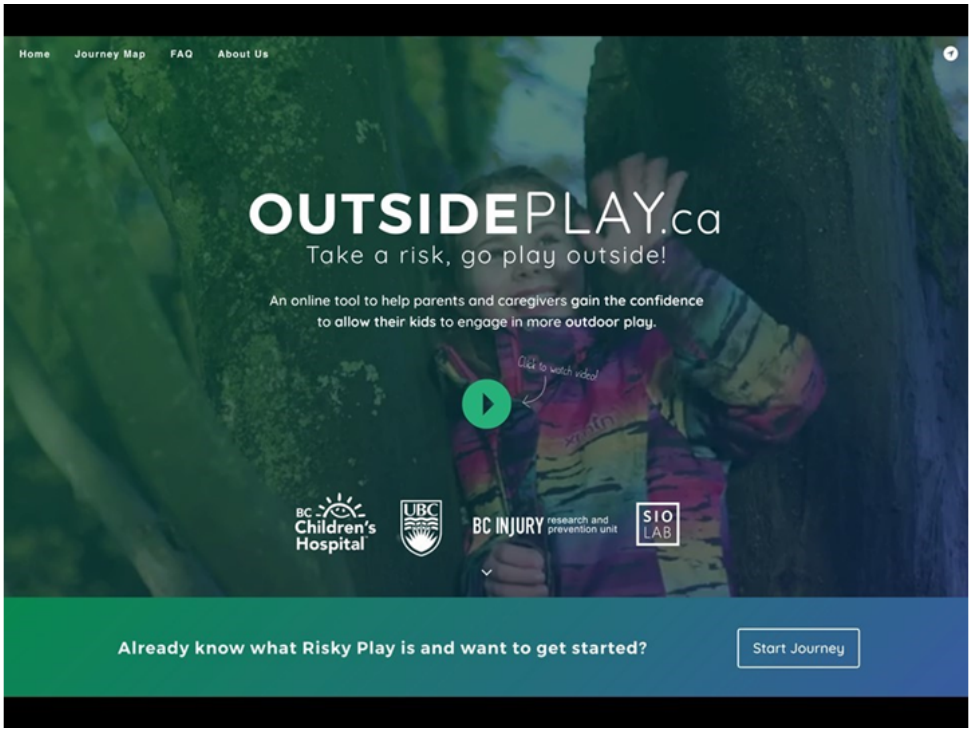
Risk Reframing Intervention

Participants in the web-based intervention were provided with a link to the fully automated web-based intervention [18] which they need to complete within 1 week. A reminder was sent via email if participants did not complete the web-based intervention within 24 hours. The final reminder was sent 48 hours after their baseline entry via email. Contact information (ie, email, phone) was provided for participants to contact the research coordinator if they had any feedback or questions. Participants in the in-person workshop were scheduled to attend the in-person workshop. Briefly, the risk reframing intervention was adapted from an in-person workshop for parents and teachers developed by Bundy and colleagues [8,19] using SCT [20] to incorporate health behavior change techniques (BCTs) as per Michie et al's taxonomy [21]. The published study protocol outlines each intervention task with the corresponding SCT construct and BCTs [14]. We sought to address common concerns about risky play and engage participants in self-reflection tasks to consider how these concepts applied to their parenting approach. The home page included a 2-minute video introduction to the topic and the tool, text defining outdoor and risky play, and outlining why they are important, as well as a brief description of the journey participants would follow in the tool. The logos of the British Columbia Children's Hospital, the University of British Columbia, British Columbia Injury Research and Prevention Unit, and the Digital Lab were prominently displayed on the home page. As per SCT, the home page focused on building knowledge and influencing outcome expectations. BCTs included credible source, and information about health, social, and emotional consequences. Participants then proceeded through 3 chapters. Chapter 1, *Reflection*, involved considering the values and traits they most desired for their child in adulthood, their child's favorite activities, their own favorite play activities at the same age, and what they got out of these childhood activities. As above, this chapter also focused on knowledge and outcome expectations. BCTs included framing/reframing and incompatible beliefs. Chapter 2, *What Would You Do?*, presented participants with 3 interactive video segments (climbing a tree, walking home from school, and building a den) and gave them the choice to allow or not allow the child to engage in the activity. If the participant chose to allow the activity, the child displayed excitement at the opportunity and a sense of achievement upon completion. If the participant chose to not allow the activity, the child displayed disappointment and dejection. Once the choice was made, the rest of the video played with the outcome of that choice. Participants were also asked to reflect on fears that influenced their choice, and things that helped them let go. In addition to the above SCT constructs (knowledge and outcome expectations), this chapter provided opportunities for observational learning, building self-efficacy, and identifying barriers and opportunities. BCTs included information about health, social and environmental, and emotional consequences; problem solving; demonstration of behavior; comparative imagining of future outcomes, framing/reframing; and focus on past success. Chapter 3, *Creating Your Plan*, allowed participants to review their journey and set a realistic goal as well as the timeline and steps to attain it. Participants could enter their own goal in a text box, or select from a list of suggested goals that included steps to attain it; for example, "letting my child play out in the yard without supervision" with the following steps: "Let your child play outside for a few minutes while you watch from the window. Gradually extend this time. Then try not watching out the window." This chapter reinforced the above SCT constructs, and also encouraged building self-efficacy and outlining intentions. BCTs included goal setting (behavior and outcome), problem solving, action planning, demonstration of behavior, prompts/cues, graded tasks, comparative imagining of future outcomes, framing/reframing, and incompatible beliefs.

The intervention was developed by the study authors. Once the web-based platform was complete, the in-person workshop presentation (prepared in Microsoft PowerPoint) and facilitator and participant manuals were developed using images from the web-based platform. Intervention content was frozen during the trial. The web-based intervention is available online [18]. Screenshot of the landing page can be seen in Figure 1 and the complete screenshots of the web-based intervention can be

accessed in Multimedia Appendix 2. The in-person workshop materials can also be accessed in Multimedia Appendix 3.

Figure 1. Intervention landing page.



Participants using the web-based intervention took between 15 and 45 minutes to complete it, depending on their movement through each task. The in-person workshops lasted approximately 45 minutes to 1.5 hours, depending on participant discussion. Participants in the control condition took 15-20 minutes to complete.

Comparison Condition

Participants in the control condition were asked to review an online version of the Position Statement on Active Outdoor Play, which includes information on research and recommendations for action in addressing barriers to outdoor play [1,22].

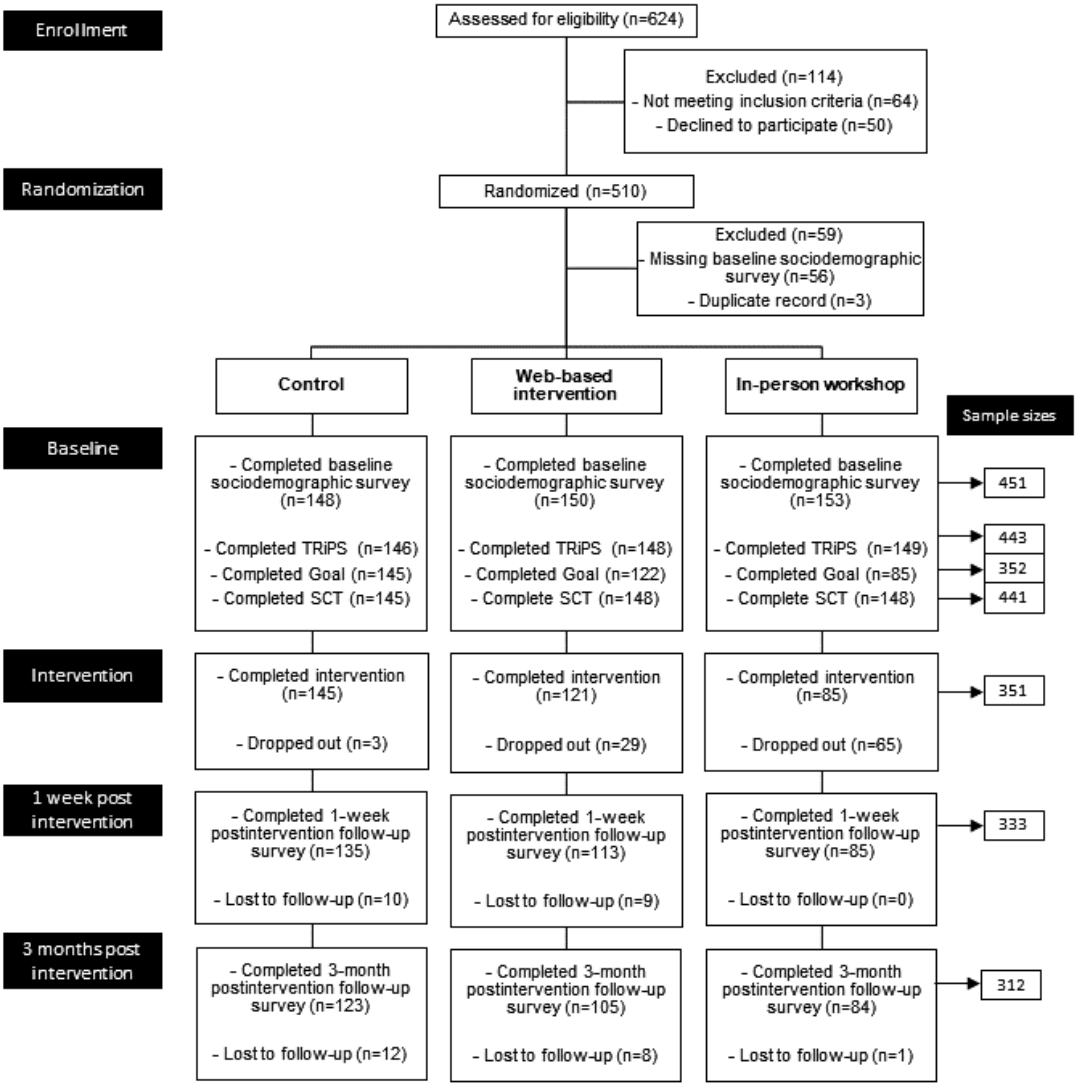
Outcome Measures

Measurements were taken at baseline, 1 week, and 3 months after the intervention. Participants in all conditions received an honorarium of Can $30 (US $24) at baseline and Can $15 (US $12) at each follow-up as a compensation for participation. Nonrespondents received 2 email reminders to complete survey data. Participants attending in-person workshops were provided with an additional honorarium of Can $30 (US $24) to compensate for expenses incurred in attending, such as travel or childcare.

The primary outcome measure was increase in the total score on the Tolerance for Risk in Play Scale (TRiPS), a 31-item measure examining adults' tolerance of risk during children's play, based on Sandseter's 6-category model of risky play [23]. An earlier version of the TRiPS scale has been validated [24]. We obtained the scale from the author AB and assessed its psychometric properties in our sample using Rasch analysis. This analysis was conducted using mirt package in R software [25]. Rasch analysis of the baseline data (N=443 completed TRiPS; Figure 2) resulted in dropping 1 item (Do you allow this child to play-fight, testing who is strongest?) due to local dependence. The remaining 30 items resulted in the following model fit: root mean square error of approximation (RMSEA)=0.051 (90% CI 0.047-0.056), standardized root mean square residual (SRMSR)=0.089, Tucker–Lewis index (TLI)=0.874, comparative fit index (CFI)=0.874, and empirical reliability=0.789. Theta standardized scores from the Rasch analysis of the final 30-item TRiPS scale ranged from –3.372 to 1.975, with a mean of 0.000 (SD 0.974). A higher standardized score indicates higher tolerance of risky outdoor play.

The secondary outcome measure was self-reported behavior change, measured by participants' self-reported progress on attaining the goal they set for themselves within the risk reframing intervention. At each follow-up, participants were reminded of their goal and asked "Did you accomplish your goal?" with "Yes" and "No" response options.

Figure 2. CONSORT flowchart.



Behavior Change Model

To assess whether the effect of the intervention was mediated by SCT as we had theorized [14], we developed measures for self-efficacy, outcome expectations, and knowledge of risky play. All measures were previously published [14]. Confirmatory factor analyses were performed on the baseline data (n=441 completed all SCT construct measures; Figure 2) to test the psychometric properties of the 3 measures, resulting in the following: (1) 4-item self-efficacy, RMSEA=0.075 (90% CI 0.020-0.139), CFI=0.987, and SRMSR=0.023; (2) 7-item outcome expectations, RMSEA=0.107 (90% CI 0.085-0.129), CFI=0.975, and SRMSR=0.023 (2 items were dropped: "promote MY CHILD's self-confidence" and "help MY CHILD become more imaginative" due to high modification indices); and (3) 7-item knowledge of risky play: RMSEA=0.129 (90% CI 0.108-0.151), CFI=0.965, and SRMSR=0.026 (2 items were dropped: "promote A CHILD's self-confidence" and "help A CHILD become more imaginative" due to high modification indices). Average scores of all remaining items were calculated for each SCT construct, respectively.

Statistical Analyses

All statistical analyses were conducted in Stata 15 (StataCorp) [26].

Power

The TRiPS is scored on a logit scale and previous research indicated that scores on the TRiPS for a sample of parents of children aged 5-13 range from 0.20 to about 1.95 with SDs in the range of 1.78-1.82 [24]. With a sample size of at least 81 mothers in each condition, a test that averaged the differences in TRiPS score from baseline to the first assessment would have 80% power at a .05 level of significance to detect a difference of 0.75 with the control condition when the SD is 1.82 and the correlation between repeated observations is 0.75.

Descriptive Analysis

To compare sociodemographic differences between conditions, for continuous variables, one-way ANOVA was used, or Kruskal–Wallis H test (if variance was not equal between conditions). For categorical variables, chi-square test was used, or Fisher exact test if single-cell numbers were small. One-way ANOVA was used to compare TRiPS scores between different

conditions at different time points. Significance level was set at $P<.05$.

Treatment Effect of the Intervention

Linear and generalized linear mixed effects models with random intercepts and unstructured covariance were fit to analyze the effects of the intervention on TRiPS scores and goal accomplishment, respectively. In other words, the mixed effects regression analysis examined (1) whether TRiPS scores changed 1 week and 3 months after the intervention, and (2) whether these changes were greater in either experimental condition (web-based intervention or in-person workshop) compared with the control condition. Intent-to-treat analysis of TRiPS scores used last-observation-carried forward as the method of imputation, because missing data were primarily in the in-person workshop condition. Because these participants only completed baseline measures and did not receive the intervention, it is reasonable to expect their scores to remain the same throughout the study. Unstandardized (ie, raw) beta coefficients (β) were reported, which are interpreted as the change of TRiPS scores when comparing the experimental conditions with the control groups at baseline.

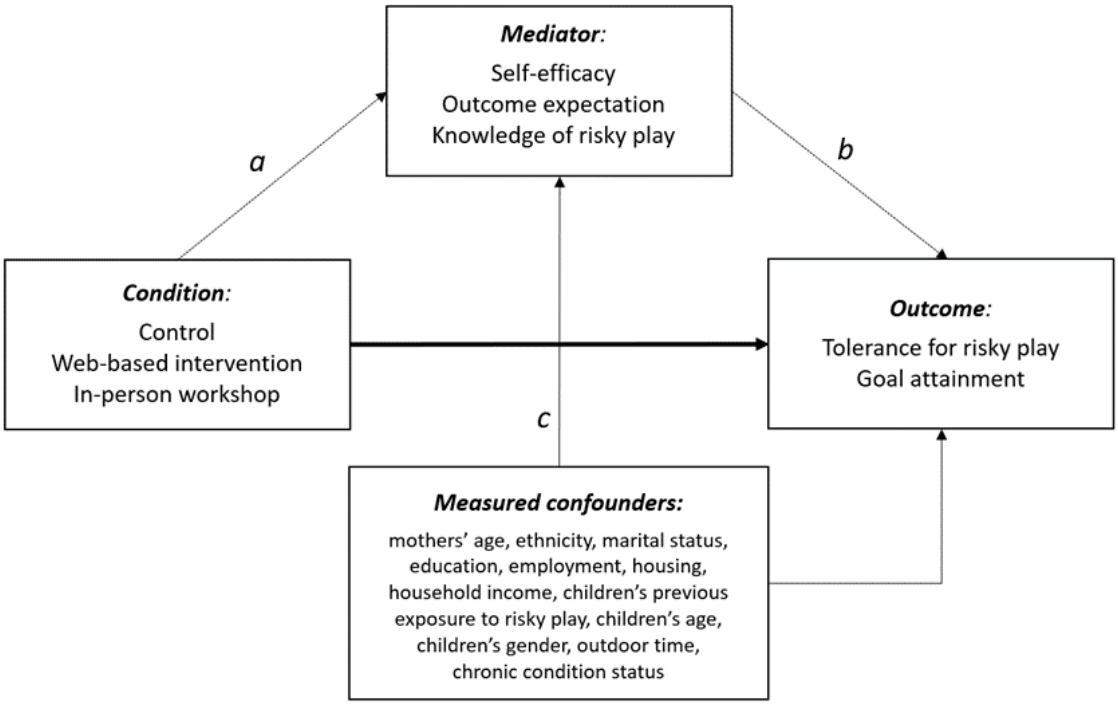
Similar to the TRiPS analyses, we conducted a generalized mixed effects regression analysis to examine the effect of the intervention on goal accomplishment, when comparing the control condition at 1 week with either of the experimental conditions at 3 months' follow-up. Intent-to-treat analysis of goal accomplishment was not performed due to the absence of baseline data. To establish a goal, participants had to complete either intervention. As a result, there was no basis to impute values of goal accomplishment. Odds ratios (ORs) were reported, which are interpreted as the odds of attaining goals for the experimental conditions at 3 months' follow-up, when comparing with the control conditions at 1 week after the intervention.

All models were adjusted by parents' age, ethnicity, marital status, education, employment, housing, and household income; and children's previous exposure to risky play, age, gender, weekday/weekend outdoor time, and chronic condition status.

Behavior Change Model Testing

To test whether SCT produced the hypothesized effect on the outcomes, we tested whether the effect of the intervention was mediated by self-efficacy, outcome expectations, and knowledge of risky play, as outlined in Figure 3. We followed the steps of mediation analysis for RCTs suggested by Whittle et al [27]. First, univariable linear regression models were fitted to the potential mediators to test whether there was an association between treatment and each mediator, respectively. Second, because a variable can only be a mediator of treatment, if there is a significant effect ($P<.05$) of treatment on the mediator (path *a*), a mediation analysis in the second step was only fitted to variables that were significantly associated with treatment. In the second step, the test of indirect (mediating) effect was performed by fitting regression models to the outcome, with treatment and the significant mediator found in step 1 included as covariates (path *ab*). This step of analysis controlled for baseline covariates, attempting to control for these as potential confounders in order to add robustness to our analyses. These confounders included mothers' age, ethnicity, marital status, education, employment, housing, household income; and children's previous exposure to risky play, children's age and gender, weekday/weekend outdoor time, and chronic condition status. Similar to the TRiPS analyses, intent-to-treat analysis was conducted using the last-observation-carried forward.

Figure 3. Mediation analysis.



Results

Overview

Figure 2 shows the study flow diagram. Recruitment began on December 1, 2017, and closed on June 18, 2018, once we had a minimum of 81 participants in each condition with complete data. A total of 510 mothers of a child between 6 and 12 years old residing in Metro Vancouver were automatically allocated to 1 of the 3 conditions by REDCap. Of these, 351 completed the intervention. While randomization produced roughly equal numbers of participants allocated to each condition, the in-person workshop condition experienced the most drop-outs (n=65). This condition involved the most time commitment for the participant and produced scheduling challenges. As such, despite the additional Can $30 (US $24) incentive to participate, it was most difficult to maintain participants in this condition. However, of the mothers attending the workshop, only 1 was lost to follow-up. We were able to ensure fidelity to the web-based intervention as we received a summary of all responses when a participant completed the intervention. These were reviewed to ensure full completion. Fidelity to the in-person workshop was established through attendance to the workshop.

Participant Characteristics

Baseline sociodemographic data from 451 mothers were included in our analyses. Participant demographics are displayed in Table 1. There were no statistically significant differences between conditions with respect to demographic characteristics at baseline (see the "*P* value column" in Table 1). There were statistically significant differences between conditions for child's disability/chronic condition at both 1 week (*P*=.03) and 3 months (*P*=.03) after the intervention, with the percentage of children having chronic conditions in the web-based intervention higher than that in the other 2 conditions (10/113 [8.8%] for 1 week after the intervention, and 9/105 [8.6%] for 3 months after the intervention, respectively). We also further compared the sociodemographic characteristics between those who completed the intervention (N=351) and those who did not complete the intervention (N=100), and found that employed mothers were more likely to complete the intervention (75.8% [266/351] versus 65.0% [65/100], respectively, *P*=.03). There were no statistical differences for other sociodemographic characteristics (see Multimedia Appendix 4).

Table 1. Participant demographics at baseline, after randomization.

Demographics	Control	Web-based intervention	In-person workshop	*P* value
Participants who completed the baseline sociodemographic assessment (N=451), n	148	150	153	.09
Age (N=450)[a], mean (SD)	40.7 (5.3)	40.8 (5.5)	39.6 (5.0)	
Ethnicity (N=451), n (%)				.24
White	101 (68.2)	112 (74.7)	117 (76.5)	
Others	47 (31.8)	38 (25.3)	36 (23.5)	
Marital status (N=451), n (%)				.39
Married/Common-law	118 (79.7)	125 (83.3)	131 (85.6)	
Others	30 (20.3)	25 (16.7)	22 (14.4)	
Education (N=447)[b], n (%)				.32
Less than university/college	36 (24.3)	36 (24.3)	33 (21.9)	
University/college	66 (44.6)	80 (54.1)	72 (47.7)	
More than university/college	46 (31.1)	32 (21.6)	46 (30.5)	
Employment (N=451), n (%)				.35
Employed for wages/self-employed	115 (77.7)	107 (71.3)	44 (28.8)	
Unemployed	33 (22.3)	43 (28.7)	109 (71.2)	
Home dwelling (N=451), n (%)				.66
Single detached	69 (46.6)	77 (51.3)	72 (47.1)	
Others	79 (53.4)	73 (48.7)	81 (52.9)	
Income (N=451), n (%)[c]				.80
<Can $63,300	36 (24.3)	40 (26.7)	33 (21.6)	
Can $63,300-Can $103,299	49 (33.1)	40 (26.7)	50 (32.7)	
≥Can $103,300	52 (35.1)	54 (36.0)	55 (35.9)	
Prefer not to answer	11 (7.4)	16 (10.7)	15 (9.8)	
Exposure to risky play information (N=451), n (%)				.15
No	22 (14.9)	28 (18.7)	36 (23.5)	
Yes	126 (85.1)	122 (81.3)	117 (76.7)	
Child age (N=451), mean (SD)	8.4 (1.7)	8.1 (1.7)	8.0 (1.9)	.10
Child sex (N=447), n (%)[d]				.78
Boy	85 (58.6)	82 (54.7)	85 (55.9)	
Girl	60 (41.4)	68 (45.3)	67 (44.1)	
Child's disability/chronic condition (N=451)[e], n (%)				.30
No	143 (96.6)	140 (93.3)	148 (96.7)	
Yes	5 (3.4)	10 (6.7)	5 (3.3)	
Outdoor time (hours) (N=451)[f], mean (SD)				
Weekday	3.0 (3.6)	2.8 (3.2)	2.6 (2.8)	.90
Weekend	2.7 (2.2)	2.8 (2.0)	2.9 (2.8)	.65

[a]One parent reported age=7 years, which is not reasonable, so we treated that data as missing.

[b]Prefer not to answer, n=4 (n=2 for web-based intervention; n=2 for in-person workshop).

[c]Can $ 1=US $0.80.

[d]A total of four children did not have information regarding their sex (n=3 in the control group, n=1 in the in-person workshop).

[e]Fisher exact test due to small sample size in single cell.

[f]Kruskal–Wallis H test due to non-equal variance between conditions.

Primary Outcome: TRiPS

Table 2 presents the description of TRiPS scores by treatment conditions and time point, without accounting for treatment effects nor adjusting for sociodemographic characteristics. There were no statistical differences in TRiPS scores between different conditions at each time point. Table 3 describes findings of the mixed effects regression analysis. Mothers who completed the web-based intervention condition had significantly higher TRiPS scores than mothers who completed the control condition at 1 week (=.26, 95% CI 0.09-0.43; *P*=.003) and 3 months (=.24, 95% CI 0.06-0.42; *P*=.01) after the intervention, indicating sustained change. Mothers who completed the in-person workshop condition had significantly higher increases in TRiPS scores than those in the control condition at 1 week after the intervention (=.19, 95% CI 0.06-0.38; *P*=.04). No statistically significant differences were found when comparing mothers who completed the in-person workshop condition with those in the control condition at 3 months after the intervention (=.09, 95% CI –0.10 to 0.29; *P*=.33).

Results of the intention-to-treat analyses for the effects of the intervention on TRiPS score largely replicated the analyses described above, indicating that mothers in the web-based intervention condition were significantly more likely to increase their TRiPS scores at 1 week (=.25, 95% CI 0.08-0.42; *P*=.004) and 3 months (=.24, 95% CI 0.06-0.42; *P*=.007) after the intervention compared with those in the control condition. Mothers in the in-person workshop condition had significantly higher increases in TRiPS scores than those in the control condition at 1 week after the intervention (=.22, 95% CI 0.03-0.40; *P*=.02). No statistically significant differences were found when comparing mothers in the in-person workshop with those in the control condition at 3 months after the intervention (=.13, 95% CI –0.06 to 0.32; *P*=.17).

Table 2. Description of TRiPS[a] scores by treatment conditions and time points.

Evaluation period	Sample size, n	Control, mean (SD)	Web-based intervention, mean (SD)	In-person workshop, mean (SD)	*P* value for 1-way ANOVA
Baseline	443	0.05 (1.03)	–0.11 (0.94)	0.06 (0.94)	.24
Completed intervention	351	0.05 (1.03)	–0.14 (0.97)	0.18 (0.87)	.05
1 week after the intervention	333	–0.09 (1.03)	–0.06 (1.05)	0.22 (0.80)	.06
3 months after the intervention	312	–0.03 (0.92)	–0.09 (0.93)	0.15 (0.80)	.19

[a]TRiPS: Tolerance for Risk in Play Scale.

Table 3. Results of mixed effects regression analysis for change in TRiPS[a] scores by treatment condition and time point.

Regression and group comparisons	Coefficients (95% CI)	*P* value for coefficients	*P* value for joint test of main effects
Raw TRiPS theta scores (N=351, for those who completed the intervention)[b]			
Treatment effects			.21
Web based versus control	–0.14 (–0.36 to 0.08)	.21	
In-person versus control	0.09 (–0.15 to 0.34)	.46	
Time effects			.86
1 week versus baseline	–0.16 (–0.27 to –0.04)	.007	
3 months versus baseline	–0.13 (–0.25 to –0.01)	.03	
Treatment by time effects			.02
Web based versus control by 1 week versus baseline	0.26 (0.09 to 0.43)	.003	
Web based versus control by 3 months versus baseline	0.24 (0.06 to 0.42)	.01	
In-person versus control by 1 week versus baseline	0.19 (0.06 to 0.38)	.04	
In-person versus control by 3 months versus baseline	0.10 (0.00 to 0.29)	.33	
Intention-to-treat analysis (imputed TRiPS theta scores) (N=443, for those who were randomized to a condition and completed baseline sociodemographic and TRiPS survey)[c]			
Treatment effects			.47
Web based versus control	–0.13 (–0.34 to 0.07)	.20	
In-person versus control	0.00 (–0.21 to 0.21)	.98	
Time effects			.90
1 week versus baseline	–0.16 (–0.27 to –0.04)	.008	
3 months versus baseline	–0.14 (–0.26 to –0.02)	.02	
Treatment by time effects			.01
Web based versus control by 1 week versus baseline	0.25 (0.08 to 0.42)	.004	
Web based versus control by 3 months versus baseline	0.24 (0.06 to 0.42)	.007	
In-person versus control by 1 week versus baseline	0.22 (0.03 to 0.40)	.02	
In-person versus control by 3 months versus baseline	0.13 (–0.06 to 0.32)	.17	

[a]TRiPS: Tolerance for Risk in Play Scale.

[b]Significant ($P<.10$) sociodemographic predictors include ethnicity, housing condition, previous exposure to risky play information, child's age, child's disability/chronic conditions, and weekend outdoor time.

[c]Significant ($P<.10$) sociodemographic predictors include ethnicity, parental employment status, housing condition, previous exposure to risky outdoor play information, child's age, child's disability/chronic conditions, and weekend outdoor time.

Secondary Outcome: Goal Attainment

Table 4 presents the results of the generalized mixed effects regression analysis. There was no statistical difference in goal attainment between mothers in the web-based intervention and the control condition at 3 months after the intervention as compared with 1 week after the intervention (OR 0.59, 95% CI 0.18-1.91; $P=.37$). There was also no statistical difference in goal attainment between mothers in the in-person workshop and the control condition at 3 months after the intervention as compared with 1 week after the intervention (OR 0.55, 95% CI 0.17-1.76; $P=.31$).

Table 4. Results of the mixed effects regression analysis for goal attainment by treatment condition and time.[a]

Regression and group comparisons	Odds ratios (95% CI)	*P* value for coefficients	*P* value for joint test of main effects
Treatment effects			.004
Web based versus control	3.36 (1.32-8.54)	.01	
In-person versus control	0.72 (0.28-1.85)	.49	
Time effects			<.001
3 months versus 1 week	7.27 (3.11-17.02)	<.001	
Treatment by time effects			.53
Web based versus control by 3 months versus 1 week	0.59 (0.18-1.91)	.37	
In-person versus control by 3 months versus 1 week	0.55 (0.17-1.76)	.31	

[a]Significant (*P*<.05) demographic predictors include ethnicity, child's disability/chronic conditions, and weekend outdoor time.

Behavior Change Model Mediation Analysis

Because only significant intervention effects were observed for TRiPS score and not goal attainment, mediation analyses were only conducted for TRiPS data. With regard to the direct effects of treatment conditions on the 3 SCT constructs, none of the 3 SCT constructs (self-efficacy, *P*=.16; outcome expectations, *P*=.24; and knowledge of risky play, *P*=.12) was associated with treatment exposures in the unadjusted models, neither for those who completed the intervention (N=351) nor for the intention-to-treat analysis sample (N=441, for those who completed baseline sociodemographic, TRiPS, and SCT survey; all *P* values were >.05). As there was no direct association between treatment and the SCT constructs (path *a*), further mediation analyses were not conducted because it was not possible for the SCT constructs to have an impact on the study outcome via mediation (ie, path *ab*). Therefore, none of the potential mediators were found to significantly mediate outcome.

Discussion

Principal Findings

Our hypotheses were partially supported. Mothers receiving both the web-based intervention and the in-person workshop intervention at 1 week after the intervention reported significantly higher increases in their tolerance for risky play than mothers in the control condition at baseline. These differences remained significant at 3 months after the intervention for mothers receiving the web-based intervention but not mothers receiving the in-person workshop intervention. There were no significant differences in goal attainment.

We did not have sufficient statistical power to test the difference in efficacy between the 2 versions of the intervention. In any case, the significant results for the web-based intervention and not for the in-person workshop at 3 months after the intervention are unexpected, as the in-person workshops provided arguably a higher intervention dose given the greater length required in participation and the more social aspect of the experience that could influence participant perceptions of social support, a construct of SCT [20]. Workshop participation required fitting into a set schedule and demanded significantly more time commitment and effort than the other conditions. Thus, it required a higher motivation level and had greater attrition. Those who ended up attending the in-person workshop might have already been fully aware of the benefits of children's outdoor play, which inadvertently left limited room for improvement. However, this possibility is not supported by baseline scores on the TRiPS, which did not differ between conditions. Another possible explanation could result from the fact that the number of participants attending each workshop varied considerably (2-12 participants), which could negatively impact the extent of discussion [28]. Further, workshops were facilitated by a professional facilitator, without content expertise in outdoor play and who was not a parent. SCT stresses the importance of relatable peers modeling behavior to encourage behavior change [28]. Thus, appropriate probing to foster discussion and participants' engagement with the topic may have been hampered.

It is not readily apparent why there was a null finding for goal attainment. It is possible that the binary yes/no outcome may have hampered participant responses in that participants may not have indicated they had attained their goal unless they perceived that all aspects were complete. In addition, goals may have been overly ambitious and did not have clear steps to attain them, such as "give my child more independence to enjoy freedom." Future iterations of the intervention should encourage users to set a more realistic and actionable goal. Sample actionable goals were provided in the tool but it is possible that these required further details on actions.

In sum, our findings indicate the efficacy of both versions of the risk reframing intervention for changing mothers' tolerance to risky play, with the web-based intervention displaying long-term effects.

Behavior Change Model

The hypothesized relationship between the constructs of SCT (self-efficacy, outcome expectation, and knowledge) and the intervention outcomes was not supported. We offer the following explanations for these findings. First, the intervention may have been of insufficient intensity or duration to influence these constructs. To reduce access and engagement barriers, we sought to develop an intervention that was efficient, would require limited time commitment, and would not necessitate repeat visits [14]. Therefore, future iterations may attempt to increase

the intervention dose to determine whether this would impact the mediators as we hypothesized. Second, it is possible that our SCT constructs were not sensitive enough to detect change over time. While we evaluated the psychometric properties of our measures, we do not know whether these measures are sensitive to change as they were developed for this study. Third, although we hypothesized that both self-efficacy and outcomes expectations would increase as a result of the intervention, it is possible that going through the intervention and attempting to change parenting practices proved more difficult than anticipated for some parents and might explain why no change was observed for the SCT constructs. We did not collect data on other potential factors that could influence parents' success, such as self-efficacy in overcoming barriers, thus limiting our insight on these results.

Strengths and Limitations

Our research is the first to use health behavior change theory and BCTs to develop an intervention to reframe mothers' perceptions and influence their parenting behaviors related to outdoor risky play. As recommended by published guidelines [29,30], we are also among the first to test the active ingredients of the intervention, examining the hypothesized relationships between SCT constructs and the outcomes. In addition, our use of RCT methods represents a gold-standard methodological evaluation technique. Previous research reported on but did not evaluate an in-person workshop intervention separately from a loose parts intervention in the playground [11]. Furthermore, our RCT evaluated 2 versions of the risk reframing intervention, providing insight into the efficacy of alternative delivery methods with important implications for practice.

There were several limitations to the study. Given the nature of the intervention, it was not possible to blind participants to their allocation, thus potentially introducing sources of bias. Furthermore, the intervention was developed based on research conducted in Western settings and was only available in English. As such, it may not reflect the perspectives and needs of other cultural and linguistic conditions and non-Western settings. Likewise, because one of our conditions necessitated in-person participation, we were only able to recruit participants within a limited geographic area, thus potentially limiting the applicability of our findings to other areas, such as rural communities.

Conclusions

Our findings provide confidence in encouraging use and broad dissemination of the web-based intervention. Given the ease of distribution, no cost to users, and low resource requirement for ongoing maintenance of the web-based tool, it is encouraging that this was an effective model and can provide the basis for further iterations and versions. Future research is necessary to examine the risk perceptions of parents from other cultural conditions and settings to facilitate the development of interventions that are applicable to their settings and realities.

Acknowledgments

This study was funded by the Lawson Foundation (Grant Nos GRT 2015-68 and GRT 2016-72). The Lawson Foundation was not involved in any aspect of study design or writing of the manuscript. MB and LM are supported by salary awards from the British Columbia Children's Hospital Research Institute. GF is supported by a Canadian Institutes of Health Research-Public Health Agency of Canada Chair in Applied Public Health. We thank Kathryn Soo and Jaspreet Dhillon, our research work study students who made the project possible. We also thank Juan Tellez, Kevin Quach, and Fahad Hasany from the Digital Lab at the University of British Columbia Department of Pediatrics for providing partnership and technical support in the development and programing of the web-based tool. We thank Dr Anita Niehues for her ideas and contributions to the development of the web-based intervention. We also thank Pam Fuselli, Heidi Campbell, Pierre Harrison, Brandy Tannenbaum, and Dr Guylaine Chabot for their review of and comments on an earlier version of the in-person workshop materials. We are further grateful to the Lawson Foundation for funding tool development and evaluation and to Christine Alden for her personal engagement in the project. Finally, we would like to express our gratitude to the participants for sharing their perspectives.

Authors' Contributions

MB conceived the study. AB and IP assisted MB with adaptation of the original risk reframing workshop for the purposes of the study. JJ led the development of the web-based platform. CH and MB developed the in-person workshop presentation and manuals. MB and LM led development of the study design, with contribution from GF, CH, and the remaining authors. MB wrote the first and subsequent drafts of this manuscript. LM and YL advised on statistical analysis. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT E-HEALTH checklist.
[PDF File (Adobe PDF File), 2730 KB - jmir_v23i4e24861_app1.pdf]

Multimedia Appendix 2

Screenshots of the web-based risk reframing intervention at OutsidePlay.ca.
[PDF File (Adobe PDF File), 4602 KB - jmir_v23i4e24861_app2.pdf]

Multimedia Appendix 3
Risk reframing intervention in-person workshop.
[PDF File (Adobe PDF File), 8681 KB - jmir_v23i4e24861_app3.pdf]

Multimedia Appendix 4
Nonstatistical differences for other sociodemographic characteristics.
[PDF File (Adobe PDF File), 55 KB - jmir_v23i4e24861_app4.pdf]

References

1. Tremblay MS, Gray C, Babcock S, Barnes J, Bradstreet CC, Carr D, et al. Position Statement on Active Outdoor Play. Int J Environ Res Public Health 2015 Jun 08;12(6):6475-6505 [FREE Full text] [doi: 10.3390/ijerph120606475] [Medline: 26062040]
2. Hansen Sandseter EB. Categorising risky play—how can we identify risk‐taking in children's play? European Early Childhood Education Research Journal 2007 Jun;15(2):237-252. [doi: 10.1080/13502930701321733]
3. Gray C, Gibbons R, Larouche R, Sandseter E, Bienenstock A, Brussoni M, et al. What Is the Relationship between Outdoor Time and Physical Activity, Sedentary Behaviour, and Physical Fitness in Children? A Systematic Review. Int J Environ Res Public Health 2015 Jun 08;12(6):6455-6474 [FREE Full text] [doi: 10.3390/ijerph120606455] [Medline: 26062039]
4. Brussoni M, Gibbons R, Gray C, Ishikawa T, Sandseter EBH, Bienenstock A, et al. What is the Relationship between Risky Outdoor Play and Health in Children? A Systematic Review. Int J Environ Res Public Health 2015 Jun 08;12(6):6423-6454 [FREE Full text] [doi: 10.3390/ijerph120606423] [Medline: 26062038]
5. Lee H, Tamminen KA, Clark AM, Slater L, Spence JC, Holt NL. A meta-study of qualitative research examining determinants of children's independent active free play. Int J Behav Nutr Phys Act 2015 Jan 24;12:5 [FREE Full text] [doi: 10.1186/s12966-015-0165-9] [Medline: 25616690]
6. Jelleyman C, McPhee J, Brussoni M, Bundy A, Duncan S. A Cross-Sectional Description of Parental Perceptions and Practices Related to Risky Play and Independent Mobility in Children: The New Zealand State of Play Survey. Int J Environ Res Public Health 2019 Jan 17;16(2):262 [FREE Full text] [doi: 10.3390/ijerph16020262] [Medline: 30658496]
7. Boxberger K, Reimers A. Parental Correlates of Outdoor Play in Boys and Girls Aged 0 to 12-A Systematic Review. Int J Environ Res Public Health 2019 Jan 11;16(2):190 [FREE Full text] [doi: 10.3390/ijerph16020190] [Medline: 30641874]
8. Niehues A, Bundy A, Broom A, Tranter P. Reframing healthy risk taking: Parents' dilemmas and strategies to promote children's well-being. Journal of Occupational Science 2016 Jul 24;23(4):449-463. [doi: 10.1080/14427591.2016.1209424]
9. McDonald N, Aalborg A. Why Parents Drive Children to School: Implications for Safe Routes to School Programs. Journal of the American Planning Association 2009 Jun 30;75(3):331-342. [doi: 10.1080/01944360902988794]
10. Council of Chief Medical Officers of Health. Active Outdoor Play Statement from the Council of Chief Medical Officers of Health. 2018. URL: http://www.phn-rsp.ca/aop-position-jae/index-eng.php#fn1 [accessed 2021-04-09]
11. Niehues A, Bundy A, Broom A, Tranter P, Ragen J, Engelen L. Everyday uncertainties: reframing perceptions of risk in outdoor free play. Journal of Adventure Education & Outdoor Learning 2013 Sep;13(3):223-237. [doi: 10.1080/14729679.2013.798588]
12. Schoeppe S, Duncan MJ, Badland HM, Alley S, Williams S, Rebar AL, et al. Socio-demographic factors and neighbourhood social cohesion influence adults' willingness to grant children greater independent mobility: A cross-sectional study. BMC Public Health 2015 Jul 22;15:690 [FREE Full text] [doi: 10.1186/s12889-015-2053-2] [Medline: 26198083]
13. Brussoni M, Olsen LL, Creighton G, Oliffe JL. Heterosexual gender relations in and around childhood risk and safety. Qual Health Res 2013 Oct;23(10):1388-1398. [doi: 10.1177/1049732313505916] [Medline: 24043348]
14. Brussoni M, Ishikawa T, Han C, Pike I, Bundy A, Faulkner G, et al. Go Play Outside! Effects of a risk-reframing tool on mothers' tolerance for, and parenting practices associated with, children's risky play: study protocol for a randomized controlled trial. Trials 2018 Mar 07;19(1):173 [FREE Full text] [doi: 10.1186/s13063-018-2552-4] [Medline: 29514699]
15. Eysenbach G. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res 2011;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]
16. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381 [FREE Full text] [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
17. sealed envelope. London, UK: Sealed Envelope Ltd URL: https://www.sealedenvelope.com/ [accessed 2021-04-12]
18. Brussoni M, Han C, Jacob J. OutsidePlay.ca: Take a risk, go play outside!. 2017. URL: https://outsideplay.ca/ [accessed 2021-04-09]
19. Bundy AC, Naughton G, Tranter P, Wyver S, Baur L, Schiller W, et al. The Sydney playground project: popping the bubblewrap--unleashing the power of play: a cluster randomized controlled trial of a primary school playground-based

intervention aiming to increase children's physical activity and social skills. BMC Public Health 2011 Sep 01;11(1):680 [FREE Full text] [doi: 10.1186/1471-2458-11-680] [Medline: 21884603]
20. Bandura A. Social cognitive theory: an agentic perspective. Annu Rev Psychol 2001;52:1-26. [doi: 10.1146/annurev.psych.52.1.1] [Medline: 11148297]
21. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95. [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
22. ParticipACTION. The Biggest Risk is Keeping Kids Indoors. The 2015 ParticipACTION Report Card on Physical Activity for Children and Youth. 2015. URL: https://participaction.cdn.prismic.io/participaction%2F61cf55e8-c1c0-42c7-ba6b-1480fd2c29b9_participaction-2015-report-card-full.pdf [accessed 2021-04-08]
23. Sandseter EBH. Characteristics of risky play. Journal of Adventure Education & Outdoor Learning 2009 Jun 24;9(1):3-21. [doi: 10.1080/14729670802702762]
24. Hill A, Bundy AC. Reliability and validity of a new instrument to measure tolerance of everyday risk for children. Child Care Health Dev 2014 Jan;40(1):68-76. [doi: 10.1111/j.1365-2214.2012.01414.x] [Medline: 22846064]
25. Chalmers RP. : A Multidimensional Item Response Theory Package for the Environment. J. Stat. Soft 2012;48(6):1-29. [doi: 10.18637/jss.v048.i06]
26. StataCorp. Stata Statistical Software: Release 15 Notes. College Station, TX: StataCorp LLC; 2017.
27. Whittle R, Mansell G, Jellema P, van der Windt D. Applying causal mediation methods to clinical trial data: What can we learn about why our interventions (don't) work? Eur J Pain 2017 Apr;21(4):614-622 [FREE Full text] [doi: 10.1002/ejp.964] [Medline: 27739626]
28. Bartholomew LK, Markham C, Ruiter RAC, Fernandez ME, Kok G, Parcel GS. Planning Health Promotion Programs: An Intervention Mapping Approach (4th ed.). San Francisco, CA: John Wiley & Sons; 2016.
29. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M, et al. Developing and evaluating complex interventions: the new Medical Research Council guidance. BMJ 2008;337:a1655 [FREE Full text] [Medline: 18824488]
30. Gottfredson DC, Cook TD, Gardner FEM, Gorman-Smith D, Howe GW, Sandler IN, et al. Standards of Evidence for Efficacy, Effectiveness, and Scale-up Research in Prevention Science: Next Generation. Prev Sci 2015 Oct;16(7):893-926 [FREE Full text] [doi: 10.1007/s11121-015-0555-x] [Medline: 25846268]

Abbreviations

BCT: behavior change technique
CFI: comparative fit index
OR: odds ratios
RCT: randomized controlled trial
RMSEA: root mean square error of approximation
SCT: social cognitive theory
SRMSR: standardized root mean square residual
TLI: Tucker–Lewis index
TRiPS: Tolerance for Risk in Play Scale

Edited by R Kukafka; submitted 08.10.20; peer-reviewed by S Wyver, C Markham; comments to author 26.10.20; revised version received 29.10.20; accepted 04.03.21; published 27.04.21.

Please cite as:
Brussoni M, Han CS, Lin Y, Jacob J, Pike I, Bundy A, Faulkner G, Gardy J, Fisher B, Mâsse L
A Web-Based and In-Person Risk Reframing Intervention to Influence Mothers' Tolerance for, and Parenting Practices Associated With, Children's Outdoor Risky Play: Randomized Controlled Trial
J Med Internet Res 2021;23(4):e24861
URL: https://www.jmir.org/2021/4/e24861
doi:10.2196/24861
PMID:33904820

Original Paper

Experiences and Factors Affecting Usage of an eHealth Tool for Self-Management Among People With Chronic Obstructive Pulmonary Disease: Qualitative Study

Sarah Marklund[1], MSc; Malin Tistad[2,3], PhD; Sara Lundell[1], PhD; Lina Östrand[3], MSc; Ann Sörlin[1], PhD; Carina Boström[3,4], PhD; Karin Wadell[1], PhD; Andre Nyberg[1], PhD

[1]Section of Physiotherapy, Department of Community Medicine and Rehabilitation, Umeå University, Umeå, Sweden

[2]School of Education, Health and Social Studies, Dalarna University, Falun, Sweden

[3]Department of Neurobiology, Care Sciences and Society, Karolinska Institutet, Stockholm, Sweden

[4]Karolinska University Hospital, Solna, Sweden

Corresponding Author:
Sarah Marklund, MSc
Section of Physiotherapy
Department of Community Medicine and Rehabilitation
Umeå University
Department of Physiotherapy
Biologihuset
Umeå, 90187
Sweden
Phone: 46 0907869835
Email: sarah.marklund@umu.se

Abstract

Background: Self-management strategies are regarded as highly prioritized in chronic obstructive pulmonary disease (COPD) treatment guidelines. However, individual and structural barriers lead to a staggering amount of people with COPD that are not offered support for such strategies, and new approaches are urgently needed to circumvent these barriers. A promising way of delivering health services such as support for self-management strategies is the use of eHealth tools. However, there is a lack of knowledge about the usage of, and factors affecting the use of, eHealth tools over time in people with COPD.

Objective: This study aimed, among people with COPD, to explore and describe the experiences of an eHealth tool over time and factors that might affect usage.

Methods: The eHealth tool included information on evidence-based self-management treatment for people with COPD, including texts, pictures, videos as well as interactive components such as a step registration function with automatized feedback. In addition to the latter, automated notifications of new content and pedometers were used as triggers to increase usage. After having access to the tool for 3 months, 16 individuals (12 women) with COPD were individually interviewed. At 12 months' access to the tool, 7 (5 women) of the previous 16 individuals accepted a second individual interview. Data were analyzed using qualitative content analysis. User frequency was considered in the analysis, and participants were divided into users and nonusers/seldom users depending on the number of logins and minutes of usage per month.

Results: Three main categories, namely, ambiguous impact, basic conditions for usage, and approaching capability emerged from the analysis, which, together with their subcategories, reflect the participants' experiences of using the eHealth tool. Nonusers/seldom users (median 1.5 logins and 1.78 minutes spent on the site per month) reported low motivation, a higher need for technical support, a negative view about the disease and self-management, and had problematic health literacy as measured by the Communicative and Critical Health Literacy Scale (median [range] 154 [5-2102]). Users (median 10 logins and 43 minutes per month) felt comfortable with information technology (IT) tools, had a positive view on triggers, and had sufficient health literacy (median [range] 5 [5-1400]). Benefits including behavior changes were mainly expressed after 12 months had passed and mainly among users.

Conclusions: Findings of this study indicate that the level of motivation, comfortability with IT tools, and the level of health literacy seem to affect usage of an eHealth tool over time. Besides, regarding behavioral changes, gaining benefits from the eHealth tool seems reserved for the users and specifically after 12 months, thus suggesting that eHealth tools can be suitable

media for supporting COPD-specific self-management skills, although not for everyone or at all times. These novel findings are of importance when designing new eHealth tools as well as when deciding on whether or not an eHealth tool might be appropriate to use if the goal is to support self-management among people with COPD.

Trial Registration: ClinicalTrials.gov NCT02696187; https://clinicaltrials.gov/ct2/show/NCT02696187

International Registered Report Identifier (IRRID): RR2-10.1136/bmjopen-2017-016851

(J Med Internet Res 2021;23(4):e25672) doi:10.2196/25672

KEYWORDS

COPD; qualitative content analysis; eHealth; self-management; primary care; chronic disease

Introduction

Chronic obstructive pulmonary disease (COPD) is a common, preventable, and treatable disease, listed as the third leading cause of death worldwide by the World Health Organization [1]. The disease most often presents with dyspnea of varying severity [2]. COPD is a complex disease with several pulmonary and extra-pulmonary manifestations, and leads to numerous negative clinical consequences such as exercise intolerance, decreased physical activity, decreased quality of life, as well as increased health care use [2,3].

Previous studies have shown that among people with COPD, those who use self-management strategies to manage their disease present with fewer symptoms and show reduced negative consequences of the disease [4-6]. Self-management strategies, such as exercise training, breathing strategies, and energy-conserving techniques during activities of daily life [7], have previously been found to reduce the need for hospitalization, increase physical activity and physical performance, as well as improve the quality of life [4-6,8,9].

Despite a high prioritization of self-management in COPD treatment guidelines [10], a staggering amount of people are not offered support for such strategies due to limitations of both individual and structural nature [10-12]. It is, therefore, crucial to find a way to circumvent this problem so that people with COPD are offered support to learn and use these COPD-specific self-management strategies.

eHealth tools, such as mobile apps and web-based platforms, represent a promising way of delivering health services such as support for self-management strategies to people with COPD. eHealth solutions are becoming increasingly common and are found to be feasible and effective within the COPD population [9,13-20]. eHealth tools have, for example, been used for educating and keeping track of a person's health and are thought to be a significant source of health-related information [14,15,19,20]. We previously found that access to the COPD Web for 3 months resulted in increased self-reported levels of physical activity among people with COPD [17]. Improved COPD-specific knowledge and altered self-management strategies were also found among the participants. However, the general use of the COPD Web varied profoundly among participants, and the vast majority of users mainly used the platform during the initial month [17]. To date, several research groups have investigated and provided valuable information on important factors to consider when designing eHealth tools [21-23], though less is known about user behavior over time. Specifically, knowledge of factors associated with the use of eHealth tools over time is sparse. The aim of this study was, therefore, to explore and describe the experiences of using an eHealth tool over time and factors that might affect usage among people with COPD.

Methods

Study Design

This exploratory qualitative study is presented in line with the Consolidated Criteria for Reporting Qualitative Research (COREQ) guidelines [24]. This qualitative study is part of a process evaluation in a parallel-group (1:1 allocation) controlled pragmatic pilot trial that was reported in line with the Consolidated Standards of Reporting Trials (CONSORT) statements for pragmatic trials and for pilot and feasibility trials [25,26]. The study was registered at ClinicalTrials.gov (NCT02696187) and ethical approval was given by the Regional Ethical Board, Umeå University, Umeå, Sweden (Dnr: 2014-319-31, 2015-457-32). Written informed consents were obtained from each participant before enrollment in the study. This study includes participants that were allocated to the intervention group and that had access to an eHealth tool, the COPD Web. To further ensure privacy of participants, all names were changed to pseudonyms during the start of the analysis, so that only interviewers (AN and MT) and SM knew their real names.

Setting and Sample

Five publicly funded primary health care units (2 situated in the north of Sweden and 3 in the middle of Sweden) were invited to participate as study sites in the pragmatic pilot trial [27]. In the main study 83 patients with COPD were included; of these, 43 were randomized to the intervention group and had access to the COPD Web. Of those randomized to the intervention group, a minimum of 3 participants at each of the 5 units were contacted and asked to participate in the individual interviews. We used purposeful sampling to ensure there is at least one male/female person with COPD at each primary health care unit. To be included in this trial, participants had to be able to read and understand Swedish or be assisted by someone with this capacity when using the eHealth tool.

The COPD Web

In brief, the COPD Web is an interactive webpage cocreated with users [16]. It consists of 2 main sections: one directed at health professionals, and another directed at people with COPD. Contents include videos, written information, images, and

helpful links. The COPD Web section aimed at people with COPD to support their self-management by increasing their knowledge about COPD and strategies to improve their health (eg, activities such as physical activity and exercise, breathing techniques, observing symptoms of exacerbations, and advice about making everyday activities less strenuous) [27]. In addition to the specific content of the COPD Web, there were a few things on the fringe of what was covered by the COPD Web that should be noted. The COPD Web also includes a function of registering physical activity (steps), for which participants received a pedometer with instructions on how to use it as well as an information leaflet on the importance of physical activity [27]. The COPD Web also had automated notifications of new publications on the website via email. It was first introduced to each participant by a health professional during a regular visit using a standardized procedure (designed to take a maximum of 10-15 minutes). During the introduction of the COPD Web, all participants were provided with a username and login. Further information about the eHealth tool is presented in the protocol [27].

Process of Data Generation

Overall, 16 participants (12 women), age range 48-86, accepted to participate in an individual face-to-face interview at 3 months after the intervention started. Follow-up interviews were done at 3 and 12 months, as both ranges of time are commonly used when investigating intervention-based effects in people with COPD [13,14]. At 3 months, 15 interviews were conducted in the participants' homes and 1 at a university (the worksite of the interviewer), according to the participant's wishes. Sociodemographic information including age, sex, civil status, occupation, physical activity, smoking habits, and educational level was obtained through a standardized questionnaire while information on lung function was obtained from medical records [27]. At the 12-month follow-up, the participants were contacted again over the telephone and 7 (5 women) of the previous 16 participants accepted a second individual interview—this time conducted over the telephone. The reason for declining a second interview was that they had not used the COPD Web at all between the 3- and 12-month follow-up period.

Experiences related to the usage of the COPD Web over time were collected through individual semistructured qualitative interviews. An interview guide was used as the framework for the interviews. It consisted of the following main areas: (1) user habits, (2) user experience, (3) accompanying parts (ie, introduction, pedometer, electronic newsletter), (4) potential effects, and (5) future use (Multimedia Appendix 1). Questions regarding all main areas were posed, albeit in varying order. The interviews at the 3-month follow-up ranged between 7 and 40 minutes (median 25 minutes) and interviews at 12 months ranged between 4 and 14 minutes (median 8 minutes).

In addition, data on characteristics of the individuals and thought to be important for their use of the COPD Web were collected. It included the impact of COPD on daily life measured with the COPD Assessment Test [28], dyspnea measured with the modified Medical Research Council Scale [29], and health literacy (ie, the degree to which individuals can find, understand, and use information and services to inform health-related decisions and actions for themselves and others) [30]. The latter was measured with the Communicative and Critical Health Literacy (CCHL) Scale questionnaire. A total score of less than 100 indicates a sufficient communicative and critical health literacy [30], a total score of more than 100 but less than 1000 indicates a problematic health literacy, and a total score of over 1000 indicates a lack in communicative and critical health literacy [30]. Data on the participants' use of the COPD Web were collected automatically from the website and included the number of visits (logins), pages used, and time spent on the website [27].

Research Team and Reflexivity

AN (PhD in physiotherapy, male, 32 years) and MT (PhD in physiotherapy, female, 43 years), conducted the interviews separately depending on geographic placement. Both interviewers were employed as postdoctoral researchers at Umeå University at the time. Prior to the study, AN had performed over 20 interviews (no specific training), and MT had performed over 30 interviews (supervision during postdoctoral employment). There was no relationship established between the researcher and the participant before the interviews.

In all of the interviews, only the researcher and the participant were present, and audio recordings were used. MT made short field notes for most of the interviews (used as a mean of recollection during analysis), AN did not take notes during or after the interviews. No immediate callbacks on the interviews were made (ie, for potential amendments or additional questions). Interesting or unexpected topics raised during the interviews were discussed and used to guide follow-up questions during the following interviews. Interview audio recordings were transcribed verbatim by an experienced third-party transcriber and verified by the authors by comparing with the audio records [31]. Transcripts were not returned to participants for comment or correction. Participants were not engaged to provide any feedback on the findings.

Analysis

Data were analyzed using an inductive approach of qualitative content analysis, as this method is useful when dealing with already gathered qualitative data where the goal is to increase the understanding of experiences of using an eHealth tool [32]. All interviews were read through several times (with the assistance of audio recordings for auditory cues). The interviews were then condensed, coded, and sorted into categories and subcategories [33]. MAXQDA 2018 software was used in the analysis process to facilitate administration of the interviews, codes, and quotes. SM was main responsible for the analysis, and researcher triangulation was used throughout the whole analysis phase to attain a higher level of credibility [16,33]. Continuous discussions among authors during the interviewing phase and making use of a semistructured interview guide were measures taken to enhance dependability [33].

To further explore usage over time, the participants were subgrouped based on their objective use of the COPD Web during the initial 3 months. *Users* were defined as those with more than 1 login/month or more than 20 minutes total spent time/month at 3 months. An individual that did not meet these

criteria were defined as a *nonuser/seldom user* at 3 months. Codes by users and nonusers/seldom users were then marked in the subcategories to analyze the data further. Furthermore, to explore usage during the initial 3-month period, we also compared responses between the 3- and 12-month interview in those accepting a follow-up interview at 12 months.

Results

Study Participants

Patients with COPD at each of the 5 included primary health care units were contacted with the goal of recruiting at least one female/male person with COPD at each unit. No male participant accepted an interview at one of the primary health care units. In total, we contacted 23 patients with COPD (12 female); of these, 7 declined (2 female), and 16 accepted participation in the interviews at 3 months. Furthermore, of those who accepted an interview at 12 months, 5 out of 7 were users at 3 months. When subgrouped based on their use of the COPD Web during the initial 3 months, users had, in median, sufficient communicative and critical health literacy. By contrast, nonusers/seldom users had problematic communicative and critical health literacy. No other apparent differences in absolute values were seen for any other participant characteristics between users and nonusers/seldom users at 3 months (Table 1).

Table 1. Participant characteristics.

Characteristics	Interview at 3 months (n=16)	User at 3 months (n=6)	Nonuser/seldom user at 3 months (n=10)
Age (years), median (range)	71 (48-86)	70 (48-86)	72 (54-81)
Sex (female), n	12	5	7
FVC[a] predicted (%), median (range)	80 (59-131)	83 (71-131)	71 (59-126)
FEV_1[b] predicted (%), median (range)	61 (30-99)	60 (54-99)	61 (30-93)
FEV_1/FVC (%), median (range)	52 (28-63)	56 (47-63)	46 (28-62)
CAT[c], median (range)	13 (2-20)	15 (2-17)	12 (2-20)
MRC[d], median (range)	1 (0-4)	1 (1-3)	1.5 (0-4)
Stage of COPD[e], n (%)			
A	3 (19)	1 (17)	3 (30)
B	7 (44)	4 (67)	3 (30)
C	2 (13)	1 (17)	1 (10)
D	3 (19)	0 (0)	3 (30)
Smoking status[f]			
Never smoker, n (%)	0 (0)	0 (0)	0 (0)
Ex-smoker, n (%)	13 (87)	6 (100)	7 (78)
Current smoker, n (%)	2 (13)	0 (0)	2 (22)
Pack-years, median (range)	24 (6-55)	18 (6-20)	29 (15-55)
Employment status[f], n (%)			
Currently working	3 (20)	2 (33)	1 (11)
Retired	12 (80)	4 (67)	8 (89)
Sickness benefits	0 (0)	0 (0)	0 (0)
Living with[f]			
Alone, n (%)	11 (73)	5 (83)	6 (67)
Family, n (%)	4 (27)	1 (17)	3 (33)
Education level[f], n (%)			
Primary	9 (60)	4 (67)	5 (56)
Secondary	3 (20)	1 (16.5)	2 (22)
Tertiary	3 (20)	1 (16.5)	2 (22)
Health literacy test			
CCHL[g] scale (score), median (range)	104 (5-2102)	5 (5-1400)	154 (5-2102)
Usage			
Minutes/month, median (range)	7.8 (0-144.6)	42.7 (26.8-144.6)	1.8 (0-16.4)
Logins, median (range)	4 (0-33)	10 (5-33)	1.5 (0-14)

[a]FVC: forced vital capacity.

[b]FEV_1: forced expiratory volume in 1 second.

[c]CAT: COPD assessment test (0-40, higher scores denote a more severe impact of COPD on patient's daily life).

[d]MRC: Medical Research Council Scale (1-5, higher score denote a higher degree of disability that breathlessness poses on day-to-day activities).

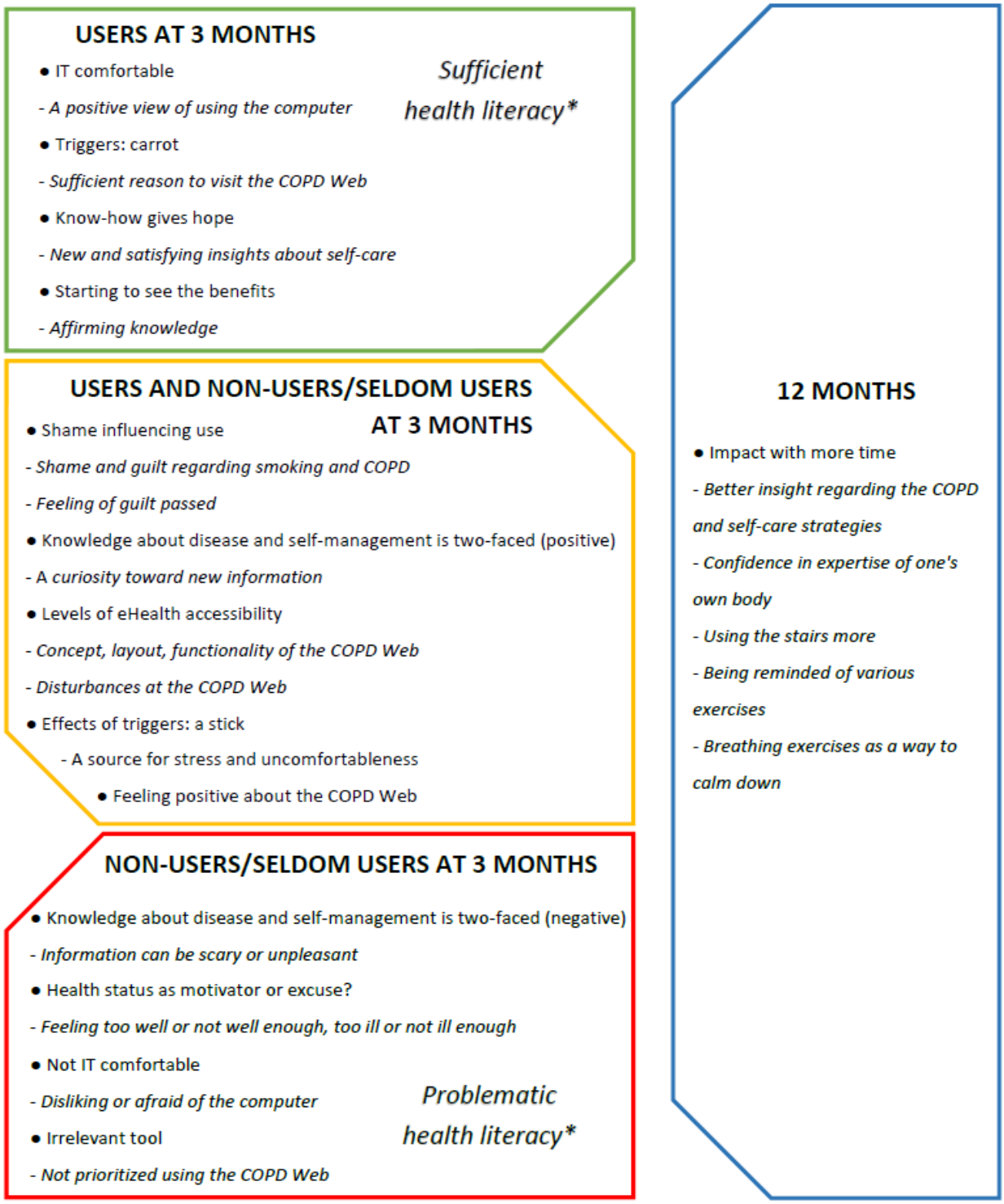
[e]COPD: chronic obstructive pulmonary disease.

[f]Data missing on 1 participant.

[g]CCHL: Communicative and Critical Health Literacy Scale, Swedish version (lower score = better health literacy).

The analysis resulted in 3 main categories, which, together with the subcategories, represent the participants' overall experiences regarding the use of the COPD Web over time and descriptions of factors that could affect the usage of the said tool (Textbox 1). An overview of the main results are seen in Figure 1.

Figure 1. Overview of user experience and factors affecting usage. "Users at three months" represent experiences mainly seen among those who used the COPD Web during the initial three months, "Non-user/seldom-users at three months" represent experiences mainly seen among those who did not use, or used at a limited extent, the COPD Web at three months, "Users and non-user/seldom-users at three months" represents experiences seen among both users and non-user/seldom-users at three months while "12 months" represents experiences mainly expressed at 12 months. *Based on the communicative and critical health literacy scale, Swedish version.



Textbox 1. Description of categories and subcategories. IT: information technology.

Ambiguous impact

- Shame influencing use
- Knowledge about disease and self-management is two-faced
- Health status as motivator or excuse?
- IT comfortable or not?

Basic conditions for usage

- Levels of eHealth accessibility
- Triggers: carrot or stick?
- Feeling positive about the COPD Web
- Irrelevant tool

Approaching capability

- Know-how gives hope
- Starting to see the benefits
- Impact with more time

Ambiguous Impact

Overview and Subcategories

In this category, participants expressed contradicting thoughts about seeking care due to, and specific information about, COPD. There were also some diverging thoughts on competence and motivation regarding the use of information technology (IT). The category includes the 4 subcategories "Shame influencing use," "Knowledge about disease and self-management is two-faced," "Health status as motivator or excuse?," and "IT comfortable or not."

Shame Influencing Use

Having COPD was sometimes experienced as something taboo, hushed up, or embarrassing which led to loneliness or isolation and prompted nonusers/seldom users to distance themselves from the eHealth tool, and users to try out the tool. A sense of not having anyone to talk to about their COPD, and feeling like COPD is rarely talked about, were mentioned as reasons for liking and welcoming the COPD Web.

> *Yes, but just talking about COPD, well...you just actually don't do that. That's why it's a good thing with this kind of page where you can get to know a little more.* [Charlotte, user, 3-month interview]

Among nonusers/seldom users experiences of shame and guilt regarding smoking and COPD were conveyed. A feeling of shame caused participants to withhold the diagnosis from family and friends and also hindered contacting health care. Having developed COPD was described as one's own fault (due to smoking), and therefore, something one should handle all by oneself. Hence using the COPD Web might be seen as getting undeserved assistance.

> *Well, this...I only have myself to blame. This is a result of my smoking. And that's how it is. I guess I'll have to battle this on my own.* [Gertrud, nonuser/seldom user, 3-month interview]

Knowledge About Disease and Self-Management Is Two Faced

Participants proclaimed a curiosity toward new information, a wish to know more, and indeed reported having actively searched for information and sources of information. Only nonusers/seldom users expressed that the urge to receive information about COPD and COPD-related subjects was not absolute. In addition, users mentioned the advantage of having access to a lot of information, whereas nonusers/seldom users felt that information could be scary or unpleasant, especially if it contradicted one's preconceptions. At the 12-month interview, participants (both users and nonusers/seldom users at 3 months) mentioned a desire to see if any new information had been posted or to refresh knowledge on something they had read about earlier as reasons for revisiting the COPD Web.

Health Status as Motivator or Excuse?

The own health status experienced was a prominent reason to visit or not to visit the COPD Web (ie, feeling good or poor made them more likely to use [or not use] the COPD Web). One view was that if one felt too well he/she would not be motivated to search for information, while another view was that one had to feel good to really be able to take in the advice and information. Besides, some thought themselves not ill enough while some proclaimed themselves not ill at all—either way, none of these conditions seemed to warrant visits to the COPD Web.

> *Well, of course, the better you feel…I believe it's more like you think it's more fun to just go in and read.* [Bella, nonuser/seldom user, 3-month interview]
>
> *[…] I kind of feel too healthy.* [Katarina, nonuser/seldom user, 3-month interview]

At 12 months it was still thought by former users to be more likely that one will use the eHealth tool if he/she feels sick.

IT Comfortable or Not?

Participants had different comfort levels regarding the use of computers and the internet. A positive view of using the computer and the internet was expressed among users. Users further mentioned the advantage of the COPD Web being accessible from anywhere (at home and any other place on the globe). Disliking or being unused to the computer or the internet was only mentioned by nonusers/seldom users as hinders for visiting the COPD Web and searching for information in general. Later, at 12 months, a nonuser/seldom user that previously stated her dislike and ignorance regarding both computers and the internet said that she was somewhat afraid of the computer:

> *Actually, I'm also a little afraid of the computer […] l am reluctant to have anything to do with it. Unfortunately. Sometimes I think that now I will…and then… "No, not now, not now, not now!"* [Penny, nonuser/seldom user at 3 months, 12-month interview]

Nonusers/seldom users also reported not prioritizing using the computer or the COPD Web, not having any interest in computers or the COPD Web, and preferring to read information as well as note (eg, step counts) on a paper instead of on the computer screen. Others explained their absence from the website as a result of being engaged in other activities or having forgotten about the computer or the COPD Web. Still, comments on lack of time for using the computer or COPD Web occurred among nonusers/seldom users. Disinterest for the eHealth tool was also reported at 12 months by both users and nonusers/seldom users.

Nonusers/seldom users expressed a need for their support system in the form of health professionals or close relatives. Support was used as a source of knowledge when it came to using the technology involved, as well as a motivational source for looking up information or learning new things and for allocating time for computer use.

> *I'm lucky to have a couple of lads I can call. "What the hell do I do now?"* [Kenneth, nonuser/seldom user, 3-month interview]

Basic Conditions for Usage

Overview and Subcategories

This category represents the technical aspects of the eHealth tool, covering both experiences of the functionality of the software and hardware and various attitudes using technical products in general. This category includes the 4 subcategories "Levels of eHealth accessibility," "Effects of triggers: a carrot or a stick?," "Feeling positive about the COPD Web," and "Irrelevant tool."

Levels of eHealth Accessibility

Generally, the concept, layout, and functionality of the COPD Web was appreciated at 3 as well as 12 months. Participants described disturbances, such as problems with their computer, frustrating and inconvenient. Sometimes strong feelings of being fed up with things not working were expressed, at both 3 and 12 months.

> *No, there's nothing else. There's my computer. I would be out on the web almost every other day looking, otherwise.* [Carol, nonuser/seldom user, 3-month interview]

> *Well, I've had a few problems with my computer, but I have been in and checked all [e-mail newsletters] that I've received.* [Cindy, user at 3 months, 12-month interview]

The COPD Web's low compatibility with smartphones was reported as a hindrance by some nonusers/seldom users at both 3 and 12 months.

Triggers: Carrot or Stick?

Overall, participants felt that the triggers (ie, automated notifications of new content, pedometer, and the step registration function) had a kind of gateway function and made using the COPD Web closer at hand. Users described themselves positively about using and subsequently revisiting the COPD Web when getting email newsletters from the eHealth tool.

The pedometer was reported to be a useful tool concerning physical activity; it was also perceived as a sufficient reason to visit the COPD Web to register steps on the site. Using the pedometer was recognized as a way to prove, in a concrete way, that you had done right (or done something at all) with regard to your level of physical activity. There was a drive, a motivation, expressed in registering steps on the COPD Web, and a pedometer helped in building this motivation.

> *My [pedometer] was my dangling carrot. All the time, I could see how much I walked … well, I thought that was really terrific.* [Patricia, user at 3 months, 12-month interview]

However, the pedometer was also recognized as a possible source for stress and uncomfortableness if you had not "done enough." Some nonusers/seldom users indeed felt like being monitored or watched when wearing it, which could be experienced as both a motivational push and something negative. Further, the pedometer was thought by some users to be of more use to someone worse off, health-wise, than oneself.

Overall, participants expressed that the introduction to the COPD Web had included a suitable type and amount of information and also that the health professionals had introduced the COPD Web well, both elements triggering interest for the COPD Web. Among nonusers/seldom users, getting valuable support during the introduction which inclined them to participate despite feeling insecure about computer use was expressed. Further, some users reported having participated due to a sense of duty. However, a somewhat contradictory experience was also conveyed, expressing that the introduction took too much time and was not relevant:

> *Because I'm not ill. So I shouldn't have to be there (at the introduction). And so I thought … good God, does this [introduction] never end?* [Gertrud, nonuser/seldom user, 3-month interview]

Feeling Positive About the COPD Web

Users of the COPD Web expressed that they were generally missing feedback from the health care provider/system regarding the actions of self-care they had already set in motion and were glad to be able to verify these with the help of the COPD Web. Possible prevention of deterioration was expressed as reason enough for using the COPD Web for some users. Users also mentioned that one can choose what to focus on, or do, on the COPD Web, and this was remarked as a very positive feature/function of the site. Further, users of the website reported liking how it can remind you about information, exercises, and other supportive tools for self-management. The COPD Web was also regarded as a source for positive extrinsic motivation by both users and nonusers/seldom users. Participants expressed curiosity for the COPD Web, as well as positive anticipatory feelings toward it. The COPD Web was described as a new concept that had not been introduced to them before and which spiked their interest.

> *[…] I thought … my God, that's great! This is about me. Naturally I needed to grip the chance. That's how it was. And the way in which [the COPD nurse] presented things and the way [the COPD nurse] described things caught my interest […]* [Kelly, user, 3-month interview]

While the thought of appreciating all help that you can get was mentioned, it did not ensure the usage of this eHealth tool. Participants mentioned that looking around in general and checking what the site was about as reasons for visiting the COPD Web.

Irrelevant Tool

Nonusers/seldom users felt that they did not have time for, had forgotten about, or had otherwise not prioritized using the COPD Web. Some users also expressed this feeling. A sense of having been "in the game" for too long, and thus already having heard all the information was expressed. This was also mentioned at 12 months. Some felt this was fine, and some were frustrated about not being able to find new information, but this ultimately seemed like a reason not to "bother" with the eHealth tool. Others felt that they were already doing what the COPD Web instructed them to do regarding self-care and so had less need for the website. Consequently, the website was expected to feel more relevant when you are newly diagnosed. In addition, visits on the website could be due to a sense of obligation (toward the study) as this former user stated at 12 months when asked how come they stopped their, formerly very active, step registration:

> *Yes, but then I asked you whether […]… should I continue registering my steps? You said, no, you do as you want to. […] So then I didn't bother with it.* [Cindy, user at 3 months, 12-month interview]

Approaching Capability

Category Overview

This category includes participants' experiences regarding getting access to COPD-specific knowledge and skills for self-management and the benefits experienced from using the COPD Web, as well as reports of actual behavioral changes within 12 months since starting to use the COPD Web. The 3 subcategories covered are "Know-how gives hope," "Starting to see the benefits," and "Impact with more time," wherein the for the last subcategory all codes are solely from the later (12-month) follow-up.

Know-How Gives Hope

The content of the COPD Web was perceived to be instructive in a concrete way, which gave new and satisfying knowledge about and insights into self-care. Users reported using the website as a sort of archive where they could search for information and tips, even regarding things that were not specific to COPD. Some users felt it was nice just to have access to this well of information, seemingly regardless of how much it was utilized. A remark on the power of habit was made where the use of the COPD Web felt helpful in changing or affecting habits in a good way because it could be revisited several times. Users also perceived that their view of COPD had become more hopeful because the COPD Web had informed them about the things that they were able to use to their advantage. Having heard depressing statements from people in one's vicinity, it felt positive to learn about actions to take and feelings of being fortified and boosted in your actions of self-care.

> *Well, do you want to know what is good? That there is … that you don't have to become worse and worse but there is something you can do about it yourself. I think that's positive* [Cindy, user, 3-month interview]

By contrast, some users conveyed a sense of no change at all in their view of the disease from having access to the COPD Web.

Starting to See the Benefits

To get an affirmation of what you had already heard or learned from other sources was an appreciated function of the COPD Web, and this is in line with the statement of some users that not all advice on the website is new news, but they can still help you. Otherwise, users felt they visited the website to improve their overall situation by learning more about the diagnosis, their body and treatments, and how to perform a specific exercise. Breathing techniques on the site were reported as helpful during, for example, travels and also as a helpful way to prepare before a particular trip.

> *I looked at this thing about going up steps. And so I went to Spain and I knew that we had to go up some steps a long way up to a church and so I actually practiced exhaling on the next step. You breathe in, take a step and then breathe out and take two steps and so on. I think this worked very well.* [Patricia, user, 3-month interview]

In addition, the COPD Web function of step registration was mentioned as a strong motivator for doing exercises.

Impact With More Time

In this subcategory, a little more time had passed, and all statements are from the 12-month follow-up. The content of the website was said to have led to a better insight regarding the COPD and the self-care strategies, as well as to essential changes in everyday life. Participants commented that the COPD Web

had helped them in understanding how vital actions such as contacting the health care in time and using the stairs were important. However, learning about COPD, in general, had also given some confidence in the expertise one holds about one's own body. Therefore, it was said, one could also feel more confident in asking for, or even demanding, care from their health care settings when something did not seem right.

> *Now, just like then when I was so wilted at Christmas time there, I was the one who stood up for myself and told them that this is the way I am and I now need to come in and perhaps get some help to stop all this. Yes. Not to say… and other times I've let things drag on more. That's just the way it is. But I do understand that it's more important to [seek care] quicker and not let things drag on.* [Kelly, user at 3 months, 12-month interview]

Others said that changes to everyday life were the actively changed patterns of body motions (eg, standing up without using one's hands), using the stairs more, and avoiding getting minor illnesses that might escalate to worse conditions due to COPD. Some felt that they had learned more from the website than from traditional health care meetings. The convenience of being reminded of various exercises and breathing exercises was especially mentioned as a way to calm down, to feel more at ease. By contrast, there were participants, including both former users and nonusers/seldom users, who after the 12 months still felt that their habits had not changed and that things were as they always had been—with or without the COPD Web.

Discussion

Principal Findings

This study aimed to, among people with COPD, explore and describe the experiences of using an eHealth tool over time and factors that might affect usage. The main results reflected study participants' experiences in the categories of *ambiguous impact*, *basic conditions for usage*, and *approaching capability*, and indicated that level of motivation, comfortability with IT tools, and the level of health literacy seem to affect usage of an eHealth tool over time. Furthermore, regarding behavioral changes**,** gaining benefits from the eHealth tool seems reserved for the users and specifically after 12 months, thus suggesting that eHealth tools can be suitable media for supporting COPD-specific self-management skills, although not for everyone or at all times. These findings support earlier qualitative studies noting that people with COPD are a heterogeneous group and that there is no one general way to reach all people with COPD [34]. Furthermore, even though technical difficulties are irritating, they do not seem to determine level of usage. Similarly, recognizing the usefulness of the eHealth tool, and even curiosity toward the tool do not seem to influence the level of usage.

Interpretation of Findings: User Experience and Factors Affecting Usage

In general, users shared positive comments about IT tools on the COPD Web, including receiving information, the electronic newsletters, and the contents on the eHealth tool. These findings are in line with previous studies in which those who used a mobile app, in general, had a positive view on eHealth tools [35,36]. Furthermore, as captured in the subcategory "Starting to see the benefits," users seemed to be able to incorporate the advice into their everyday life, and they felt that the eHealth tool helped them to improve their overall situation. This suggests that the eHealth tool provided participants with at least some of the tools needed to manage their condition [7].

As knowledge and skill on how to integrate the demands of the disease into the daily routine are crucial to enable behavior modification [7,37], findings from this study are of importance. Specifically, in a similar way, as previously reported after self-management interventions for people with COPD, users in our trial reported increased knowledge on how to perform specific exercises and use of breathing techniques in daily life [38]. Some users also mentioned that they gained a sense of hope, as the eHealth tool had informed them about their disease so that they were able to utilize this knowledge to their advantage. The latter is of importance as a person's belief in his or her ability to manage the disease is a powerful and well-recognized predictor of health-related behavior changes [7,39]. Our findings also suggest that the eHealth tool was beneficial as a self-management tool for the users, as it seemed to have resulted in more substantial confidence in their ability to manage their disease, which is an essential factor to influence specific health behaviors [40-43]. The sense of hope in itself could also be a reason for the continued use of the eHealth tool among users because the perception of a clear benefit has been highlighted as a critical facilitator for eHealth usage.

In addition to the positive view on information and IT tools, when exploring potential differences in baseline characteristics among users and nonusers/seldom users, we saw that users also had sufficient communicative and critical health literacy as measured by the CCHL Scale, Swedish version. Health literacy has been defined as "the degree to which individuals can obtain, process, and understand basic health information and services needed to make appropriate health decisions." [44]. It is, in other words, a set of fundamental skills for obtaining precisely the kind of information that is communicated with the help of eHealth tools, which might help explain some basic differences between those who used and did not use the COPD Web. By contrast, nonusers/seldom users of the eHealth tool had a problematic communicative and critical health literacy. Thus, the level of health literacy might be a contributing factor to our observed findings because this is important in taking in new information as well as highlighting a need for education and training. The latter has also been identified as a critical barrier to eHealth use in daily practice [30]. Limited health literacy toward eHealth was likewise the top mentioned barrier for implementing eHealth services in a recent systematic review and meta-analysis [36]. Low health literacy is also widespread among patients with COPD. For example, Puente-Maestu et al [45] found that more than 50% of their patients with COPD had low health literacy. Similarly, as evident among nonusers/seldom users in our trial, people with both COPD and low health literacy have previously also been reported to be more dependent on others and to be more concerned about their illness [45,46]. With this in mind, our findings further support previous views

that health literacy should be considered when designing self-management support programs for people with COPD [47] and, specifically, that low health literacy seems to be an important barrier to using eHealth tools over time and other strategies should be considered. Moreover, participants expressed that the COPD Web would have been especially relevant when they were newly diagnosed and that the COPD Web would have enabled them to circumvent underestimating the importance of self-care–related information. This observation is similar to the findings made by Ansari et al [48] who demonstrated that newly diagnosed patients with COPD had difficulty recognizing the impact of COPD on their health, mostly due to low awareness of the disease and its long-term implications.

Furthermore, nonusers/seldom users reported a problematic attitude toward IT tools or a completely nonfunctioning computer or internet connection at their disposal. However, even though nonusers/seldom users had access to and utilized help when needed, they still did not use the eHealth tool, which implies that their lower level of interest (low motivation) or faith in new information and technology had more to do with not using the tool more. Triggers, such as the automatic newsletters or pedometer, did not seem to affect usage in itself. The electronic newsletters, for example, did help in getting people to use the eHealth tool actively—but still, it only seemed to make a positive impact for those already more inclined to receive new information, use the computer/internet, and prioritizing an eHealth tool such as this. The same goes for the pedometer and the step registration function that were expressed as motivating and useful by users. These findings could be related to the Fogg Behavior Model (FBM) [37]. The FBM suggests that acquiring a targeted behavior, in our case, using the eHealth tool, requires sufficient motivation, ability, and triggers. Thus, the low motivation, the low ability to use IT tools, and the opposing view on used triggers help us understand why nonusers/seldom users did not use the tool.

Nevertheless, as reported above and captured in the subcategory "IT comfortable or not," the expressed need for assistance could also be linked to the low health literacy among our nonusers/seldom users [45,46,49]. Further, no apparent differences in disease severity or other sociodemographic characteristics such as age, sex, smoking history, living conditions, or basic educational level could be seen between users and nonusers/seldom users of the eHealth tool. This further indicates that our population either used their health status as an excuse for not using the eHealth tool or that their perceptions of their health status varied no matter the objective measurements. Our findings, however, still indicate that the nonusers/seldom users were in that group more due to insufficient self-competence and motivation, and a more negative general attitude toward getting informed. For example, in contrast to users that only expressed a positive attitude toward receiving new information, nonusers/seldom users also expressed that information could be scary or unpleasant. Again, these findings could be linked to the lower health literacy of nonusers/seldom users [45,46,49]. In this study, both users and nonusers/seldom users expressed that the eHealth tool was easy to use. This indicates that even though the ease of use has been stated as the most successful factor for implementing eHealth services [36], it was not a key factor explaining usage in our sample.

In the interviews performed at 12 months, we could see a shift in the type of information provided in the interviews, despite using the same interview guide as during the 3-month interviews. Specifically, as highlighted in the category "Assembling know-how," the reported changes in behavior (eg, standing up without using one's hands and using the stairs more) were mainly expressed at the 12-month interviews. Thus, knowledge from the eHealth tool was put into practice, and behavioral changes were underway at this time. Even though participants, mainly users, also expressed benefits since having access to the eHealth tool after the initial 3 months, these findings highlight that time might be an essential factor when evaluating the potential benefit of eHealth tools. These findings are in line with the results from a recent previous study suggesting that a continuous self-management program helps people with COPD to perform given self-health behaviors, thus indicating that achieving changes in health behavior takes time [38]. Further support for the latter is also seen in a qualitative systematic review from 2018 [34]. Other than this observation, no apparent differences were seen between the 3- and 12-month interviews, which also supports the consistency of our 3-month findings [31,50].

Strengths and Limitations

Strengths of this study are the direct comparison of users and nonusers/seldom users, as well as the design of the study following the COREQ guidelines, increasing the credibility of our findings [24]. Trustworthiness has been strived for by utilizing triangulation and recurrent discussions regarding the material through all phases [51]. Furthermore, interviews were conducted via both face-to-face meeting and telephone due to practical choices. The 2 interview methods are, however, equally credible according to Ward et al [52] who found that interviewees felt free to divulge sensitive information over the telephone and that the communicative cues were to be regarded as equal to those used within face-to-face interviews. Interview lengths are varying and at times, very short. However, a specific duration is not a guarantee for richness [51,53], and we interpreted our data to be rich enough for the analysis that was carried out. To counteract the potential loss of information by removing the researchers from this process by using a professional transcriber [54], the transcripts were then checked for deviations between audio and transcript [31]. Besides, knowledge and notes from the interviewers were utilized when needed during analysis. In addition, we continuously consulted the audio recordings when triangulation indicated risks of interpretational differences of a transcript [33]. Even though the number of interviews in itself is not a crucial criterion in qualitative methods, it should be noted that the number of informants, especially at 12-month interview, was relatively few. Besides, even though there was a variation in age, sex, education, stage of COPD, and living conditions, the vast majority of included participants were females, living alone, and with overall low disease severity as evident by low CAT and MRC scores, as well as moderate forced expiratory volume in 1 second (%FEV_1) of predicted values. It should also be noted

that the majority of the interviewees were women, highlighting a potential selection bias [55], as more men than women declined to participate in the interviews. Analyzing whether the results would have been similar if more men had participated was outside the scope of the study, but should be considered for future projects. In addition, of importance, the cut-offs used to define users and nonusers/seldom users were not predetermined and set by the researchers after having access to information about the usage of the COPD Web. However, these cut-offs were set by a researcher not involved in the primary analysis, and before codes were tracked back to individual informants. Lastly, objective data on the use of the eHealth tool were only available during the initial 3 months. Thus, we cannot elaborate on whether the users at 3 months also were users at 12 months.

Conclusions

To our knowledge, this study was among the first to explore experiences and factors that affect the usage of an eHealth tool aiming at improved self-management strategies in people with COPD over time. The novel findings of this study indicate that usage of an eHealth tool, to support self-management, is affected by the view on the information. That is, curiosity or fear for new information can influence whether to use an eHealth tool or not. In addition, a low motivation, a higher need for technical support, and a problematic health literacy were seen among nonusers/seldom users. By contrast, users were comfortable with IT tools, had a positive view on triggers, and had sufficient health literacy. Lastly, reaping benefits from an eHealth tool seems to be reserved for the users, and primarily after 12 months. These findings suggest that eHealth tools can be suitable media for supporting COPD-specific self-management skills, although they are not for everyone or at all times.

Furthermore, factors such as motivation, IT comfortability, and level of health literacy are important to consider when deciding on whether or not an eHealth tool might be appropriate to use if the goal is to support self-management among individuals with COPD. In addition, our findings highlight that time might be an essential factor when evaluating the potential benefit of eHealth tools aiming at behavioral changes regarding self-management strategies.

Acknowledgments

This work was supported by the Swedish Research Council (grant number K2014-99X-22572-01-4) and the Swedish Heart and Lung Foundation (grant number 20130331). The funders had no role in study design, data collection and analysis, preparation of the manuscript, and decision to publish.

Authors' Contributions

SM has made direct and substantial contribution to this work by performing the analyses, interpretation of data, and drafting of the manuscript. MT made a direct and substantial contribution to this work by playing a leading role in the design of the study, data collection, interpretation of data, and by providing critical revisions that are important for the intellectual content of the manuscript. SL, LÖ, AS, and CB made a direct and substantial contribution to this work by playing a significant role in the analysis and interpretation of data, and by providing critical revisions that are important for the intellectual content of the manuscript. KW is the principal investigator and has made a direct and substantial contribution to this work by providing the project idea, conception, design of the study, and critical revisions that are important for the intellectual content of the manuscript. AN made a direct and substantial contribution to this work by playing a leading role in the design of the study, data collection, interpretation of data, writing the manuscript, and critical revisions that are important for the intellectual content of the manuscript. All the authors have read and approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Interview guide for people with COPD that have used the COPD Web.
[PDF File (Adobe PDF File), 124 KB - jmir_v23i4e25672_app1.pdf]

References

1. World Health Organization. WHO - The Top 10 Causes of Death. 2018. URL: https://www.who.int/en/news-room/fact-sheets/detail/the-top-10-causes-of-death [accessed 2021-04-14]
2. Agusti A, Soriano JB. COPD as a systemic disease. COPD 2008 Apr;5(2):133-138. [doi: 10.1080/15412550801941349] [Medline: 18415812]
3. Wouters EFM. Chronic obstructive pulmonary disease. 5: systemic effects of COPD. Thorax 2002 Dec;57(12):1067-1070 [FREE Full text] [doi: 10.1136/thorax.57.12.1067] [Medline: 12454303]
4. Gadoury M, Schwartzman K, Rouleau M, Maltais F, Julien M, Beaupré A, Chronic Obstructive Pulmonary Disease axis of the Respiratory Health Network, Fonds de la recherche en santé du Québec (FRSQ). Self-management reduces both

short- and long-term hospitalisation in COPD. Eur Respir J 2005 Nov;26(5):853-857 [FREE Full text] [doi: 10.1183/09031936.05.00093204] [Medline: 16264046]
5. Kruis AL, Smidt N, Assendelft WJJ, Gussekloo J, Boland MRS, Rutten-van MM, et al. Cochrane corner: is integrated disease management for patients with COPD effective? Thorax 2014 Nov;69(11):1053-1055 [FREE Full text] [doi: 10.1136/thoraxjnl-2013-204974] [Medline: 24415716]
6. McCarthy B, Casey D, Devane D, Murphy K, Murphy E, Lacasse Y. Pulmonary rehabilitation for chronic obstructive pulmonary disease. Cochrane Database Syst Rev 2015 Feb 23(2):CD003793. [doi: 10.1002/14651858.CD003793.pub3] [Medline: 25705944]
7. Bourbeau J, Nault D, Dang-Tan T. Self-management and behaviour modification in COPD. Patient Educ Couns 2004 Mar;52(3):271-277. [doi: 10.1016/S0738-3991(03)00102-2] [Medline: 14998597]
8. Steurer-Stey C, Dalla Lana K, Braun J, Ter Riet G, Puhan MA. Effects of the "Living well with COPD" intervention in primary care: a comparative study. Eur Respir J 2018 Jan;51(1):1701375 [FREE Full text] [doi: 10.1183/13993003.01375-2017] [Medline: 29301921]
9. Murphy LA, Harrington P, Taylor SJ, Teljeur C, Smith SM, Pinnock H, et al. Clinical-effectiveness of self-management interventions in chronic obstructive pulmonary disease: An overview of reviews. Chron Respir Dis 2017 Aug;14(3):276-288 [FREE Full text] [doi: 10.1177/1479972316687208] [Medline: 28774200]
10. Spruit MA, Singh SJ, Garvey C, ZuWallack R, Nici L, Rochester C, et al. An official American Thoracic Society/European Respiratory Society statement: key concepts and advances in pulmonary rehabilitation. Am J Respir Crit Care Med 2013 Oct 15;188(8):e13-e64. [doi: 10.1164/rccm.201309-1634ST] [Medline: 24127811]
11. Socialstyrelsen. Vård vid astma och KOL. 2015. URL: https://www.socialstyrelsen.se/globalassets/sharepoint-dokument/artikelkatalog/nationella-riktlinjer/2015-11-2.pdf> [accessed 2020-05-25]
12. Arne M, Emtner M, Lisspers K, Wadell K, Ställberg B. Availability of pulmonary rehabilitation in primary care for patients with COPD: a cross-sectional study in Sweden. Eur Clin Respir J 2016;3:31601 [FREE Full text] [doi: 10.3402/ecrj.v3.31601] [Medline: 27900930]
13. Jácome C, Marques F, Paixão C, Rebelo P, Oliveira A, Cruz J, et al. Embracing digital technology in chronic respiratory care: Surveying patients access and confidence. Pulmonology 2020;26(1):56-59 [FREE Full text] [doi: 10.1016/j.pulmoe.2019.05.001] [Medline: 31160235]
14. van Zelst C, Van Noort E, Chavannes N, In 'T Veen H, Kasteleyn M. Blended care results in an improved adherence of an eHealth Platform by COPD patients. Eur Respir J 2020 Oct 28;56:1360 [FREE Full text] [doi: 10.1183/13993003.congress-2020.1360]
15. Van Noort E, Kasteleyn M, In 'T Veen H, Mennema B, Chavannes N. Selfmanagement by eHealth in asthma and COPD patients: with or without the professional. European Respiratory Society 2017 Dec 06;50(Suppl_61):PA1606 PA1606. [doi: 10.1183/1393003.congress-2017.PA1606]
16. Tistad M, Lundell S, Wiklund M, Nyberg A, Holmner, Wadell K. Usefulness and Relevance of an eHealth Tool in Supporting the Self-Management of Chronic Obstructive Pulmonary Disease: Explorative Qualitative Study of a Cocreative Process. JMIR Hum Factors 2018 Oct 26;5(4):e10801 [FREE Full text] [doi: 10.2196/10801] [Medline: 30368440]
17. Nyberg A, Tistad M, Wadell K. Can the COPD web be used to promote self-management in patients with COPD in swedish primary care: a controlled pragmatic pilot trial with 3 month- and 12 month follow-up. Scand J Prim Health Care 2019 Mar;37(1):69-82 [FREE Full text] [doi: 10.1080/02813432.2019.1569415] [Medline: 30700230]
18. Escoffery C. Gender Similarities and Differences for e-Health Behaviors Among U.S. Adults. Telemed J E Health 2017 Aug 16:335-343. [doi: 10.1089/tmj.2017.0136] [Medline: 28813630]
19. Morrison D, Mair FS, Yardley L, Kirby S, Thomas M. Living with asthma and chronic obstructive airways disease: Using technology to support self-management - An overview. Chron Respir Dis 2017 Nov;14(4):407-419 [FREE Full text] [doi: 10.1177/1479972316660977] [Medline: 27512084]
20. Kummervold PE, Chronaki CE, Lausen B, Prokosch H, Rasmussen J, Santana S, et al. eHealth trends in Europe 2005-2007: a population-based survey. J Med Internet Res 2008;10(4):e42 [FREE Full text] [doi: 10.2196/jmir.1023] [Medline: 19017584]
21. Kujala S. User involvement: A review of the benefits and challenges. Behaviour & Information Technology 2010 Nov 08;22(1):1-16 [FREE Full text] [doi: 10.1080/01449290301782]
22. Kayser L, Kushniruk A, Osborne RH, Norgaard O, Turner P. Enhancing the Effectiveness of Consumer-Focused Health Information Technology Systems Through eHealth Literacy: A Framework for Understanding Users' Needs. JMIR Hum Factors 2015 May 20;2(1):e9 [FREE Full text] [doi: 10.2196/humanfactors.3696] [Medline: 27025228]
23. Vorrink S, Huisman C, Kort H, Troosters T, Lammers J. Perceptions of Patients With Chronic Obstructive Pulmonary Disease and Their Physiotherapists Regarding the Use of an eHealth Intervention. JMIR Hum Factors 2017 Sep 19;4(3):e20 [FREE Full text] [doi: 10.2196/humanfactors.7196] [Medline: 28928110]
24. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357 [FREE Full text] [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]

25. Eldridge SM, Chan CL, Campbell MJ, Bond CM, Hopewell S, Thabane L, et al. CONSORT 2010 statement: extension to randomised pilot and feasibility trials. BMJ 2016 Oct 24;355:i5239 [FREE Full text] [Medline: 27777223]
26. Zwarenstein M, Treweek S, Gagnier JJ, Altman DG, Tunis S, Haynes B, et al. Improving the reporting of pragmatic trials: an extension of the CONSORT statement. BMJ 2008 Nov 11;337:a2390 [FREE Full text] [Medline: 19001484]
27. Nyberg A, Wadell K, Lindgren H, Tistad M. Internet-based support for self-management strategies for people with COPD-protocol for a controlled pragmatic pilot trial of effectiveness and a process evaluation in primary healthcare. BMJ Open 2017 Aug 01;7(7):e016851 [FREE Full text] [doi: 10.1136/bmjopen-2017-016851] [Medline: 28765136]
28. Jones PW, Harding G, Berry P, Wiklund I, Chen W, Kline LN. Development and first validation of the COPD Assessment Test. Eur Respir J 2009 Sep;34(3):648-654 [FREE Full text] [doi: 10.1183/09031936.00102509] [Medline: 19720809]
29. Mahler DA, Wells CK. Evaluation of clinical methods for rating dyspnea. Chest 1988 Mar;93(3):580-586. [doi: 10.1378/chest.93.3.580] [Medline: 3342669]
30. Wångdahl JM, Mårtensson LI. The communicative and critical health literacy scale--Swedish version. Scand J Public Health 2014 Feb;42(1):25-31. [doi: 10.1177/1403494813500592] [Medline: 23982461]
31. Grossoehme DH. Overview of qualitative research. J Health Care Chaplain 2014;20(3):109-122 [FREE Full text] [doi: 10.1080/08854726.2014.925660] [Medline: 24926897]
32. Cho J, Lee E. Reducing Confusion about Grounded Theory and Qualitative Content Analysis: Similarities and Differences. TQR 2014 Nov 08;19(32):1-20. [doi: 10.46743/2160-3715/2014.1028]
33. Graneheim UH, Lundman B. Qualitative content analysis in nursing research: concepts, procedures and measures to achieve trustworthiness. Nurse Educ Today 2004 Feb;24(2):105-112. [doi: 10.1016/j.nedt.2003.10.001] [Medline: 14769454]
34. Russell S, Ogunbayo OJ, Newham JJ, Heslop-Marshall K, Netts P, Hanratty B, et al. Qualitative systematic review of barriers and facilitators to self-management of chronic obstructive pulmonary disease: views of patients and healthcare professionals. NPJ Prim Care Respir Med 2018 Jan 17;28(1):2. [doi: 10.1038/s41533-017-0069-z] [Medline: 29343739]
35. Ritterband LM, Thorndike FP, Cox DJ, Kovatchev BP, Gonder-Frederick LA. A behavior change model for internet interventions. Ann Behav Med 2009 Aug;38(1):18-27 [FREE Full text] [doi: 10.1007/s12160-009-9133-4] [Medline: 19802647]
36. Schreiweis B, Pobiruchin M, Strotbaum V, Suleder J, Wiesner M, Bergh B. Barriers and Facilitators to the Implementation of eHealth Services: Systematic Literature Analysis. J Med Internet Res 2019 Nov 22;21(11):e14197 [FREE Full text] [doi: 10.2196/14197] [Medline: 31755869]
37. Fogg B. A Behavior Model for Persuasive Design. In: Proceedings of the Fourth International Conference on Persuasive Technology. New York, NY: ACM Press; 2009 Presented at: Fourth International Conference on Persuasive Technology, PERSUASIVE 2009; April 26-29, 2009; Claremont, CA, USA p. 09. [doi: 10.1145/1541948.1541999]
38. Nault D, Dagenais J, Perreault V, Pépin J, Labrecque S, Séguin M. Qualitative evaluation of a disease specific self-managementprogram “Living well with COPD©”. Eur Respir J 2000(16):317.
39. Bandura A. The self system in reciprocal determinism. American Psychologist 1978 Apr;33(4):344-358. [doi: 10.1037/0003-066X.33.4.344]
40. Kohler C, Fish L, Greene P. The relationship of perceived self-efficacy to quality of life in chronic obstructive pulmonary disease. Health Psychology 2002;21(6):610-614. [doi: 10.1037/0278-6133.21.6.610] [Medline: 12433014]
41. Lau-Walker M. Predicting self-efficacy using illness perception components: a patient survey. Br J Health Psychol 2006 Nov;11(Pt 4):643-661. [doi: 10.1348/135910705X72802] [Medline: 17032489]
42. Lorig KR, Ritter P, Stewart AL, Sobel DS, Brown BW, Bandura A, et al. Chronic disease self-management program: 2-year health status and health care utilization outcomes. Med Care 2001 Nov;39(11):1217-1223. [Medline: 11606875]
43. Marks R, Allegrante J, Lorig K. A review and synthesis of research evidence for self-efficacy-enhancing interventions for reducing chronic disability: implications for health education practice (part I). Health Promot Pract 2005 Jan;6(1):37-43. [doi: 10.1177/1524839904266790] [Medline: 15574526]
44. Sugawara E, Nikaido H. Properties of AdeABC and AdeIJK efflux systems of Acinetobacter baumannii compared with those of the AcrAB-TolC system of Escherichia coli. Antimicrob Agents Chemother 2014 Dec;58(12):7250-7257 [FREE Full text] [doi: 10.1128/AAC.03728-14] [Medline: 25246403]
45. Puente-Maestu L, Calle M, Rodríguez-Hermosa JL, Campuzano A, de Miguel Díez J, Álvarez-Sala JL, et al. Health literacy and health outcomes in chronic obstructive pulmonary disease. Respir Med 2016 Dec;115:78-82 [FREE Full text] [doi: 10.1016/j.rmed.2016.04.016] [Medline: 27215508]
46. Kale MS, Federman AD, Krauskopf K, Wolf M, O'Conor R, Martynenko M, et al. The Association of Health Literacy with Illness and Medication Beliefs among Patients with Chronic Obstructive Pulmonary Disease. PLoS One 2015;10(4):e0123937 [FREE Full text] [doi: 10.1371/journal.pone.0123937] [Medline: 25915420]
47. Stellefson M, Paige SR, Alber JM, Chaney BH, Chaney D, Apperson A, et al. Association Between Health Literacy, Electronic Health Literacy, Disease-Specific Knowledge, and Health-Related Quality of Life Among Adults With Chronic Obstructive Pulmonary Disease: Cross-Sectional Study. J Med Internet Res 2019 Jun 06;21(6):e12165 [FREE Full text] [doi: 10.2196/12165] [Medline: 31172962]

48. Ansari S, Hosseinzadeh H, Dennis S, Zwar N. Patients' perspectives on the impact of a new COPD diagnosis in the face of multimorbidity: a qualitative study. NPJ Prim Care Respir Med 2014 Aug 14;24:14036 [FREE Full text] [doi: 10.1038/npjpcrm.2014.36] [Medline: 25119845]
49. Omachi TA, Sarkar U, Yelin EH, Blanc PD, Katz PP. Lower health literacy is associated with poorer health status and outcomes in chronic obstructive pulmonary disease. J Gen Intern Med 2013 Jan;28(1):74-81 [FREE Full text] [doi: 10.1007/s11606-012-2177-3] [Medline: 22890622]
50. Leung L. Validity, reliability, and generalizability in qualitative research. J Family Med Prim Care 2015;4(3):324-327. [doi: 10.4103/2249-4863.161306] [Medline: 26288766]
51. Elo S, Kääriäinen M, Kanste O, Pölkki T, Utriainen K, Kyngäs H. Qualitative Content Analysis. SAGE Open 2014 Feb 11;4(1):1-10. [doi: 10.1177/2158244014522633]
52. Ward K, Gott M, Hoare K. Participants' views of telephone interviews within a grounded theory study. J Adv Nurs 2015 Dec;71(12):2775-2785. [doi: 10.1111/jan.12748] [Medline: 26256835]
53. Graneheim UH, Lindgren B, Lundman B. Methodological challenges in qualitative content analysis: A discussion paper. Nurse Educ Today 2017 Sep;56:29-34. [doi: 10.1016/j.nedt.2017.06.002] [Medline: 28651100]
54. Easton KL, McComish JF, Greenberg R. Avoiding common pitfalls in qualitative data collection and transcription. Qual Health Res 2000 Sep;10(5):703-707. [doi: 10.1177/104973200129118651] [Medline: 11066874]
55. Robinson O. Sampling in Interview-Based Qualitative Research: A Theoretical and Practical Guide. Qualitative Research in Psychology 2013 Nov 18;11(1):25-41. [doi: 10.1080/14780887.2013.801543]

Abbreviations

CAT: COPD Assessment Test
CCHL: Communicative and Critical Health Literacy Scale
COPD: chronic obstructive pulmonary disease
COREQ: Consolidated Criteria for Reporting Qualitative Research
FBM: Fogg Behavioral Model
FEV: forced expiratory volume
FVC: forced vital capacity
MRC: Medical Research Council
IT: information technology

Edited by R Kukafka; submitted 11.11.20; peer-reviewed by C Jacombe, S Alghamdi; comments to author 30.12.20; revised version received 23.02.21; accepted 18.03.21; published 30.04.21.

Please cite as:
Marklund S, Tistad M, Lundell S, Östrand L, Sörlin A, Boström C, Wadell K, Nyberg A
Experiences and Factors Affecting Usage of an eHealth Tool for Self-Management Among People With Chronic Obstructive Pulmonary Disease: Qualitative Study
J Med Internet Res 2021;23(4):e25672
URL: https://www.jmir.org/2021/4/e25672
doi:10.2196/25672
PMID:33929327

Original Paper

CANreduce 2.0 Adherence-Focused Guidance for Internet Self-Help Among Cannabis Users: Three-Arm Randomized Controlled Trial

Christian Baumgartner[1], MSc; Michael Patrick Schaub[1], PhD; Andreas Wenger[1], MSc; Doris Malischnig[2], PhD; Mareike Augsburger[1], PhD; Marc Walter[3], MD; Thomas Berger[4], PhD; Lars Stark[5], MD; David Daniel Ebert[6], PhD; Matthew T Keough[7], PhD; Severin Haug[1], PhD

[1]Swiss Research Institute for Public Health and Addiciton, University of Zurich, Zürich, Switzerland

[2]Institute for Addiction Prevention, Addiction and Drug Coordination Vienna, Vienna, Austria

[3]University Psychiatric Clinics, University of Basel, Basel, Switzerland

[4]Department of Clinical Psychology and Psychotherapy, University of Bern, Bern, Switzerland

[5]Arud Centre for Addiction Medicine, Zurich, Switzerland

[6]Department for Sport and Health Sciences, Chair for Psychology and Digital Mental Health Care, Technical University Munich, Munich, Germany

[7]Department of Psychology, York University, Toronto, ON, Canada

Corresponding Author:
Christian Baumgartner, MSc
Swiss Research Institute for Public Health and Addiciton
University of Zurich
Konradstrasse 32
Zürich, 8005
Switzerland
Phone: 41 444481160
Email: christian.baumgartner@isgf.uzh.ch

Abstract

Background: Despite increasing demand for treatment among cannabis users in many countries, most users are not in treatment. Internet-based self-help offers an alternative for those hesitant to seek face-to-face therapy, though low effectiveness and adherence issues often arise.

Objective: Through adherence-focused guidance enhancement, we aimed to increase adherence to and the effectiveness of internet-based self-help among cannabis users.

Methods: From July 2016 to May 2019, cannabis users (n=775; male: 406/575, 70.6%, female: 169/575, 29.4%; age: mean 28.3 years) not in treatment were recruited from the general population and were randomly assigned to (1) an adherence-focused guidance enhancement internet-based self-help intervention with social presence, (2) a similar intervention with an impersonal service team, and (3) access to internet as usual. Controls who were placed on a waiting list for the full intervention after 3 months underwent an assessment and had access to internet as usual. The primary outcome measurement was cannabis-use days over the preceding 30 days. Secondary outcomes included cannabis-dependence severity, changes in common mental disorder symptoms, and intervention adherence. Differences between the study arms in primary and secondary continuous outcome variables at baseline, posttreatment, and follow-up were tested using pooled linear models.

Results: All groups exhibited reduced cannabis-use days after 3 months (social presence: –8.2 days; service team: –9.8 days; internet as usual: –4.2 days). The participants in the service team group (P=.01, d=.60) reported significantly fewer cannabis-use days than those in the internet as usual group; the reduction of cannabis use in the social presence group was not significant (P=.07, d=.40). There was no significant difference between the 2 intervention groups regarding cannabis-use reduction. The service team group also exhibited superior improvements in cannabis-use disorder, cannabis-dependence severity, and general anxiety symptoms after 3 months to those in the internet as usual group.

Conclusions: The adherence-focused guidance enhancement internet-based self-help intervention with an impersonal service team significantly reduced cannabis use, cannabis-use disorder, dependence severity, and general anxiety symptoms.

Trial Registration: ISRCTN Registry ISRCTN11086185; http://www.isrctn.com/ISRCTN11086185

(J Med Internet Res 2021;23(4):e27463) doi:10.2196/27463

KEYWORDS

cannabis; common mental disorders; adherence; social presence; internet; cognitive behavioral therapy; motivational interviewing; therapy; mental health; mental disorder; adherence; guidance; self-help; drug abuse; randomized controlled trial

Introduction

Cannabis is the most consumed illicit drug in Europe, having witnessed a steady increase in recent years, evidenced by roughly 24.7 million European users in 2019 [1]. The global number of cannabis users was estimated as 178 million people in 2017 [2]. As more and more countries consider decriminalization or outright legalization, it seems unlikely that the increase in cannabis users will stagnate soon [3]. However, only a minority of cannabis users seem to develop cannabis dependence; in general population surveys, the risk of becoming dependent on cannabis appears to be between 10% and 11% of all cannabis users [4,5]. However, for cannabis users who start at a young age, the risks of cannabis dependence [6] and cannabis use problems [7] are significantly higher. In addition, poorer mental and physical health, lower educational attainment, and reduced cognitive performance than non-cannabis users are common among daily cannabis users [8]. Numerous studies [9] also point to a broad range of often co-occurring mental health disorders, such as depression, anxiety, and posttraumatic stress disorder, during the treatment of problematic cannabis use.

Treatment demand in Europe for first-time admissions with cannabis listed as the main problem substance has been increasing steadily, having almost doubled from roughly 45,000 in 2006 to approximately 83,000 in 2017 [1]. However, it is clear that, although the number of clients seeking treatment has increased, they still account for just a small minority of cannabis users who could potentially benefit from treatment, with or without comorbid mental health problems [9]. Similarly, only a few consumers seek professional medical assistance [10], suggesting that a broader range of treatment options should be provided [11]. Various potential barriers prevent people from seeking treatment, including poor accessibility to treatment centers, the lack of awareness of negative health consequences, the wish to reduce cannabis use on their own [12], and fear of stigmatization as a drug addict, which seems to be a major factor [13,14]. Facilitators of treatment, on the other hand, include improving available information, increased access to cannabis-specific services, providing additional treatment options, and making admissions easier [15], all of which many internet-based interventions could provide.

Studies on web-based interventions for which participants were recruited from the general adult population (>18 years old) have been shown to draw a cannabis-using population that is different than those entering outpatient addiction treatment centers, not only in terms of having a higher level of education and being older, but also in terms of reporting more frequent cannabis use [14,16]. However, poor adherence to the intervention is often found in these studies [17,18]. Moreover, a recent meta-analysis [19] on internet-based treatments for cannabis users yielded significant but only small effect sizes for the reduction of cannabis use (mostly frequency) in the short term (15 comparisons, Hedges g=0.12) that could not be maintained longer term (12 months). The effects of multisession interventions, such as those combining cognitive behavioral therapy [20] with motivational interviewing [21], produced larger effect sizes (6 comparisons, Hedges g=0.18) than single-session interventions using approaches like brief interventions [22] and motivational interviewing (13 comparisons, Hedges g=0.09). Among the studies assessing multisession interventions, only 2 took symptoms of possible co-occurring mental health disorders into account [14,23].

In previous studies [13,14], called CANreduce 1.0, we were able to show that additional professional chat sessions increased the effectiveness of an internet-based self-help program designed to reduce cannabis use. The study [14] also found that participants who had the opportunity but did not participate in these chat sessions, nevertheless reduced their cannabis use more than those who only received internet-based self-help from the beginning. It seems that, on its own, having a professional therapist send chat invitations helped to reduce cannabis use in cannabis users. Since only a quarter of the participants in the treatment arm with chat took part in at least one chat appointment, we wondered whether the same effect could be achieved by replacing the professional therapist with a virtual eCoach. We also found that almost half (44.8%) of participants screened positive for clinically relevant depression symptoms at baseline [14]. Comorbidity of depressive symptoms and substance use and its hindrance on positive treatment outcomes has repeatedly been demonstrated [24].

CANreduce 2.0, a minimally guided internet-based self-help intervention for cannabis users, is designed to overcome the issues of low intervention adherence and effectiveness, as well as to address frequently co-occurring mental health disorders. This intervention is based on adherence-focused guidance which has, to date, never been tested as a component of an internet intervention for individuals with a substance use disorder but has been documented to be effective at increasing adherence to web-based self-help for the reduction of stress and depression symptoms [25,26]. The concept of adherence-focused guidance enhancement is primarily based on the supportive-accountability model of guidance in web-based interventions [27], which argues that adherence to internet-based interventions relies on an online coach (eCoach) who is seen as trustworthy, benevolent, and having expertise, and who has clear, process-oriented expectations in a reciprocal eCoach–participant relationship. In addition to an eCoach, we incorporated cognitive behavioral therapy–based approaches [28-31] into the program to target issues that potentially help to ameliorate overlapping common mental disorder symptoms, such as inactivity, depressed mood, excessive rumination, and difficulty relaxing.

The primary goal of this study was to investigate whether intervention effectiveness and program adherence can be increased by implementing adherence-focused guidance and

emphasizing the social presence factor of a personal eCoach when compared with a general support team implementation.

Methods

Study Design

This study was a 3-arm, randomized controlled trial that compared 2 versions of a minimally guided web-based self-help intervention for cannabis users based on adherence-focused guidance, cognitive behavioral therapy, motivational interviewing, and social presence factor—one combined with a personal eCoach (social presence), and one with a general support team (service team)—to a wait-list control group that underwent an assessment and had access to internet as usual (internet as usual) for the purpose of reducing cannabis use and associated mental health problems. The internet as usual group was able to use the internet to search for additional support and information regarding cannabis use from other online resources. Each intervention lasted for 6 weeks and was followed by a posttreatment survey and a follow-up survey 3 months postbaseline.

Participants were randomized, by computer, to the 3 conditions in a 1:1:1 ratio. Participants in the social presence and service team groups did not know to which program version they had been assigned, while participants in the internet as usual group knew they had been assigned to treatment as usual. The study was approved by the ethics committee of the Canton of Zurich on July 4, 2016 (BASEC 2016-00264) and registered (ISRCTN11086185). A detailed study protocol has been published [32].

Recruitment and Inclusion and Exclusion Criteria

We recruited participants from August 2016 through May 2019 with 2 websites, advertisements in relevant internet forums and newspapers (or online versions thereof), and search engine website advertisements. Study inclusion and exclusion criteria, and the rationale behind them, are summarized in Table 1.

Table 1. Inclusion and exclusion criteria and underlying rationale.

Criteria	Reasoning
Inclusion	
Informed consent via the web form	To ensure knowledge of procedures and the declaration of consent
Minimum age: 18 years	To ensure a minimum age of participation
Cannabis use at least once weekly over the last 30 days	To include participants with less than daily cannabis use, increase validity
At least once weekly internet access and a valid email address	To ensure at least some access to the intervention
Good command of the German language	To ensure that participants will be able to understand the information provided
Exclusion	
Participation in other psychosocial or pharmacological treatments for the reduction or cessation of cannabis use	To avoid confounding treatment effects
Current pharmacologically treated psychiatric disease or any history of psychosis, schizophrenia, bipolar type I disorder or significant current suicidal or homicidal thoughts	To avoid having participants with these problems enter the study

For compensation, all participants who completed the final follow-up evaluation were offered a choice of either an online voucher worth 30 € (approximately US $35.85) or donating that amount to charity.

Sample Size Calculation

We anticipated that a Cohen *d*=0.30 was appropriate, by employing previous internet-based studies in cannabis users [16,32] recruited from the general population, and adherence-focused guidance enhancement internet-based studies among individuals with stress or depression [25,26] to estimate effect-size differences between the adherence-focused guidance-enhanced version with (social presence) versus without (service team) a personal eCoach. This resulted in a sample size of 176 for each study arm (n=528 in total) to detect a small effect size with 80% power and an alpha error of 5% (2-tailed testing).

Treatment Arms

Both active interventions—social presence and service team—consisted of a dashboard and 8 self-help intervention modules that included stories of 6 fictional companions who appeared within the modules at key points, with the goal of encouraging reflection on potential questions raised by the modules. Table 2 provides an overview of the modules' contents and underlying therapeutic approaches. Both active interventions also incorporated a use and activity diary, weekly semiautomated motivational and adherence-focused guidance–based email feedback, and a section containing educational information on cannabis and health. The semiautomated motivational emails were triggered by a moderator, depending on how participants responded in exercises. These feedback emails also included module suggestions, dealing with high-risk situations, cravings, or the pros and cons of their use. Participants in both active intervention groups also were invited to ask questions of either their eCoach (social presence group) or support team (service team group) whenever they felt the need.

Table 2. Modules.

Module	Content	Therapeutic approach
Module 1: Introduction	• General overview • Introduction of fictional companions • Reflection on personal cannabis use	Based on motivational interviewing techniques [21]
Module 2 : Identifying risk situations	• Identifying personal high-risk situations • Recognizing seemingly irrelevant, but triggering decisions	Cognitive behavioral therapy approach to relapse Prevention [28]
Module 3: Working on needs	• Strengthening social contacts • Decreasing excessive ruminations • Developing healthier sleeping habits	Behavioral activation approach [29]
Module 4: Craving	• Concept of craving • Ways to deal with feelings of craving	Based on cognitive behavioral therapy [30]
Module 5: Dealing with relapses	• Relapse prevention • Dealing with relapses	Cognitive behavioral therapy approach to Relapse Prevention [28]
Module 6 : Working on problems	• Relationships between use, problems, and depressive symptoms • Skills to deal with solvable and unsolvable problems	Social problem-solving approach [31]
Module 7: Saying “no”; refusal skills	• Strengthening refusal skills for use in high-risk situations	Based on cognitive behavioral therapy [30]
Module 8: Preserving achievements	• Review of program • List of 5 personalized points to help secure achievements after the program is complete	Based on motivational interviewing techniques [21]

The only difference between the social presence and service team interventions was that, for the social presence group, a semiautomated eCoach, with short personal introduction videos (Multimedia Appendix 1) that preceded most of the modules and a picture of the female eCoach displayed on the dashboard, was used, while for the service team group, an anonymous semiautomated support team, with no pictures, videos, or any other kind of social presence, was used.

The control group had access to the internet as usual, since it was deemed impossible and unethical to prevent participants in this group from seeking out other internet support or face-to-face treatment options during the waiting period. A detailed description of the groups and their technical specifications is provided in the study protocol [32].

CANreduce 2.0 is regarded as a medical device and is CE (*Conformité Européenne*) certified.

Measurements

Table 3 provides an overview of the measurement instruments. The primary outcome of interest was the number of days of cannabis use over the preceding 30 days, in accordance with the timeline follow-back method [33,34]. Secondary outcomes included the severity of cannabis-use disorder assessed using the Cannabis Use Disorders Identification Test-Revised (CUDIT-R [35]); the Severity of Dependence Scale (SDS [36]); quantity of cannabis use over the previous 30 days using the timeline follow-back method, quantified in individually standardized cannabis joint sizes (detailed description in the study protocol [32]); the use of alcohol, tobacco, or other illicit drugs besides cannabis (with questions derived from the European Adaptation of a Multidimensional Assessment Instrument for Drug and Alcohol Dependence [37]); change in depression (Centre of Epidemiologic Studies of Depression, CES-D, scale [38], anxiety (Generalized Anxiety Disorder–7, GAD-7 [39]), and attention deficit and hyperactivity symptoms (adult Attention Deficit and Hyperactivity Self-Report Scale, ASRS, version 1.1) [40]; Short Screening Scale for lifetime Diagnostic and Statistical Manual of Mental Disorders, fourth edition, Posttraumatic Stress Disorder [41]; client satisfaction questionnaire [42]; Working Alliance Inventory adapted for web-based interventions [43], and treatment adherence (finished modules, time spent on modules). Furthermore, the occurrence of any negative intervention effects was identified using a questionnaire [44] at the 3-month follow-up assessment. We asked all participants if they had used any treatment other than CANreduce during the 3 months and, if so, to identify it from a predefined list of services. Details regarding study measures are reported in the study protocol [32].

Table 3. Assessment instruments.

Assessment instruments	Assessment
Sociodemographic data	Baseline
Center for Epidemiologic Studies Depression scale (range 0-60)	Baseline
Short Screening Scale for DSM-IV[a] Posttraumatic Stress Disorder (range 7-28)	Baseline
General Anxiety Disorder–7 (range 0-21)	Baseline
Adult ADHD[b] Self-Report Scale version 1.1 (range 0-24)	Baseline
Frequency of cannabis use (last 30 days)	Baseline, posttreatment, 3-months follow-up
Cannabis use according to the timeline follow-back method	Baseline, posttreatment, 3-months follow-up
Cannabis Use Disorder Identification Test-Revised (range 0-40)	Baseline, posttreatment, 3-months follow-up
Severity of Dependence Scale (range 0-15)	Baseline, posttreatment, 3-months follow-up
Fragebogen Substanzanamnese (substance use questionnaire)	Baseline, posttreatment, 3-months follow-up
Client Satisfaction Questionnaire-I (range 8-32)	Posttreatment
Intervention adherence (continuous assessment over 6-week treatment)	Posttreatment
WAI-TECH[c] (range 10-70)	Posttreatment
Negative effects [44]	3 months follow-up

[a]DSM-IV: Diagnostic and Statistical Manual of Mental Disorders, fourth edition.

[b]ADHD: attention deficit hyperactivity disorder.

[c]Working Alliance Inventory adapted for web-based interventions.

Statistical Analysis

Data were analyzed according to intention-to-treat (ITT). To address missing data for the ITT analyses, we applied multiple imputation procedures using the multivariate imputation by chained equations software package [45] in R (version 3.6.1; R Foundation for Statistical Computing), a minor deviation from the study protocol, which involves specifying a multivariate distribution for the missing data and drawing imputations from their conditional distributions using Markov chain Monte Carlo techniques. As recommended, 20 imputation sets were employed [45]. All sociodemographic, as well as primary and secondary outcome variables that had been assessed in all 3 groups, were included in the imputation. Reported outcomes use the ITT results from the imputed data sets, but complete case analysis results are also reported.

Cohen *d* effect size was used to compare change scores between the 3 treatment arms. As suggested elsewhere, Cohen d=0.2 indicates a small effect, Cohen d=0.5 indicates a medium effect, and Cohen d=0.8 indicates a large effect [46].

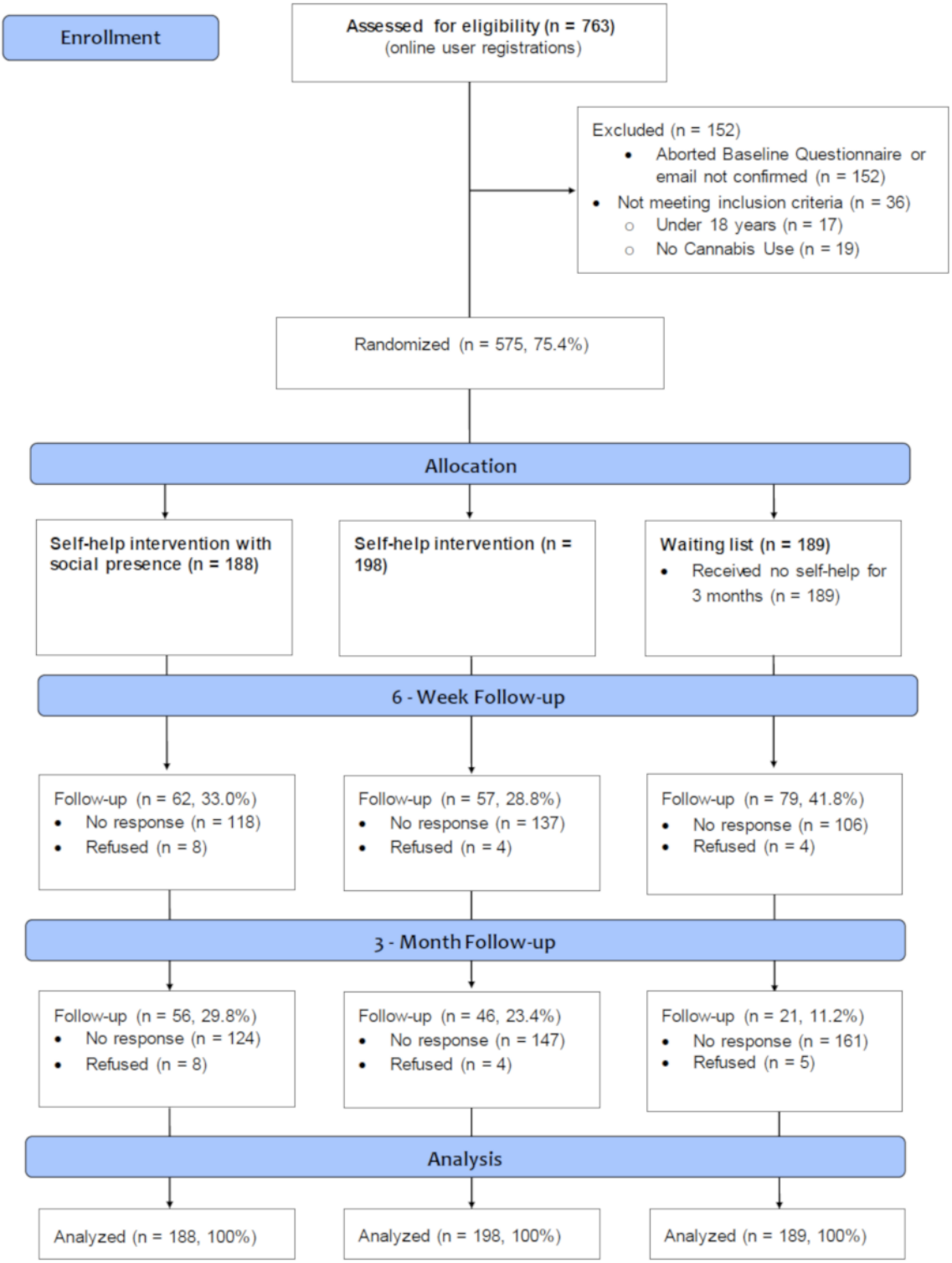
Differences between the study arms in primary and secondary continuous outcome variables at baseline, posttreatment, and follow-up were tested using pooled linear models. Change scores from baseline for primary and secondary outcomes were dependent variables, and study condition was the independent variable; all controlled for the baseline value of the respective outcome variable. Effect sizes were calculated for changes from baseline to follow-up (within-group effect size d_w) and between the 2 intervention groups (social presence, service team) and internet as usual. All P values are 2-sided with no adjustment made for multiple comparisons, which was deemed unnecessary based on CONSORT recommendations [47].

Results

Participation

Between July 2016 and May 2019, a total of 763 people registered online for the program, among whom 575 were randomized to the 3 study arms (Figure 1). All participants received email reminders for follow-ups and subsequent telephone calls if they did not complete the survey. We were able to reach 198 participants (34.4% of the initial sample) posttreatment, and this number dropped to 123 (21.4%) for the final assessment 3 months postbaseline. A coding bug arose within the email system which affected distribution of the final assessment questionnaire to the internet as usual group (only telephone follow-ups were performed). There were significant differences in follow-up rate between the study groups (χ^2=20.16, P<.001) that may have been caused by this bug.

Figure 1. Participation flowchart.



Baseline Characteristics

Of the 575 participants, 406 (70.6%) were male, and the average age was 28.3 (SD 7.9). Most (n=234, 40.7%) were from Switzerland, followed closely by 216 (37.6%) from Austria, and 121 (21.0%) from Germany. The average participant had used cannabis almost daily (25.7 days, SD 5.9) over the preceding 30 days. Complete case analyses of baseline data and study group comparisons are summarized in Table 4.

Table 4. Baseline participant data.

Characteristic	Social presence (n=188)	Service team (n=198)	Internet as usual (n=189)	All (N=575)	*F* test (*df1*,*df2*)	*P* value
Gender, n (%)					0.57 (2,575)[a]	.75
Female	52 (27.7)	58 (29.3)	59 (31.2)	169 (29.4)		
Male	136 (72.3)	140 (70.7)	130 (68.8)	406 (70.6)		
Age, mean (SD)	28.0 (7.4)	27.9 (7.9)	28.9 (8.3)	28.3 (7.9)	1.01 (2,572)	.36
Highest education, n (%)					14.20 (10,575)[a]	.16
Primary school	12 (6.4)	17 (8.6)	10 (5.3)	39 (6.8)		
Apprenticeship	31 (16.5)	36 (18.2)	53 (28.0)	120 (20.9)		
Secondary school	64 (34.0)	56 (28.3)	46 (24.3)	166 (28.9)		
Technical college	25 (13.3)	32 (16.2)	27 (14.3)	84 (14.6)		
University	49 (26.1)	48 (24.2)	49 (25.9)	146 (25.4)		
Not specified	7 (3.7)	9 (4.5)	4 (2.1)	20 (3.5)		
Country of origin, n (%)					9.73 (6,575)[a]	.14
Switzerland	74 (39.4)	77 (38.9)	83 (43.9)	234 (40.7)		
Austria	67 (35.6)	72 (36.4)	77 (40.7)	216 (37.6)		
Germany	45 (23.9)	49 (24.7)	27 (14.3)	121 (21.0)		
Other	2 (1.1)	0 (0.0)	2 (1.1)	4 (0.7)		
Centre for Epidemiological Studies Depression scale, mean (SD)	20.4 (10.0)	23.5 (11.1)	21.7 (10.1)	21.9 (10.5)	4.48 (2,572)	.01
Generalized Anxiety Disorder–7, mean (SD)	7.4 (4.8)	7.9 (5.0)	7.6 (4.5)	7.7 (4.8)	0.55 (2,571)	.58
Cannabis-use disorder, mean (SD)	20.8 (5.5)	20.7 (5.8)	21.6 (5.3)	21.0 (5.5)	1.41 (2,572)	.24
Severity of Dependence scale, mean (SD)	7.4 (3.1)	8.1 (3.3)	8.1 (3.2)	7.9 (3.2)	3.16 (2,572)	.04
Adult ADHD[b] Self-Report Scale, mean (SD)	10.7 (3.9)	10.8 (4.1)	10.8 (4.2)	10.8 (4.1)	0.09 (2,572)	.92
Short Screening Scale for PTSD[c], mean (SD)	13.2 (4.5)	12.8 (5.2)	13.8 (5.6)	13.2 (5.1)	0.56 (2,174)	.57
Number of cannabis joints, mean (SD)	22.6 (16.0)	21.3 (15.2)	23.6 (17.7)	22.5 (16.3)	0.89 (2,572)	.41
Number of cannabis-use days, mean (SD)	24.9 (6.7)	26.1 (5.3)	26.2 (5.5)	25.7 (5.9)	2.65 (2,572)	.07
Number of years of use, mean (SD)						
Cannabis	8.5 (7.2)	7.6 (6.6)	9.1 (6.7)	8.4 (6.8)	2.62 (2,569)	.07
Alcohol	5.5 (7.0)	4.7 (6.2)	5.4 (7.1)	5.2 (6.8)	0.82 (2,525)	.44
Alcohol risky use[d]	1.7 (3.9)	1.6 (3.7)	1.9 (4.3)	1.7 (3.9)	0.19 (2,500)	.83
Cocaine	0.6 (2.6)	0.4 (2.2)	0.3 (1.0)	0.4 (2.0)	1.14 (2,496)	.32

[a]Chi-square test (*df*).

[b]ADHD: attention deficit hyperactivity disorder.

[c]PTSD: posttraumatic stress disorder.

[d]Risky use is defined as 5 or more standard drinks per day on at least 3 days a week. A standard drink is defined as 50 mL spirits, 150-200 mL wine, or 330-450 mL beer.

Primary Outcome: Cannabis-Use Days

Immediately posttreatment, both intervention groups (social presence: mean 8.0, SD 9.3, d_w=.89; service team: mean 10.7 days, SD 9.5, d_w=1.18) reduced their cannabis use significantly more than internet as usual (mean 3.8, SD 8.1, d_w=.55) (social presence: B=–4.34, CI –7.21 to –1.47, *P*=.004, between-group effect size *d*=.48; service team: B=–6.43, CI –9.87 to –2.97, *P*<.001, *d*=.71). These effects persisted 3 months postbaseline, with participants in the service team (mean 9.8, SD 9.9, d_w=1.18) group still reducing their cannabis-use days significantly more (B=–5.70, CI –10.09 to –1.30, *P*=.01, *d*=.60) than in the control group (mean 4.2 days, SD 8.8, d_w=.55). Similarly, there was a significantly greater reduction in the social presence group (mean 8.2 days, SD 9.8, d_w=.93) than in the control group (B=–4.41, CI –9.19 to 0.37, *P*=.07, *d*=.40). There was no significant difference between the 2 intervention groups immediately posttreatment (*P*=.26) or 3 months postbaseline (*P*=.44) (Figure 2; Table 5).

Figure 2. Cannabis use in the previous 30 days. AFGE-SP: adherence-focused guidance enhancement with social presence; AFGE-ST: adherence-focused guidance enhancement with service team; IAU: internet as usual.

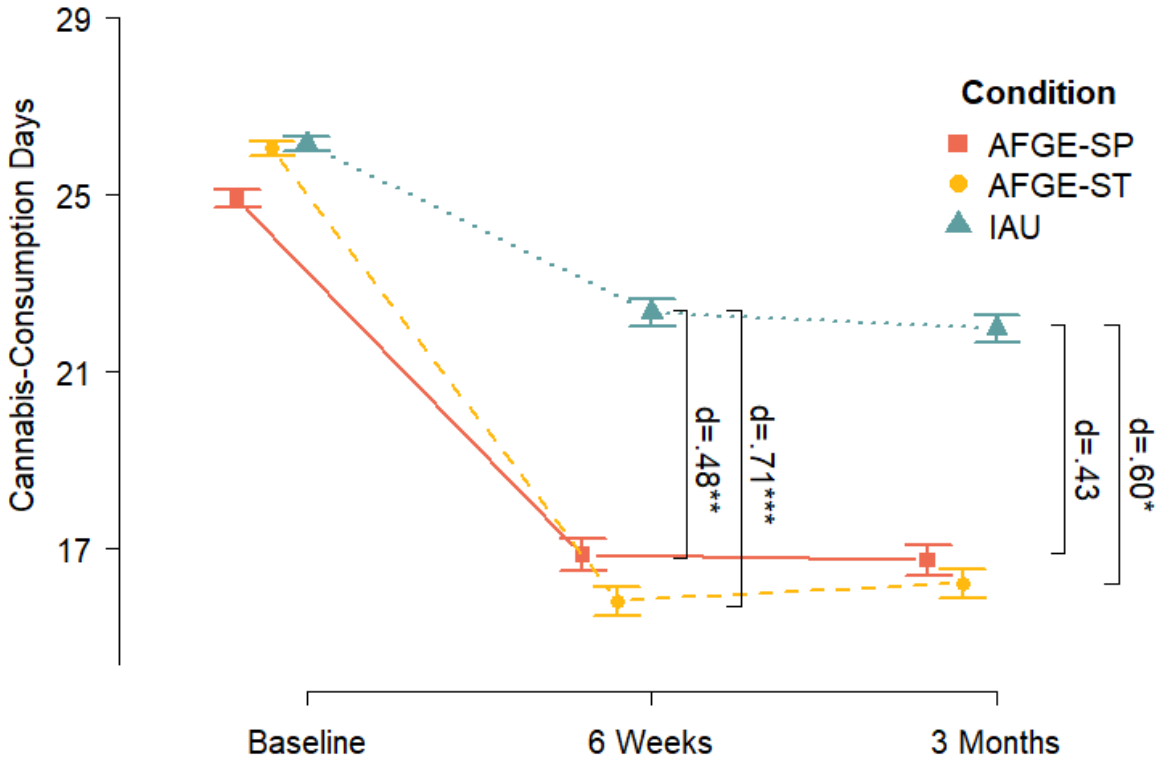


Table 5. Regression analysis (intention to treat).

Variable	Social presence		Service team	
	B (95% CI)	*P* value	B (95% CI)	*P* value
Consumption days[a]	–4.41 (–9.19, 0.37)	.07	–5.70 (–10.09, –1.32)	.01
Cannabis Use Disorder Identification Test	–2.15 (–4.49, 0.18)	.07	–3.39 (–5.96, –0.83)	.01
Severity of Dependence Scale	–1.49 (–3.45, 0.47)	.13	–2.16 (–3.90, –0.42)	.02
Generalized Anxiety Disorder–7	–2.07 (–3.59, –0.55)	.001	–2.87 (–4.43, –1.31)	<.001
Adult ADHD[b] Self-Report Scale	0.63 (–1.68, 2.94)	.57	–0.51 (–2.59, 1.56)	.61
Centre for Epidemiological Studies Depression scale	0.41 (–3.49, 4.32)	.83	–1.76 (–5.99, 2.47)	.40

[a]Previous 30 days according to the timeline follow-back method.

[b]ADHD: attention deficit hyperactivity disorder.

Secondary Outcomes

At follow-up 3 months postbaseline, a significant difference was noted in the decrease in cannabis-use disorder severity between the service team group and controls (B=–3.39, CI –5.96 to –0.83, *P*=.01, *d*=.52), but no significant difference was detected between the social presence and the internet as usual group (*P*=.07). Additionally, there was a significantly greater reduction in the severity of dependence in the service team group compared to the internet as usual group (B=–2.16, CI –3.90 to –0.42, *P*=.02, *d*=.60). There were significantly greater reductions in general anxiety disorder symptoms in both groups (social presence: B=–2.70, CI –3.59 to –0.55, *P*=.001, *d*=.41; service team: B=–2.87, CI –4.43 to –1.31, *P*<.001, *d*=.51) compared to the internet as usual group. All groups decreased their mean ASRS score, but no significant intergroup differences were detected (social presence: mean 0.9, SD 3.9, *P*=.57; service team: mean 2.1, SD 4.2, *P*=.61; internet as usual: mean 1.6, SD 4.0). Similarly, mean CES-D scores decreased in all groups (social presence: mean 3.6, SD 9.6, *P*=.83; service team: mean 7.5, SD 10.3, *P*=.40; internet as usual: mean 4.7, SD 10.4) with no significant between-group differences (Table 6 and Table 7). Upon review, the number of standard joints was deemed unreliable as a measurement and was dropped from analysis.

Table 6. Complete case analysis.

Group and measure	Baseline	Posttreatment		Follow-up	
	Mean (SD)	Mean (SD)	*d*[a] (95% CL)	Mean (SD)	*d* (95% CL)
Internet as usual (n=189)					
Use days[b]	26.16 (5.53)	22.58 (9.29)	N/A[c]	21.81 (8.30)	N/A
CUDIT[d]	21.57 (5.30)	20.11 (6.49)	N/A	18.48 (5.61)	N/A
SDS[e]	8.12 (3.16)	7.79 (3.41)	N/A	7.00 (3.33)	N/A
GAD-7[f]	7.61 (4.54)	N/A	N/A	7.14 (2.53)	N/A
ASRS[g]	10.84 (4.21)	N/A	N/A	9.71 (4.22)	N/A
CES-D[h]	21.69 (10.12)	N/A	N/A	15.67 (6.19)	N/A
Social presence (n=188)					
Use days	24.92 (6.67)	15.74 (10.95)	0.53 (0.18, 0.85)	16.54 (10.51)	0.32 (–0.19, 0.71)
CUDIT	20.79 (5.46)	16.90 (6.14)	0.54 (0.19, 0.87)	14.88 (6.73)	0.61 (0.09, 1.11)
SDS	7.38 (3.10)	5.85 (3.26)	0.37 (0.03, 0.70)	5.14 (3.15)	0.39 (–0.12, 0.89)
GAD-7	7.42 (4.79)	N/A	N/A	5.26 (3.74)	0.19 (–0.32, 0.69)
ASRS	10.69 (3.89)	N/A	N/A	9.46 (3.66)	–0.13 (–0.63, 0.37)
CES-D	20.35 (10.00)	N/A	N/A	14.68 (9.76)	–0.26 (–0.76, 0.25)
Service team (n=198)					
Use days	26.06 (5.25)	15.97 (10.51)	0.73 (0.36, 1.06)	17.37 (10.44)	0.34 (–0.18, 0.86)
CUDIT	20.71 (5.76)	17.19 (6.00)	0.52 (0.17, 0.86)	14.63 (6.66)	0.69 (0.15, 1.20)
SDS	8.06 (3.26)	6.29 (3.34)	0.63 (0.27, 0.97)	5.38 (3.24)	0.48 (–0.06, 0.99)
GAD-7	7.92 (4.97)	N/A	N/A	5.20 (4.24)	0.57 (0.04, 1.09)
ASRS	10.84 (4.11)	N/A	N/A	8.96 (3.92)	0.31 (–0.22, 0.82)
CES-D	23.51 (11.06)	N/A	N/A	16.10 (10.23)	0.26 (–0.27, 0.77)

[a]Effect size Cohen *d* based on differences between the intervention and control groups.

[b]Previous 30 days according to the timeline follow-back method.

[c]N/A: not applicable.

[d]CUDIT: Cannabis Use Disorder Identification Test.

[e]SDS: Severity of Dependence Scale.

[f]GAD-7: Generalized Anxiety Disorder–7.

[g]ASRS: Adult Attention Deficit and Hyperactivity Disorder Self-Report Scale.

[h]CES-D: Centre for Epidemiological Studies Depression scale.

Table 7. Imputed data analysis.

Group and measure	Baseline		Posttreatment		Follow-up
	mean (SD)	mean (SD)	d[a] (95% CL)	mean (SD)	d (95% CL)
Internet as usual (n=189)					
Use days[b]	26.16 (5.53)	22.34 (9.47)	N/A[c]	21.98 (9.33)	N/A
CUDIT[d]	21.57 (5.30)	20.09 (6.31)	N/A	18.48 (6.31)	N/A
SDS[e]	8.12 (3.16)	7.67 (3.42)	N/A	7.38 (3.56)	N/A
GAD-7[f]	7.61 (4.54)	N/A	N/A	7.62 (3.84)	N/A
ASRS[g]	10.84 (4.21)	N/A	N/A	9.27 (4.04)	N/A
CES-D[h]	21.69 (10.12)	N/A	N/A	16.97 (9.66)	N/A
Social presence (n=188)					
Use days	24.92 (6.67)	16.88 (10.93)	0.48 (0.27, 0.68)	16.74 (10.54)	0.43 (0.22, 0.63)
CUDIT	20.79 (5.46)	17.07 (6.27)	0.50 (0.28, 0.69)	15.84 (6.68)	0.31 (0.10, 0.51)
SDS	7.38 (3.10)	5.58 (3.22)	0.46 (0.25, 0.66)	5.52 (3.34)	0.32 (0.12, 0.52)
GAD-7	7.42 (4.79)	N/A	N/A	5.48 (3.84)	0.42 (0.21, 0.62)
ASRS	10.69 (3.89)	N/A	N/A	9.82 (3.92)	–0.18 (–0.38, 0.03)
CES-D	20.35 (10.00)	N/A	N/A	16.74 (9.97)	–0.11 (–0.31, 0.11)
Service team (n=198)					
Use days	26.06 (5.25)	15.82 (10.59)	0.71 (0.50, 0.91)	16.21 (10.57)	0.60 (0.39, 0.80)
CUDIT	20.71 (5.76)	17.08 (6.46)	0.47 (0.26, 0.66)	14.56 (7.18)	0.49 (0.28, 0.68)
SDS	8.06 (3.26)	5.76 (3.32)	0.62 (0.40, 0.81)	5.16 (3.32)	0.60 (0.38, 0.79)
GAD-7	7.92 (4.97)	N/A	N/A	4.87 (3.90)	0.67 (0.46, 0.87)
ASRS	10.84 (4.11)	N/A	N/A	8.75 (3.82)	0.13 (–0.08, 0.32)
CES-D	23.51 (11.06)	N/A	N/A	16.04 (10.07)	0.27 (0.06, 0.43)

[a]Effect size Cohen *d* based on differences between the intervention and control groups.

[b]Previous 30 days according to the timeline follow-back method.

[c]N/A: not applicable.

[d]CUDIT: Cannabis Use Disorder Identification Test.

[e]SDS: Severity of Dependence Scale.

[f]GAD-7: Generalized Anxiety Disorder–7.

[g]ASRS: Adult Attention Deficit and Hyperactivity Disorder Self-Report Scale.

[h]CES-D: Centre for Epidemiological Studies Depression scale.

Adherence and User Satisfaction

Participants in the social presence group completed an average of 2.6 (SD 2.6) modules versus 2.4 modules (SD 2.5) completed in the service team group (t_{374}=0.85, *P*=.20). Social presence group members (mean 46.9 minutes) spent significantly more time than service team users (37.5 minutes) on the program (t_{374}=2.04, *P*=.02). Social presence group members exhibited the highest retention rate (29.8% versus 23.4% in the service team group), but this difference was not significant (χ^2=1.81 *P*=.18).

There was no significant difference between the 2 intervention groups in level of user satisfaction (*P*=.83), but there were significant differences between each of the 2 intervention groups and internet as usual (internet as usual: mean 12.9, SD 6.9; social presence: mean 24.6, SD 5.4, t_{139}=11.36, *P*<.001; service team: mean 25.5, SD 4.6, t_{133}=12.74, *P*<.001).

Participants in the social presence group (mean 52.7, SD 13.6) scored significantly lower (t_{114}=–2.81, *P*=.005) on the Working Alliance Inventory than those in the service team group (mean 59.6, SD 13.0).

Adverse Effects

Among the 123 participants who completed the final follow-up assessment, 82 completed the questionnaire on adverse intervention effects (social presence: n=44, service team: n=24, internet as usual: n=14). Of these, 66 (80.5%) reported no negative effects during the study, while 10 (12.2%) answered that an adverse effect had affected them “somewhat negatively,” 4 (4.9%) answered that an adverse effect had affected them

"quite negatively," and 2 (2.4%) answered that an adverse effect had affected them "to a great extent." However, there was no significant difference between the 3 treatment arms (χ^2=1.33, *P*=.27).

Dropout Analysis

Participants who dropped out scored significantly higher on the CES-D scale (t_{197}=–2.21, *P*=.03), GAD-7 (t_{218}=–2.92, *P*=.004), and ASRS (t_{216}=–2.12, *P*=.04) scales; reported a greater number of risky alcohol use years (t_{211}=–2.37, *P*=.02); finished fewer modules (t_{140}=9.23, *P*<.001); and spent less time on the program (t_{145}=7.15, *P*<.001) than those who completed the final follow-up evaluation. Full dropout analysis is summarized in Multimedia Appendix 2.

Discussion

Principal Findings

In this study, participants in intervention group service team (*d*=.60; *P*=.01) reported significantly greater reductions in cannabis use than those in the internet as usual group immediately after treatment and 3 months postbaseline. A reduction in the social presence group that was significant immediately after treatment (*d*=.48; *P*=.004) was no longer significant (*d*=.40, *P*=.07) at follow-up. Additionally, there were reductions in cannabis-use days 3 months after baseline in all 3 groups. There was no significant difference between the 2 active interventions.

The intervention group service team clearly outperformed all the internet-based interventions previously studied (Cohen *d* between 0 and 0.37 at follow-up 3 months from the start of treatment [19]), in terms of reducing cannabis use in general population samples, and the effects achieved were maintained 3 months after baseline. This persistence of effects aligns with the results of 2 previously published studies: one, an evaluation of video-based self-help [16], and the other, our own previous CANreduce 1.0 [14] study, in which we compared the efficacy of internet-based self-help with and without professional chat sessions to a control group. Within-group effect sizes for reducing cannabis use frequency (social presence: d_w=.93; service team d_w=.1.18) were better than those found in our previous study (d_w=.75) [14] and similar to those found by Rooke et al (d_w=1.08) [16]. Interestingly, the intervention group in the study [16] involved extended videos with an eCoach that, in our opinion, represents, apart from the reciprocal eCoach-participant relationship, the most important component of adherence-focused guidance. Thus, it appears that adherence-focused guidance also makes a difference during internet-based self-help, in terms of reducing cannabis use in cannabis users, which may allow findings from studies on stress and depression symptom reduction [25,26] to be expanded to the web-based cannabis interventions.

Our new programs also performed better than web-based interventions for brief personalized feedback (d_w=.85) and extended personalized feedback (d_w=.89) [48]; however, the study did not use a control group. Two recent meta-analyses [49,50] on brief interventions for cannabis reduction found little to no evidence of significant reductions in use or frequency. Even though some in-person brief interventions yielded small effects, the evidence consistently favored more intense, longer interventions [51]. There is evidence that brief interventions are beneficial for mild to moderate cases [52], while our program generated good effects in more severe cases. Current literature seems to indicate that more rather than less comprehensive interventions fare better in treating cannabis-use disorders. A combined stepped-care model with a range of varied intense treatment options could reach more users than a one-size-fits-all approach, as the majority of users are not in treatment [10].

Our a priori hypothesis that the social presence of a personal eCoach would outperform an impersonal study team was not confirmed. To the contrary, participants in the impersonal study team group had a significantly higher working alliance score with significant cannabis-use days reduction at 3 months (after baseline), indicating a stronger bond with the impersonal study team. We suspect that this is because the 2 adherence-focused guidance enhancement versions differed only slightly—the only differences being the presence versus absence of introduction videos and eCoach picture—and because support-team group participants may have perceived that there was an entire team being there to help them. Nevertheless, participants with access to an eCoach exhibited significantly greater adherence (in time spent, *P*=.02) and nonsignificantly greater retention (*P*=.18) than those with a support team, besides users in the latter group performing better overall in the primary outcome. Additionally, participants in the 2 intervention groups differed in their baseline scores for severity of dependence and depression. Thus, our findings did not support the existence of a linear relationship between adherence and treatment success. We found that greater adherence led to better retention and, with this, to greater data availability, and thus, more robust results to guide future research. A number of participants dropped out immediately after the start of the program, with 27% (102/386) not finishing a single module. This may stem from a discrepancy between what participants expected and what the program actually offered. Future programs should provide more information (eg, pictures or videos) to interested participants, so that they have a better idea of what to expect.

Our implementation of a single eCoach did not seem to achieve the level of social presence that was intended, even though it led to greater engagement. There may have been different intrapersonal aspects that could have affected how the participants perceived the eCoach or the study team. As we develop this program further, we intend to increase the program's social presence and offer a variety of eCoaches to foster more personal freedom and choice. We nonetheless note that, in both intervention groups, the program was well received by users.

Among secondary outcomes, we found significant differences between the service team group and internet as usual were detected, in terms of reducing the severity of cannabis dependence (*P*=.02) and reducing cannabis-use disorder severity (*P*=.01). To our knowledge, this finding has only been reported once before, in a study [16] in which a similar program was evaluated.

We expected that the active interventions would significantly alleviate the symptoms of common comorbid mental health disorders more than internet access as usual was only partially correct, with greater reductions observed for general anxiety disorder.

In a meta-analysis [53], Kedzior and Laeber identified a positive association between both cannabis use and cannabis-use disorders, and anxiety disorders. Our findings that decreased cannabis use was accompanied by decreased anxiety symptoms were consistent with those of the meta-analysis [53]. It appears that cannabis has a bidirectional effect on anxiety [54]. On one hand, some individuals with anxiety may experience a degree of acute relief from their symptoms if they use cannabis infrequently and in low doses. On the other hand, regular and heavier use could lead to a cannabis-use disorder, thereby worsening anxiety symptoms. Interestingly, the acute effects of cannabis use, such as panic attacks, resemble the symptoms of anxiety disorders [55], which could have increased the anxiety score in our sample. This said, attributing anxiety to either an anxiety disorder or cannabis use does not make much of a difference to the person suffering from anxiety.

Measuring the success of the program in cannabis-use days may be not as specific as number of joints but corresponding results from CUDIT and the severity of cannabis dependence scale support the effectiveness of the program. Future usage quantity measures should also account for cannabidiol, as well as the potency of the tetrahydrocannabinol consumed. Measuring potency is difficult, as it requires either regulated products or toxicological testing, both of which seem unlikely to be feasible in Switzerland, Austria, or Germany for the foreseeable future. Future studies on adherence-focused guidance among cannabis users need to also start investigating long-term intervention effects (12 months or longer).

Strengths and Limitations

CANreduce 2.0 was associated with comparable rates of adherence and retention but greater effect sizes than either the chat-enhanced or self-help only versions of CANreduce 1.0 [13], while being fully automated and requiring little to no human support, which in turn decreased the intervention's complexity and costs of widespread implementation. Additionally, the program reached almost twice as many women (29.4% vs 17.4%), an older population (mean age 28.3, SD 7.9 vs 21.6 years, SD 8.4), and more severe users (daily use: 74.7% vs 31.2%) than Swiss outpatient treatment monitoring statistics [56]. These differences also were observed with the previous CANreduce 1.0 program [13], supporting our belief that such programs reach a different population of cannabis users than that reached by traditional outpatient treatment facilities.

This study has several limitations. First, technical difficulties decreased the number of participants who could be successfully followed up in the internet as usual group, which likely contributed to the overall high attrition rate (452/575, 70%), though this rate is common in these types of intervention [14,16]. To reduce any bias created by dropouts, multiple imputations were used, but large variances that decreased the chance of finding smaller effects still existed. Second, the introduction and rising popularity of low tetrahydrocannabinol (less than 1%) and high cannabidiol joints in the market may have caused some confusion in our study, as it is not clear how these were counted by participants; consequently, we had to drop our personalized standard joint measurement, in which participants chose from a range of different predefined joints to match their own consumed joints [32]. Third, all measures were self-reported and could not be validated externally, though there is evidence that the internet enables people to be more open and honest and to offer more accurate self-evaluations regarding their problems [57]. Last, the study only followed patients for a relatively short time period (3 months) meaning that we can make no claims of long-term treatment success; this said, beneficial effects were maintained for at least 3 months after the baseline was completed.

Conclusions

The internet-based self-help interventions with an impersonal service team based upon adherence-focused guidance enhancement reduced cannabis use, severity of dependence, and general anxiety symptoms. The program reached a different group of treatment seekers in the general population than that reached by typical outpatient treatment options. The program is fully automated and requires little human support, which may render such programs cost-effective additions to the general health care system.

Acknowledgments

The study was funded by the Swiss National Science Foundation (grant 10001C_192301) and internal funds from the Swiss Research Institute for Public Health and Addiction. The funding sources had no role in the writing of the manuscript or the decision to submit for publication. Content is solely the responsibility of the authors and does not necessarily represent the official views of the Swiss National Science Foundation or the Swiss Research Institute for Public Health and Addiction.

Authors' Contributions

MS had the initial idea for this study. CB prepared the first draft of the paper and final manuscript. AW, DE, LS, and MS developed the interventions for study arms 1 and 2. AW and CB programmed and implemented the CANreduce 2.0 study websites. DM helped to develop and adapt the Austrian version of the website. DE, DM, LS, MA, MW, MK, SH, and TB provided continuous feedback on the development of the interventions and the manuscript. CB performed statistical analysis. CB, MA, MK, MS, and SH thoroughly revised the first version of the manuscript. All authors approved the final version of the manuscript. CB is the guarantor.

Conflicts of Interest

DE has served as a consultant to/on the scientific advisory boards of Sanofi, Novartis, Minddistrict, Lantern, Schoen Kliniken, Ideamed and German health insurance companies (BARMER, Techniker Krankenkasse) and a number of federal chambers for psychotherapy. He is also stakeholder of the Institute for health training online (formerly GET.ON/ nowHelloBetter), which aims to implement scientific findings related to digital health interventions into routine care.

Multimedia Appendix 1
Introduction video of eCoach Lena.
[PNG File , 1033 KB - jmir_v23i4e27463_app1.png]

Multimedia Appendix 2
Complete dropout analysis.
[DOCX File , 15 KB - jmir_v23i4e27463_app2.docx]

Multimedia Appendix 3
CONSORT-eHEALTH checklist (V1.6.1).
[PDF File (Adobe PDF File), 688 KB - jmir_v23i4e27463_app3.pdf]

References

1. European drug report 2019. European Monitoring Centre for Drugs and Drug Addiction. Lisbon: Publications Office of the European Union; 2019 Jun 01. URL: https://www.emcdda.europa.eu/publications/edr/trends-developments/2019_en [accessed 2021-04-14]
2. World drug report 2019. United Nations iLibrary. 2019 Jun 01. URL: https://www.un-ilibrary.org/docserver/fulltext/9789210041744/a4dd519a-en.pdf [accessed 2021-04-14]
3. Smart R, Pacula R. Early evidence of the impact of cannabis legalization on cannabis use, cannabis use disorder, and the use of other substances: Findings from state policy evaluations. Am J Drug Alcohol Abuse 2019;45(6):644-663 [FREE Full text] [doi: 10.1080/00952990.2019.1669626] [Medline: 31603710]
4. Wagner F, Anthony J. From first drug use to drug dependence; developmental periods of risk for dependence upon marijuana, cocaine, and alcohol. Neuropsychopharmacology 2002 Apr;26(4):479-488. [doi: 10.1016/S0893-133X(01)00367-0] [Medline: 11927172]
5. Degenhardt L, Hall W, Lynskey M. The relationship between cannabis use and other substance use in the general population. Drug Alcohol Depend 2001 Nov;64(3):319-327. [doi: 10.1016/s0376-8716(01)00130-2]
6. The epidemiology of cannabis dependence. In: The Nature of Cannabis Dependence. Cambridge, UK: Cambridge University Press; 2006:58.
7. Fischer R, Clair C, Studer J, Cornuz J, Gmel G. Prevalence and factors associated with use of smokeless tobacco in young Swiss men. Eur J Public Health 2013 Jun 28;24(3):459-464. [doi: 10.1093/eurpub/ckt086]
8. Fischer B, Jeffries V, Hall W, Room R, Goldner E, Rehm J. Lower risk cannabis use guidelines for Canada x(LRCUG): a narrative review of evidence and recommendations. Can J Public Health 2011 Sep 1;102(5):324-327. [doi: 10.1007/bf03404169]
9. Lev-Ran S, Le Foll B, McKenzie K, George TP, Rehm J. Cannabis use and cannabis use disorders among individuals with mental illness. Compr Psychiatry 2013 Aug;54(6):589-598. [doi: 10.1016/j.comppsych.2012.12.021] [Medline: 23375264]
10. Teesson M, Slade T, Swift W, Mills K, Memedovic S, Mewton L, et al. Prevalence, correlates and comorbidity of DSM-IV cannabis use and cannabis use disorders in Australia. Aust N Z J Psychiatry 2012 Dec 14;46(12):1182-1192. [doi: 10.1177/0004867412460591] [Medline: 22984111]
11. Sobesky M, Gorgens K. Cannabis and adolescents: Exploring the substance misuse treatment provider experience in a climate of legalization. Int J Drug Policy 2016 Jul;33:66-74. [doi: 10.1016/j.drugpo.2016.02.008] [Medline: 26992485]
12. Fernández-Artamendi S, Fernández-Hermida JR, García-Fernández G, Secades-Villa R, García-Rodríguez O. Motivation for change and barriers to treatment among young cannabis users. Eur Addict Res 2013;19(1):29-41. [doi: 10.1159/000339582] [Medline: 22948385]
13. Schaub MP, Haug S, Wenger A, Berg O, Sullivan R, Beck T, et al. Can reduce--the effects of chat-counseling and web-based self-help, web-based self-help alone and a waiting list control program on cannabis use in problematic cannabis users: a randomized controlled trial. BMC Psychiatry 2013 Nov 14;13:305 [FREE Full text] [doi: 10.1186/1471-244X-13-305] [Medline: 24228630]
14. Schaub MP, Wenger A, Berg O, Beck T, Stark L, Buehler E, et al. A web-based self-help intervention with and without chat counseling to reduce cannabis use in problematic cannabis users: three-arm randomized controlled trial. J Med Internet Res 2015 Oct 13;17(10):e232 [FREE Full text] [doi: 10.2196/jmir.4860] [Medline: 26462848]
15. Gates P, Copeland J. Barriers to treatment seeking for cannabis dependence. In: Preedy VR, editor. Handbook of Cannabis and Related Pathologies. Cambridge, Massachusetts: Academic Press; Jan 06, 2017:1025-1029.

16. Rooke S, Copeland J, Norberg M, Hine D, McCambridge J. Effectiveness of a self-guided web-based cannabis treatment program: randomized controlled trial. J Med Internet Res 2013 Feb 15;15(2):e26 [FREE Full text] [doi: 10.2196/jmir.2256] [Medline: 23470329]
17. Eysenbach G. The law of attrition. J Med Internet Res 2005 Mar 31;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
18. Kelders SM, Kok RN, Ossebaard HC, Van Gemert-Pijnen JE. Persuasive system design does matter: a systematic review of adherence to web-based interventions. J Med Internet Res 2012 Nov 14;14(6):e152 [FREE Full text] [doi: 10.2196/jmir.2104] [Medline: 23151820]
19. Boumparis N, Loheide-Niesmann L, Blankers M, Ebert DD, Korf D, Schaub MP, et al. Short- and long-term effects of digital prevention and treatment interventions for cannabis use reduction: a systematic review and meta-analysis. Drug Alcohol Depend 2019 Jul 01;200:82-94 [FREE Full text] [doi: 10.1016/j.drugalcdep.2019.03.016] [Medline: 31112834]
20. Naar S, Safren S. Motivational Interviewing and CBT Combining Strategies for Maximum Effectiveness. New York City: Guilford Press; Jun 30, 2017:242.
21. Miller W, Rollnick S. Motivational Interviewing: Preparing People to Change Addictive Behaviour. New York City: The Guilford Press; Aug 09, 1991:348.
22. Center for Substance Abuse Treatment. Brief Interventions and Brief Therapies for Substance Abuse. Rockville: Substance Abuse and Mental Health Services Administration; 1999:260.
23. Kay-Lambkin FJ, Baker AL, Kelly B, Lewin TJ. Clinician-assisted computerised versus therapist-delivered treatment for depressive and addictive disorders: a randomised controlled trial. Med J Aust 2011 Aug 01;195(3):S44-S50. [doi: 10.5694/j.1326-5377.2011.tb03265.x] [Medline: 21806518]
24. Hides L, Samet S, Lubman D. Cognitive behaviour therapy (CBT) for the treatment of co-occurring depression and substance use: current evidence and directions for future research. Drug Alcohol Rev 2010 Sep;29(5):508-517. [doi: 10.1111/j.1465-3362.2010.00207.x] [Medline: 20887574]
25. Zarski A, Lehr D, Berking M, Riper H, Cuijpers P, Ebert DD. Adherence to internet-based mobile-supported stress management: a pooled analysis of individual participant data from three randomized controlled trials. J Med Internet Res 2016 Jun 29;18(6):e146 [FREE Full text] [doi: 10.2196/jmir.4493] [Medline: 27357528]
26. Ebert DD, Lehr D, Heber E, Riper H, Cuijpers P, Berking M. Internet- and mobile-based stress management for employees with adherence-focused guidance: efficacy and mechanism of change. Scand J Work Environ Health 2016 Sep 01;42(5):382-394 [FREE Full text] [doi: 10.5271/sjweh.3573] [Medline: 27249161]
27. Mohr DC, Cuijpers P, Lehman K. Supportive accountability: a model for providing human support to enhance adherence to eHealth interventions. J Med Internet Res 2011 Mar 10;13(1):e30 [FREE Full text] [doi: 10.2196/jmir.1602] [Medline: 21393123]
28. Marlatt G, Donovan D. Relapse Prevention: Maintenance Strategies In The Treatment Of Addictive Behaviors 2nd edition. New York City: Guilford Press; 2005:416.
29. Kanter JW, Puspitasari AJ, Santos MM, Nagy GA. Behavioural activation: history, evidence and promise. Br J Psychiatry 2012 May;200(5):361-363. [doi: 10.1192/bjp.bp.111.103390] [Medline: 22550329]
30. Jaffee WB, D'Zurilla TJ. Personality, problem solving, and adolescent substance use. Behav Ther 2009 Mar;40(1):93-101. [doi: 10.1016/j.beth.2008.03.001] [Medline: 19187820]
31. Platt JJ, Husband SD. An overview of problem-solving and social skills approaches in substance abuse treatment. Psychotherapy (Chic) 1993;30(2):276-283. [doi: 10.1037/0033-3204.30.2.276]
32. Amann M, Haug S, Wenger A, Baumgartner C, Ebert DD, Berger T, et al. The effects of social presence on adherence-focused guidance in problematic cannabis users: protocol for the CANreduce 2.0 randomized controlled trial. JMIR Res Protoc 2018 Jan 31;7(1):e30 [FREE Full text] [doi: 10.2196/resprot.9484] [Medline: 29386176]
33. Hjorthøj CR, Hjorthøj AR, Nordentoft M. Validity of timeline follow-back for self-reported use of cannabis and other illicit substances--systematic review and meta-analysis. Addict Behav 2012 Mar;37(3):225-233. [doi: 10.1016/j.addbeh.2011.11.025] [Medline: 22143002]
34. Robinson SM, Sobell LC, Sobell MB, Leo GI. Reliability of the timeline followback for cocaine, cannabis, and cigarette use. Psychol Addict Behav 2014 Mar;28(1):154-162. [doi: 10.1037/a0030992] [Medline: 23276315]
35. Annaheim B, Scotto TJ, Gmel G. Revising the Cannabis Use Disorders Identification Test (CUDIT) by means of item response theory. Int J Methods Psychiatr Res 2010 Sep 19;19(3):142-155 [FREE Full text] [doi: 10.1002/mpr.308] [Medline: 20812291]
36. Gossop M, Darke S, Griffiths P, Hando J, Powis B, Hall W, et al. The Severity of Dependence Scale (SDS): psychometric properties of the SDS in English and Australian samples of heroin, cocaine and amphetamine users. Addiction 1995 May;90(5):607-614. [doi: 10.1046/j.1360-0443.1995.9056072.x] [Medline: 7795497]
37. Kokkevi A, Hartgers C. EuropASI: European adaptation of a multidimensional assessment instrument for drug and alcohol dependence. Eur Addict Res 1995;1(4):208-210. [doi: 10.1159/000259089]
38. Cole JC, Rabin AS, Smith TL, Kaufman AS. Development and validation of a Rasch-derived CES-D short form. Psychol Assess 2004 Dec;16(4):360-372. [doi: 10.1037/1040-3590.16.4.360] [Medline: 15584795]

39. Löwe B, Decker O, Müller S, Brähler E, Schellberg D, Herzog W, et al. Validation and standardization of the Generalized Anxiety Disorder Screener (GAD-7) in the general population. Med Care 2008 Mar;46(3):266-274. [doi: 10.1097/MLR.0b013e318160d093] [Medline: 18388841]
40. Daigre Blanco C, Ramos-Quiroga J, Valero S, Bosch R, Roncero C, Gonzalvo B, et al. Adult ADHD Self-Report Scale (ASRS-v1.1) symptom checklist in patients with substance use disorders. Actas Esp Psiquiatr 2009;37(6):299-305 [FREE Full text] [Medline: 20066581]
41. Breslau N, Peterson EL, Kessler RC, Schultz LR. Short screening scale for DSM-IV posttraumatic stress disorder. Am J Psychiatry 1999 Jun;156(6):908-911. [doi: 10.1176/ajp.156.6.908] [Medline: 10360131]
42. Boß L, Lehr D, Reis D, Vis C, Riper H, Berking M, et al. Reliability and validity of assessing user satisfaction with web-based health interventions. J Med Internet Res 2016 Aug 31;18(8):e234 [FREE Full text] [doi: 10.2196/jmir.5952] [Medline: 27582341]
43. Kiluk BD, Serafini K, Frankforter T, Nich C, Carroll KM. Only connect: the working alliance in computer-based cognitive behavioral therapy. Behav Res Ther 2014 Dec;63:139-146 [FREE Full text] [doi: 10.1016/j.brat.2014.10.003] [Medline: 25461789]
44. Rozental A, Boettcher J, Andersson G, Schmidt B, Carlbring P. Negative effects of internet interventions: a qualitative content analysis of patients' experiences with treatments delivered online. Cogn Behav Ther 2015 Feb 23;44(3):223-236. [doi: 10.1080/16506073.2015.1008033] [Medline: 25705924]
45. van Buuren S, Groothuis-Oudshoorn K. MICE: multivariate imputation by chained equations in R. J Stat Soft 2011;45(3):67. [doi: 10.18637/jss.v045.i03]
46. Cohen J. Statistical Power Analysis for the Behavioral Sciences. New York: Routledge; 1988:579.
47. Juszczak E, Altman DG, Hopewell S, Schulz K. Reporting of multi-arm parallel-group randomized trials: extension of the CONSORT 2010 statement. JAMA 2019 Apr 23;321(16):1610-1620. [doi: 10.1001/jama.2019.3087] [Medline: 31012939]
48. Copeland J, Rooke S, Rodriquez D, Norberg MM, Gibson L. Comparison of brief versus extended personalised feedback in an online intervention for cannabis users: short-term findings of a randomised trial. J Subst Abuse Treat 2017 May;76:43-48. [doi: 10.1016/j.jsat.2017.01.009] [Medline: 28162849]
49. Halladay J, Scherer J, MacKillop J, Woock R, Petker T, Linton V, et al. Brief interventions for cannabis use in emerging adults: a systematic review, meta-analysis, and evidence map. Drug Alcohol Depend 2019 Nov 01;204:107565. [doi: 10.1016/j.drugalcdep.2019.107565] [Medline: 31751868]
50. Imtiaz S, Roerecke M, Kurdyak P, Samokhvalov AV, Hasan OSM, Rehm J. Brief interventions for cannabis use in healthcare settings: systematic review and meta-analyses of randomized trials. J Addict Med 2020;14(1):78-88. [doi: 10.1097/ADM.0000000000000527] [Medline: 32012140]
51. Gates P, Sabioni P, Copeland J, Le Foll B, Gowing L. Psychosocial interventions for cannabis use disorder. Cochrane Database Syst Rev 2016 May 05(5):CD005336 [FREE Full text] [doi: 10.1002/14651858.CD005336.pub4] [Medline: 27149547]
52. Parmar A, Sarkar S. Brief interventions for cannabis use disorders: a review. Addict Disord Their Treat 2017;16(2):80-93. [doi: 10.1097/ADT.0000000000000100]
53. Kedzior KK, Laeber LT. A positive association between anxiety disorders and cannabis use or cannabis use disorders in the general population- a meta-analysis of 31 studies. BMC Psychiatry 2014 May 10;14(1):136. [doi: 10.1186/1471-244x-14-136]
54. Leweke F, Koethe D. Cannabis and psychiatric disorders: it is not only addiction. Addict Biol 2008 Jun;13(2):264-275. [doi: 10.1111/j.1369-1600.2008.00106.x] [Medline: 18482435]
55. Crippa JA, Zuardi AW, Martín-Santos R, Bhattacharyya S, Atakan Z, McGuire P, et al. Cannabis and anxiety: a critical review of the evidence. Hum Psychopharmacol 2009 Oct;24(7):515-523. [doi: 10.1002/hup.1048] [Medline: 19693792]
56. Maffli E. act- info - SAMBAD Ambulante Suchthilfe. Sucht Schweiz. 2019. URL: https://www.suchtschweiz.ch/fileadmin/user_upload/DocUpload/2021/SAMBAD--D19.pdf [accessed 2021-04-14]
57. Fullwood C, Sheehan N, Nicholls W. Blog function revisited: a content analysis of MySpace blogs. Cyberpsychol Behav 2009 Dec;12(6):685-689. [doi: 10.1089/cpb.2009.0138] [Medline: 19788375]

Abbreviations

ADHD: attention deficit hyperactivity disorder
ASRS: attention deficit and hyperactivity symptoms self-report scale
CES-D: Centre of Epidemiologic Studies of Depression
CUDIT-R: Cannabis Use Disorders Identification Test Revised
DSM-IV: Diagnostic and Statistical Manual of Mental Disorders, 4th edition
GAD-7: Generalized Anxiety Disorder 7
ITT: intention-to-treat
MI: motivational interviewing
PTSD: posttraumatic stress disorder

SDS: Severity of Dependence Scale

Edited by G Eysenbach; submitted 26.01.21; peer-reviewed by A Berman, G Rutigliano, D Szinay; comments to author 11.02.21; revised version received 18.03.21; accepted 03.04.21; published 30.04.21.

Please cite as:
Baumgartner C, Schaub MP, Wenger A, Malischnig D, Augsburger M, Walter M, Berger T, Stark L, Ebert DD, Keough MT, Haug S
CANreduce 2.0 Adherence-Focused Guidance for Internet Self-Help Among Cannabis Users: Three-Arm Randomized Controlled Trial
J Med Internet Res 2021;23(4):e27463
URL: https://www.jmir.org/2021/4/e27463
doi:10.2196/27463
PMID:33929333

Original Paper

Validation of the Withings ScanWatch as a Wrist-Worn Reflective Pulse Oximeter: Prospective Interventional Clinical Study

Romain Kirszenblat[1*], MS; Paul Edouard[1*], MS

Withings, Issy-Les-Moulineaux, France

[*]all authors contributed equally

Corresponding Author:
Romain Kirszenblat, MS
Withings
2 rue Maurice Hartmann
Issy-Les-Moulineaux, 92130
France
Phone: 33 141460460
Email: romain.kirszenblat@withings.com

Abstract

Background: A decrease in the level of pulse oxygen saturation as measured by pulse oximetry (SpO_2) is an indicator of hypoxemia that may occur in various respiratory diseases, such as chronic obstructive pulmonary disease (COPD), sleep apnea syndrome, and COVID-19. Currently, no mass-market wrist-worn SpO_2 monitor meets the medical standards for pulse oximeters.

Objective: The main objective of this monocentric and prospective clinical study with single-blind analysis was to test and validate the accuracy of the reflective pulse oximeter function of the Withings ScanWatch to measure SpO_2 levels at different stages of hypoxia. The secondary objective was to confirm the safety of this device when used as intended.

Methods: To achieve these objectives, we included 14 healthy participants aged 23-39 years in the study, and we induced several stable plateaus of arterial oxygen saturation (SaO_2) ranging from 100%-70% to mimic nonhypoxic conditions and then mild, moderate, and severe hypoxic conditions. We measured the SpO_2 level with a Withings ScanWatch on each participant's wrist and the SaO_2 from blood samples with a co-oximeter, the ABL90 hemoximeter (Radiometer Medical ApS).

Results: After removal of the inconclusive measurements, we obtained 275 and 244 conclusive measurements with the two ScanWatches on the participants' right and left wrists, respectively, evenly distributed among the 3 predetermined SpO_2 groups: $SpO_2 \leq 80\%$, $80\% < SpO_2 \leq 90\%$, and $90\% < SpO_2$. We found a strong association and a high level of agreement between the measurements collected from the devices, with high Pearson correlation coefficients of r=0.944 and r=0.954 on the correlation plots, low Pearson correlation coefficients of r=0.083 (P=.17) and r=0.23 (P=.001) on Bland-Altman plots, biases of 0.98% (95% CI 0.65-1.32) and 1.56% (95% CI 1.24-1.87), and root mean square errors of 2.97% and 3.00% from the participants' right and left hands, respectively.

Conclusions: In conclusion, the Withings ScanWatch is able to measure SpO_2 levels with adequate accuracy at a clinical grade. No undesirable effects or adverse events were reported during the study.

Trial Registration: ClinicalTrials.gov NCT04380389; http://clinicaltrials.gov/ct2/show/NCT04380389

(J Med Internet Res 2021;23(4):e27503) doi:10.2196/27503

KEYWORDS

connected watch; COPD; COVID-19; neural network; pulse oxygen saturation; reflective pulse oximeter; sleep apnea syndrome; SpO2; Withings ScanWatch; wearable; respiratory; oxygen; respiratory disease; oximeter; validation; accuracy; safety

Introduction

Prevalence of Respiratory Diseases

According to the World Health Organization, respiratory diseases are medical conditions affecting the airways and other structures of the lungs. Three of these include chronic obstructive pulmonary disease (COPD), sleep apnea syndrome (SAS), [1] and COVID-19 [2]. COPD is a chronic inflammatory lung disease that causes persistent and progressive airflow limitation [3,4]. Patients with COPD often experience dyspnea,

cough, sputum production, and exacerbation, defined as a worsening of the previously cited symptoms. The global prevalence of COPD in 1990 was 10.7% of persons aged at least 30 years, with 227.3 million cases, and it rose to 11.7% in 2010, with 384 million reported cases. By 2030, COPD is predicted to be one of the most common specific causes of death worldwide [5]. COPD is generally caused by tobacco smoke and exposure to outdoor and indoor pollution [3]. COPD is not a life-threatening condition but is associated with multiple comorbidities, such as cardiovascular diseases, hypertension, diabetes mellitus, and osteoporosis, all of which increase the risk of mortality [4,6]. SAS is a sleep disorder in which pauses in breathing or periods of shallow breathing during sleep occur more often than normal [1]. Each pause can last for a few seconds to a few minutes, and this occurs many times per night [1]. SAS affects 1%-6% of adults and 2% of children [7]. It affects males approximately twice as often as females [7]. Although people can be affected at any age, SAS occurs most commonly among those aged 55-60 years [7]. The prevalence of SAS is highly underestimated, as 80% of patients are undiagnosed [8]. SAS increases risks of hypertension and cardiovascular and cerebrovascular diseases. SAS also negatively affects patients' quality of life by causing daytime sleepiness and impairing cognitive function, which can lead to accidents [8,9]. COVID-19 is a contagious coronavirus disease that was first detected in 2019, caused by SARS-CoV-2, and has led to an ongoing pandemic. As of December 23, 2020, 76 million cases and 1.7 million deaths linked to COVID-19 had been confirmed, straining health care systems throughout the world [3].

Oxygen Saturation as a Detection and Monitoring Tool

Detecting SAS, COPD, and COVID-19 earlier in people can help reduce potential damage from these diseases, and accomplishing this requires a more thorough examination of oxygen levels. Oxygen saturation, defined as the fraction of oxygen-saturated hemoglobin relative to total blood hemoglobin, is measured either through an invasive method by sampling arterial blood to analyze the arterial oxygen saturation (SaO_2) via a co-oximeter or with readings collected noninvasively using a pulse oximeter to measure the peripheral or pulse oxygen saturation (SpO_2) level [10]. A pulse oximeter measures oxygen saturation in peripheral arterial blood by illuminating the skin and measuring the light absorption of oxygenated (oxyhemoglobin) and deoxygenated (reduced hemoglobin) blood using two light wavelengths: 660 nm (red) and 940 nm (infrared). The ratio of light absorbance between oxyhemoglobin and the sum of oxyhemoglobin plus deoxyhemoglobin is calculated and compared with previously calibrated direct measurements of SaO_2 to establish an estimated measure of SpO_2 [11,12]. Physiologically, oxygen saturation levels range from 95%-100% in a healthy person, and a decrease, also referred to as oxygen desaturation, can be suggestive of a respiratory disease [12]. Indeed, SpO_2 levels drop to less than 90% in patients with COPD and COVID-19 as they experience extensive periods of hypoxemia and in patients with SAS when they endure repetitive oxygen desaturation events [12-14]. Although spirometry and polysomnography (PSG) are the gold standards for diagnosing COPD and SAS, respectively, these medical devices are intrusive and are not suitable for long-term monitoring and mass screening because of their high cost and lack of accessibility [15]. In the case of COVID-19, it has been observed that some patients experience oxygen levels below 90% without dyspnea, a condition termed "silent hypoxemia;" therefore, the detection of low SpO_2 levels is even more useful for early detection and monitoring of COVID-19 [16]. Furthermore, during a pandemic that has caused a worldwide health crisis, pulse oximeters are a low-cost and easy-to-use solution to identify problems at an early stage and monitor patients at home after hospitalization [17].

Benefits of Connected Devices

Smartwatches exhibit a high degree of satisfaction and growing popularity among the general population for health monitoring [18]. These wrist-worn devices provide a wide range of personalized health features for users, such as heart rate monitoring, sleep quality control, and oxygen saturation measurements, which can assist the prevention and long-term monitoring of diseases [19]. The Withings ScanWatch is a high-end watch with an embedded heart rate sensor that uses a reflective pulse oximeter on its caseback. The device is worn as a regular watch and is connected to a smartphone through the Withings Health Mate application. Its software, the Withings Scan Monitor, measures, transfers, records, and displays the wearer's functional SpO_2 level. The Withings ScanWatch is intended for intermittent measurements and can be used in hospital, sleep laboratory, long-term care, and home use environments. Therefore, the Withings ScanWatch is an alternative option to finger pulse oximeters and offers several advantages to users. It enables the user to avoid the discomfort of wearing a finger pulse oximeter (especially during sleep), and it is more firmly attached to the user's wrist than the finger pulse oximeter, which can slip on the finger. Therefore, the Withings ScanWatch has the potential to ensure better compliance than a finger oximeter.

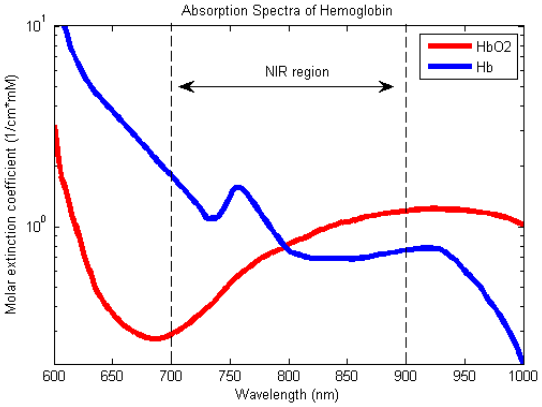
Sleep Apnea

In particular, because the ScanWatch is a wrist-worn watch that can measure SpO_2 continuously along with heart rate and, potentially, breathing rate [20,21], it could provide a less costly and time-consuming alternative to PSG. More specifically, due to the high cost of and lack of access to PSG, a subject cannot be monitored by PSG over the long term, although night-to-night variability of the apnea-hypopnea index has been found [22]. ScanWatch would enable physicians to follow patients over several weeks, at home, and in a way that does not disturb their sleep, as can occur with intrusive PSG devices. Withings ScanWatch may therefore be an alternative for screening, diagnosing, and monitoring SAS.

Biophysics of Pulse Oximetry

Pulse oximetry relies on the differential absorption by blood of red and infrared light. Figure 1 shows the absorption spectra of these two dominant forms of hemoglobin, namely oxygenated (HbO_2) and deoxygenated (Hb).

Figure 1. Absorption spectra of oxygenated and deoxygenated hemoglobin. Licensed from Adrian Curtin/CC BY-SA. Hb: deoxygenated hemoglobin; HbO_2: oxygenated hemoglobin.



Two wavelengths are chosen so that the ratios of the absorption rates of Hb and HbO_2 are respectively maximal and minimal. Thus, customary choices are red light (at 660 nm) and infrared light (at 940 nm). From the intensities of the transmitted or reflected light at these two wavelengths, one can deduce the concentrations of both forms of hemoglobin, [HbO_2] and [Hb], and the functional oxygen saturation of arterial blood (neglecting carboxyhemoglobin and methemoglobin).

In practice, a ratio of ratios (also called modulation ratio), R, is computed as follows [23]:

Alternating current (AC) refers to the amplitude of the pulsatile component of the PPG signal, and direct current (DC) refers to the value of its baseline. The ratio of AC/DC is commonly called the perfusion ratio.

An analytical relationship between the perfusion index R and the fraction of oxygenated hemoglobin SpO_2 can be obtained with a model based on the Beer-Lambert law and a simplified model for the light path taken by both wavelengths. This simple model, which neglects the multiple scattering events of light in the skin, applies to some extent to pulse oximeters operating in transmission where tissues are located between the light-emitting diode (LED) and photodiode (PD). However, the model is unable to explain how reflective pulse oximeters operate, where the LED and PD are on the same side. Therefore, an experimental calibration of R to SpO_2 during a hypoxia study is necessary.

Reflective PPG on the Wrist and Its Challenges

Measuring SpO_2 with a reflective sensor is more challenging on the wrist than on the fingertip. Due to the thickness of the wrist and the presence of bones, it is impossible to measure SpO_2 in transmission. In reflectance mode, the signal-to-noise ratio (SNR) is smaller than typically found in transmission-mode finger pulse oximeters for two reasons. First, the light emitted by the LEDs penetrates the various layers of the skin, but only a small fraction will find its way back (upward) to the PD adjacent to the LED after multiple scattering events inside the skin. Second, blood perfusion is dramatically lower on the back of the wrist than on the finger. During our calibration studies, we compared blood perfusion on the wrist and on the finger using ScanWatch sensors, and we found empirically that the SNR for the measurements taken at the finger was approximately 10 times as high as the SNR for the measurements taken at the wrist.

A second challenge is that optical measurements on the wrist are more prone to artifacts. Motion artifacts are much more frequent than on the fingertip because of the presence of tendons and bones. Light-skin coupling of finger pulse oximeters is robust due to the clamp design. In the absence of a clamp, the optical sensor can lose contact with the skin more easily, both breaking the optical coupling and letting ambient light in; this tends to unpredictably modify the DC levels.

Objectives

The main objective of this study was to clinically test and validate the accuracy of the reflective pulse oximeter function of the Withings ScanWatch to measure SpO_2 levels during mild, moderate, and severe hypoxia compared to a co-oximeter, the ABL90 hemoximeter (Radiometer Medical ApS), in accordance with the ISO 80601-2-61:2017 standard and US Food & Drug Administration (FDA) guidance. The secondary objective was to confirm the safety of the device when used as intended.

Methods

Recruitment

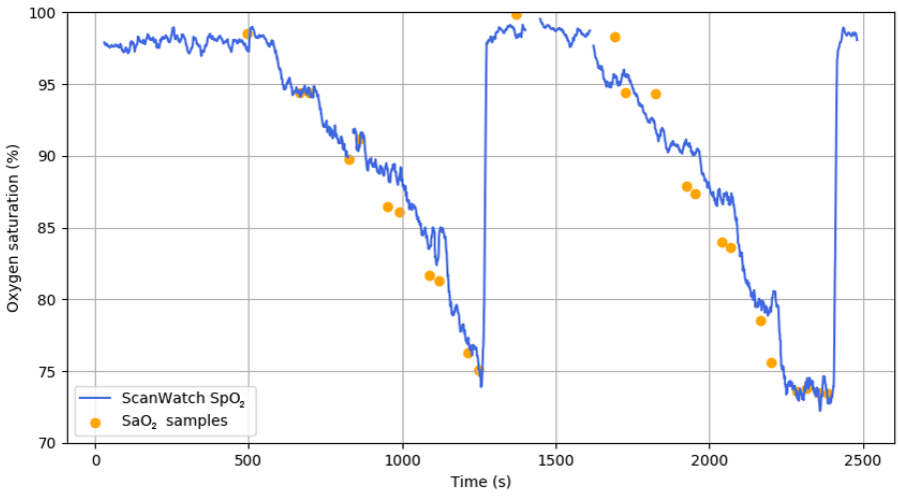
This monocentric and prospective clinical study with a double-blind analysis was conducted on 14 healthy participants in a hypoxia laboratory at the University of California San Francisco in March 2020 at an altitude of 122 m and at a room temperature of 25 °C. Inclusion criteria were being between the ages of 18 and 50 years, having a healthy status with no evidence of any medical problems, and having both wrist circumferences between 14 and 22 cm. Current smokers, women who were pregnant, lactating, or trying to get pregnant, and participants with obesity (BMI >30 kg/m^2) or who had an injury, deformity, or abnormality at the sensor sites and piercings that might cause air leaks during the test were excluded from the study. Participants with a known history of heart, lung, kidney, or liver disease; diabetes; clotting disorder; hemoglobinopathy or history

of anemia; sensitivity to local anesthesia; or fainting or vasovagal response or any other serious systemic illness were also excluded, as well as those diagnosed with asthma, sleep apnea and Raynaud disease. Exclusion criteria also included participants with a resting heart rate of over 120 beats per minute, a systolic blood pressure over 150 mm Hg, a diastolic blood pressure over 90 mm Hg, a room air SpO_2 under 94%, and a carboxyhemoglobin level over 3%.

Study Design

All selected participants in the study met the inclusion criteria. Before starting the study, two Withings ScanWatch units were placed on each participant's wrist to measure their SpO_2, while a 22-gauge catheter was placed in their left radial artery to measure their SaO_2. In addition, two reference finger pulse oximeters were placed on each participant to facilitate the identification of plateaus and any discrepancies between the hands. Each participant was then asked to lie in a semisupine position, to remain still, and to breathe a mixture of gas through a mouthpiece while a nose clip blocked their nose. After a 5-minute rest period, they were instructed to hyperventilate 2-3 times deeper and faster than normal during the runs to speed alveolar gas equilibration. The target SpO_2 at each run was chosen to be evenly balanced over the 70%-100% SpO_2 range. On the initial step of the run, two blood samples of 1-2 mL were first collected 30 seconds apart from the participant at the same oxygen saturation level as the room air. Then, approximately 10 seconds later, the inspired oxygen was abruptly changed to reduce oxygen saturation to the next SpO_2 target level. On the next steps of the run, two blood samples were collected 30 seconds apart from the participant when the reference finger pulse oximeters both measured a stabilized level of oxygen saturation, with less than 1% of difference for 35 seconds. Note that the second sampling was not performed if the oxygen saturation level was destabilized between measurements. Then, approximately 10 seconds later, inspired oxygen was abruptly changed again to reach the next SpO_2 target level. Overall, these 75-second periods of stable oxygen saturation were defined as "plateaus," and every participant was subjected to 5 plateaus (typically 92%, 87%, 82%, 77%, and 70%) before being brought back to a high oxygen saturation level (100% O_2) by breathing oxygen-enriched air for 2 minutes. After this run was repeated a second time, the collected blood samples (approximately 20-25 samples per participant) were immediately analyzed with the co-oximeter (ABL90 multiwavelength hemoximeter), then compared to the data recorded with the Withings ScanWatches (Figure 2).

Figure 2. Comparison of SpO_2 measured by the ScanWatch and hemoximeter for one subject. SpO_2: pulse oxygen saturation as measured by pulse oximetry.



Registration

The study was registered at ClinicalTrials.gov (NCT04380389).

Materials

General Description of the Withings ScanWatch

The Withings ScanWatch is an analog battery-operated watch consisting of (1) a metal case with a connected movement, (2) three hands, with two hands indicating the time and one hand indicating a cumulated activity level, (3) an adjustable silicone band to fit any user's wrist between 14 and 22 cm, and (4) a reflective pulse oximeter composed of three LEDs (red, infrared, and green), one broadband photodiode, and one infrared-cut photodiode (Figure 3). It includes all the necessary hardware to record and transmit SpO_2 measurements via Bluetooth to the Health Mate application (available on Android version 6 or later and iOS version 10 or later). For the purposes of this clinical study, instead of using the original software, which displays only a single value of SpO_2 after a 30-second measurement, we used a derived version of the Withings Scan Monitor to continuously compute, display, and record SpO_2 levels. All the other aspects of the Withings Scan Monitor, such as its algorithm and averaging window, remained identical to its original version.

Figure 3. Front view (left) and back view (right) of the Withings ScanWatch.

Design of the ScanWatch Pulse Oximeter Sensor

In a reflective design, the positions of the emitters and receptors must be carefully chosen to maximize the SNR for both wavelengths. This is achieved with optical simulations using Monte Carlo methods [24]. On average, rays detected by the photodiode follow a banana-shaped path, the depth of which depends on the wavelength [25]. On the one hand, simulations show that for a given wavelength, a higher PD-LED spacing will enable the light to pass through deeper tissues, which are more likely to be pulsatile than the superficial skin. Therefore, a longer and deeper path should increase the pulsatile modulation (or perfusion ratio) of the light signal and therefore improve the SNR.

On the other hand, increasing the LED-PD distance also comes with disadvantages. Because only a small fraction of the total light emitted by the LEDs actually reaches the photodiode, increasing this distance requires a higher light intensity, which consumes more energy. In addition, LEDs have a maximum current they can sustain without damage, so it becomes counterproductive to increase the LED-PD spacing above a certain value at which too little light would be received by the photodiode to provide a usable signal. Based on this trade-off, we found that spacings between 7 and 9 mm were optimal.

Algorithm

The Withings ScanWatch embeds a real-time self-contained algorithm implemented in C language designed to estimate SpO_2. The machine learning components of the algorithm (the neural network and linear regressions) were trained using two hypoxia calibration studies, totaling 34 subjects. This algorithm comprises three parts, as described below.

First, a signal processing part estimates AC and DC for each wavelength as well as the modulation ratio. DC levels are obtained by applying a moving average filter on the raw signals, and AC levels are obtained by filtering the raw signals with a bandpass filter centered around 1 Hz and computing the moving standard deviation on the resulting signals. Thus, the perfusions (AC/DC ratios) can be computed for each wavelength, and the modulation ratio (perfusion in the red divided by the perfusion in the infrared) can be estimated. This modulation ratio is supposed to be linearly correlated with SpO_2; however, in the case of wrist-worn oximeters, the frequent occurrence of movement and respiratory artefacts often cause that ratio to be unreliable.

To counteract the problem of the modulation ratio not being linearly correlated with SpO_2, a second SpO_2 estimator is used independently from the first part using a 1D convolutional neural network [26]. The network takes as an input an 8-second window over the three LED signals and outputs a direct SpO_2 estimation. The green LED signal is provided to the network to help the network denoise the red and infrared channels. The advantage of this method over the standard one is that the advanced filter bank built inside the network can be trained to better handle movement and respiratory artefacts present on the signals, which are particularly present in reflective PPG measurements.

Finally, the modulation ratio computed in the first part and the SpO_2 estimation calculated in the second part are merged via a linear regression to obtain a final SpO_2 estimation.

Several indicators were used to determine whether a SpO_2 measurement was conclusive or inconclusive. First, two algorithms use the PPG and accelerometer signals to assess if the watch is worn and if the user is still. These algorithms are heuristics that rely on simple filtering and thresholding, and they were calibrated on separate data sets acquired specifically for this purpose. The stillness of the user was derived from the absence of variation in the accelerometer signal, and the presence of a pulse on the PPG signal was the main factor to determine that the watch was worn.

In parallel, a second neural network with the same topology (8-second window over the three LED signals) detected and eliminated signals of poor quality. The calibration hypoxia studies were manually annotated to provide labels on which to train the neural network.

An SpO_2 measurement was considered to be inconclusive if the watch was not worn, the user was moving, or the measurement was classified as being of poor signal quality.

Statistical Analysis

A statistical analysis of the collected data was performed with the software Python 3.6.9 on a frozen database. A separate analysis was conducted for each wrist. The bias (mean error) and root mean square error (RMSE) between the SpO_2 and SaO_2 values were calculated for each range of SpO_2 values ($SpO_2 \leq 80\%$, $80\% < SpO_2 \leq 90\%$, and $90\% < SpO_2$) and for the whole range of 70%-100%. We used the Pearson correlation coefficient to determine the strength of the association between

the SpO_2 values collected from the Withings ScanWatch and the ABL90 hemoximeter. We used Bland-Altman plots to measure the agreement between the Withings ScanWatch and the ABL-90 hemoximeter.

Because a time offset exists between the Withings ScanWatch and the blood sample measured by the co-oximeter, due in part to physiological considerations (eg, distance between the arms, the wrists, and the depth of the arteries) and in part to the delay inherent to the Withings ScanWatch algorithm, we applied a plateau-matching algorithm in accordance with the recommendations of the ISO 80601-2-61:2017 standard before comparing the readings. We used a cross-correlation method to determine the delay between the measurements collected by the devices on each patient's wrists and the blood samples. Table 1 provides an overview of the mean, median, and standard deviation values of the time offsets applied on the Withings ScanWatch measurements for plateau matching.

Table 1. Time offsets applied on the Withings ScanWatch for plateau matching. A negative offset indicates that Withings ScanWatch lags behind the co-oximeter.

Offset	Mean (s)	Median (s)	SD (s)
Right hand	–4.2	–4.5	9.5
Left hand	–5.2	–6.5	10.9

Results

User Statistics

The 14 participants in our study included 8 men and 6 women aged 23-39 years with various skin tones: fair, medium, and dark skin (Table 2). Their body mass index and blood pressure values before and after the study are also reported in this table.

Table 2. Demographic data of the study participants.

Subject	Age (years)	Gender	Skin pigmentation	BMI (kg/m^2)	BP1[a] (mm Hg)	BP2[b] (mm Hg)
1	25	Male	Dark	22.2	91/73	111/72
2	26	Male	Medium	22.0	119/63	112/67
3	23	Female	Dark	20.8	125/63	112/67
4	23	Female	Medium	23.0	112/57	-
5	26	Female	Light	21.7	123/61	126/74
6	26	Male	Medium	28.1	113/67	120/65
7	25	Male	Medium	22.7	117/57	121/73
8	39	Male	Light	28.1	106/71	112/81
9	28	Female	Light	22.7	107/63	101/67
10	26	Male	Medium	22.4	108/64	110/55
11	28	Female	Medium	25.4	127/70	111/67
12	30	Male	Medium	20.8	138/91	132/88
13	28	Male	Dark	23.5	128/70	120/68
14	26	Female	Light	25.4	109/60	107/70

[a]BP1: blood pressure before the study.

[b]BP2 : blood pressure after the study.

Signal Quality

Of the 322 oxygen saturation measurements collected by the Withings ScanWatches placed on the participants' right and left wrists, 275 (85.4%) and 244 (75.8%) samples, respectively, were classified as conclusive measurements and were subsequently included for further data analysis. The remaining samples, 47/322 (14.6%) and 78/322 (24.2%) taken from the participants' right and left wrists, respectively, were classified as inconclusive measurements (poor signal quality, motion detected, or watch not worn) and were excluded.

Oxygen Saturation

The SpO_2 levels collected from the conclusive measurements ranged from 70% to 100% and were evenly distributed into 3 groups: $SpO_2 \leq 80\%$, $80\% < SpO_2 \leq 90\%$, and $90\% < SpO_2$ (Table 3). In the $SpO_2 \leq 80\%$, $80\% < SpO_2 \leq 90\%$, and $90\% < SpO_2$ groups, we found biases of 0.75%, 2.02%, and 0.14% and RMSEs of 3.29%, 3.24%, and 2.41%, respectively, when the Withings ScanWatch was placed on the participants' right wrists, and biases of 2.25%, 2.41%, and 0.31% and RMSEs of 3.74%, 3.21%, and 1.90%, respectively, when the Withings ScanWatch was placed on the participants' left wrists. Overall, we found

RMSEs of 3.00% and 2.97% and a bias of 0.98% (95% CI 0.65-1.32) and 1.56% (95% CI 1.24-1.87) from the participants' right and left wrists, respectively (Table 4).

The correlation plots show a positive strong correlation between the SpO_2 values collected from the participants' right and left hands, with high Pearson correlation coefficients of r=0.944 and r=0.954, respectively (Figure 4). The Bland-Altman plots show a high level of agreement, with Pearson correlation coefficients of r=0.083 (P=.17) and r=0.23 (P=.001) for the devices on the right and left wrists, respectively, and 95% lower and upper limits of agreement of –4.66% to 6.62% and –3.46% to 6.58%, respectively (Figure 5).

Table 3. Distribution among the SpO_2 groups of the conclusive measurements collected by the Withings ScanWatch. The corresponding distribution of SaO_2 values given by the ABL90 hemoximeter are also reported.

Group	Values, n (%)		
	Withings ScanWatch (SpO_2[a])		ABL90 hemoximeter (SaO_2[b])
	Right wrist (n=275)	Left wrist (n=244)	Blood samples (n=322)
$SpO_2 \leq 80\%$	87 (31.6)	72 (29.5)	103 (32.0)
$80\% < SpO_2 \leq 90\%$	95 (34.5)	78 (32.0)	109 (33.9)
$90\% < SpO_2$	93 (33.8)	94 (38.5)	110 (34.2)

[a]SpO_2: pulse oxygen saturation as measured by pulse oximetry.

[b]SaO_2: arterial oxygen saturation.

Table 4. Bias and RMSE found from the Withings ScanWatches placed on the participants' right and left wrists.

Group	Values (%)			
	Right wrist		Left wrist	
	Bias	RMSE[a]	Bias	RMSE
$SpO_2 \leq 80\%$	0.75	3.29	2.25	3.75
$80\% < SpO_2 \leq 90\%$	2.02	3.24	2.41	3.21
$90\% < SpO_2$	0.14	2.41	0.31	1.90
Total	0.98	3.00	1.56	2.97

[a]RSME: root mean square error.

Figure 4. Correlation plots for the Withings ScanWatches versus the ABL90 hemoximeter from the participants' right (A) and left (B) hands.

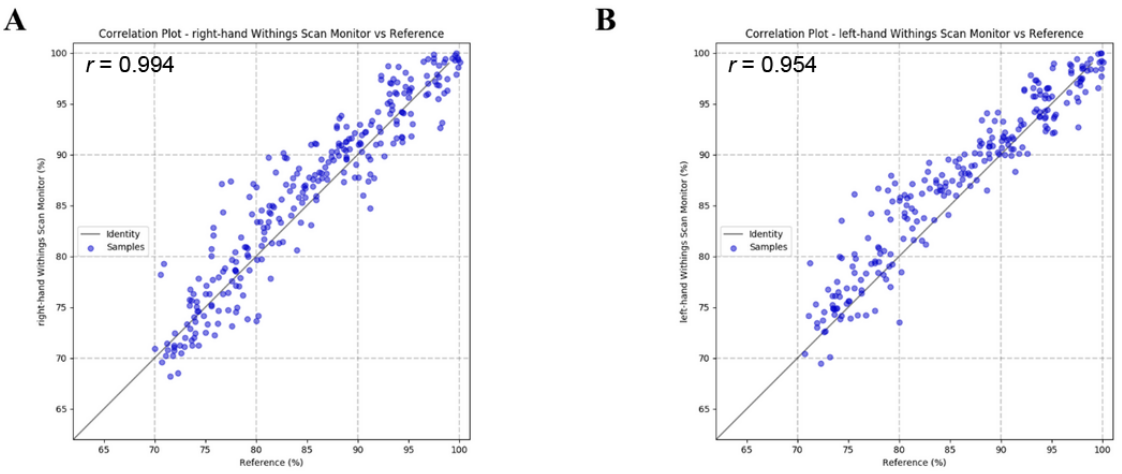
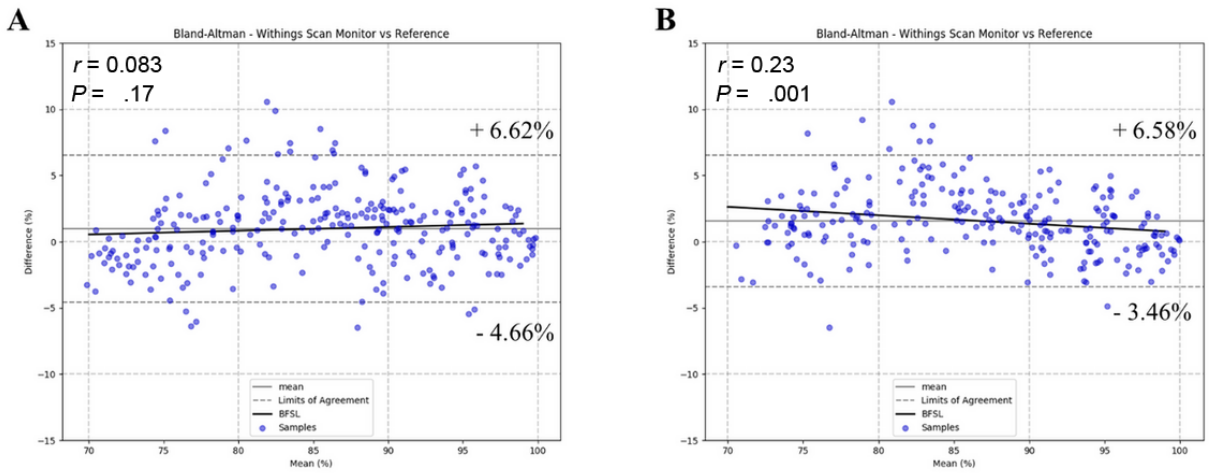


Figure 5. Bland-Altman plots for the Withings ScanWatches and the ABL90 hemoximeter from the participants' right (A) and left (B) hands. BFSL: best fit straight line.



Safety

No undesirable effects or adverse events were reported during the study.

Discussion

Principal Results

Since 2014, the popularity of smartwatches has grown considerably, particularly in the health care and biomedical industries [18]. With their integrated biosensors, these wrist-worn devices have the potential to provide valuable health care data to users, thus opening new opportunities for clinical applications. Indeed, instead of going to medical facilities, which is time-consuming, costly, and requires highly trained personnel, users can now monitor their physiological conditions themselves and report any abnormalities to physicians [18,19]. The Withings Scan Monitor is the software embedded in the Withings ScanWatch, a smartwatch that displays a reflective pulse oximeter function to measure, transfer, record, and display SpO_2 levels. Here, we tested this functionality according to the ISO 80601-2-61:2017 standard and FDA guidelines to validate its accuracy in measuring SpO_2 levels at a clinical-grade level [27,28]. To this end, we compared the performance of the Withings ScanWatch with that of the ABL90 hemoximeter, a co-oximeter routinely used in the medical field for measuring SaO_2 levels in blood samples, to measure the SpO_2 levels of 14 participants who underwent plateaus of oxygen desaturation in the laboratory.

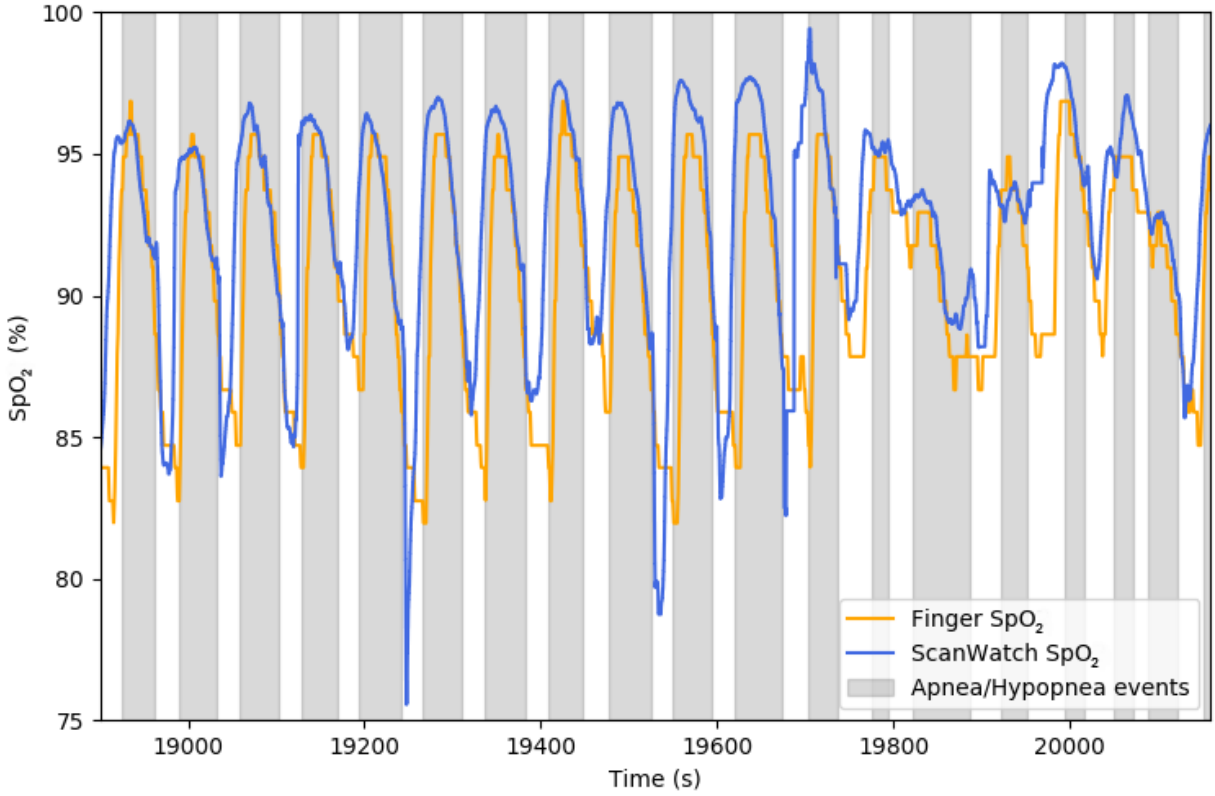
In this study, we observed an inequality in the distribution of the 325 measurements collected, with a higher rate of inconclusive measurements taken from the participants' left wrists (78/322, 24.2%) than from their right wrists (47/322, 14.6%). This imbalance was mainly caused by unwanted movements when taking blood from the participants' left arms. We also observed that the interactions between the catheter and the device on the left wrist interfered with the collection of the conclusive measurements when SpO_2 levels were below 90% (Table 3). Indeed, the distribution of the samples was slightly unbalanced, with 38% of measurements collected in the 90%<SpO_2 group. However, this difference was not significant. Thus, the ratio between the measurements distributed among the three SpO_2 groups is acceptable. Overall, we obtained sufficient conclusive measurements to meet one of the requirements set by the ISO 80601-2-61:2017 standard and FDA guidelines, which requires at least 200 paired samples from 10 participants to validate the reflective pulse oximeter function of a device.

Next, we examined the strength of the association and the agreement between the conclusive measurements taken from the reflective pulse oximeter function of the Withings ScanWatch and the ABL90 hemoximeter. We found high Pearson correlation coefficients of r=0.944 and r=0.954 on the correlation plots and r=0.083 (P=.17) and r=0.23 (P=.001) on the Bland-Altman plots, as well as RMSEs of 2.97% and 3.00% on the participants' right and left hands, respectively (Figure 4 and Figure 5). Taken together, these results indicate a strong association and a high level of agreement between the measurements collected from the devices. Therefore, we have demonstrated that the reflective pulse oximeter function of the Withings ScanWatch is adequate to measure the SpO_2 levels at different stages of hypoxia.

Detection of Sleep Apneas

In addition to the accuracy of 3% found in the hypoxia study, the ScanWatch SpO_2 algorithm possesses adequate resolution and dynamics to identify apnea and hypopnea events when worn during sleep (Figure 6). Indeed, when used in a continuous monitoring mode, the ScanWatch can detect short SpO_2 variations that occur in apneic patients, paving the way for automatic sleep apnea detection for patients at home without using intrusive and costly polysomnography setups.

Figure 6. SpO_2 measured by ScanWatch and a finger pulse oximeter during apnea/hypopnea events.



Limitations

The clinical study was conducted in a controlled environment with a well-established protocol and methodology to collect SpO_2 measurements during stable plateaus of SpO_2 in healthy participants aged 23-39 years. Its design may limit the generalizability of the results to real-world situations. Indeed, the ability of the Withings ScanWatch reflective pulse oximeter to dynamically monitor the evolution of SpO_2 in a subject is unknown in the real world because participants were exposed to stable plateaus of SpO_2 between 70% and 100% in this study. Finally, given the risks induced by hypoxia on older subjects or patients with respiratory conditions, we were unable to test the ability of the Withings ScanWatch to measure SpO_2 in these populations. In future work, the accuracy of the Withings ScanWatch reflective pulse oximeter should therefore be tested in real-life conditions (including in the home, at a hospital ward, and during rehabilitation), on a specific population such as patients with COPD or obstructive sleep apnea, or to diagnose and monitor patients with respiratory diseases.

Conclusions

FDA guidance and the ISO 80601-2-61:2017 standard require RMSEs below 3.5% and 4% for reflectance pulse oximeter approval, respectively [27,28]. These criteria were recently fulfilled by a wrist-sensor pulse oximeter, the Oxitone 1000, in a study in which its precision and accuracy were tested and an RMSE of 3% was reported [29]. Here, out of the conclusive measurements collected from the 14 participants, we have shown that the Withings ScanWatch exhibited acceptable RMSE levels for SpO_2 that were below the thresholds defined by these authorities. According to our data, we have thereby demonstrated that the Withings ScanWatch fulfills the requirements set by both the FDA guidelines and the ISO 80601-2-61 standard. The reflective pulse oximeter function of the Withings ScanWatch is thus validated and is accurate in measuring SpO_2 levels at a clinical-grade level. No undesirable effects or adverse events were reported during the study.

Acknowledgments

This work was supported by Withings (Issy-les-Moulineaux, 92130, France). We thank Guillaume Ha, PhD, for his medical writing assistance.

Conflicts of Interest

Both authors are employees of Withings, which manufactures ScanWatch, the connected watch studied in this article.

References

1. Chronic respiratory diseases. World Health Organization. URL: https://www.who.int/health-topics/chronic-respiratory-diseases [accessed 2020-10-02]
2. Bent B, Dunn JP. Wearables in the SARS-CoV-2 Pandemic: What Are They Good for? JMIR mHealth uHealth 2020 Dec 22;8(12):e25137 [FREE Full text] [doi: 10.2196/25137] [Medline: 33315580]
3. WHO Coronavirus Disease (COVID-19) Dashboard. World Health Organization. URL: https://covid19.who.int/ [accessed 2021-01-15]
4. Ntritsos G, Franek J, Belbasis L, Christou MA, Markozannes G, Altman P, et al. Gender-specific estimates of COPD prevalence: a systematic review and meta-analysis. COPD 2018 May;13:1507-1514. [doi: 10.2147/copd.s146390]
5. Barnes PJ, Celli BR. Systemic manifestations and comorbidities of COPD. Eur Respir J 2009 May 30;33(5):1165-1185 [FREE Full text] [doi: 10.1183/09031936.00128008] [Medline: 19407051]
6. Adeloye D, Chua S, Lee C, Basquill C, Papana A, Theodoratou E, Global Health Epidemiology Reference Group (GHERG). Global and regional estimates of COPD prevalence: systematic review and meta-analysis. J Glob Health 2015 Dec;5(2):020415 [FREE Full text] [doi: 10.7189/jogh.05-020415] [Medline: 26755942]
7. Global Alliance against Chronic Respiratory Diseases. Global surveillance, prevention and control of chronic respiratory diseases: a comprehensive approach. World Health Organization. URL: https://www.who.int/gard/publications/GARD_Manual/en/ [accessed 2021-04-19]
8. American Academy of Sleep Medicine (AASM). URL: https://aasm.org/ [accessed 2021-04-19]
9. Young T, Peppard PE, Gottlieb DJ. Epidemiology of obstructive sleep apnea: a population health perspective. Am J Respir Crit Care Med 2002 May 01;165(9):1217-1239. [doi: 10.1164/rccm.2109080] [Medline: 11991871]
10. Jubran A. Pulse oximetry. Crit Care 2015 Jul 16;19:272 [FREE Full text] [doi: 10.1186/s13054-015-0984-8] [Medline: 26179876]
11. Wukitsch MW, Petterson MT, Tobler DR, Pologe JA. Pulse oximetry: analysis of theory, technology, and practice. J Clin Monit 1988 Oct;4(4):290-301. [doi: 10.1007/BF01617328] [Medline: 3057122]
12. Pretto JJ, Roebuck T, Beckert L, Hamilton G. Clinical use of pulse oximetry: official guidelines from the Thoracic Society of Australia and New Zealand. Respirology 2014 Jan;19(1):38-46. [doi: 10.1111/resp.12204] [Medline: 24251722]
13. Hang L, Wang H, Chen J, Hsu J, Lin H, Chung W, et al. Validation of overnight oximetry to diagnose patients with moderate to severe obstructive sleep apnea. BMC Pulm Med 2015 Mar 20;15:24 [FREE Full text] [doi: 10.1186/s12890-015-0017-z] [Medline: 25880649]
14. Shah SA, Velardo C, Farmer A, Tarassenko L. Exacerbations in chronic obstructive pulmonary disease: identification and prediction using a digital health system. J Med Internet Res 2017 Mar 07;19(3):e69 [FREE Full text] [doi: 10.2196/jmir.7207] [Medline: 28270380]
15. Behar JA, Palmius N, Li Q, Garbuio S, Rizzatti FPG, Bittencourt L, et al. Feasibility of single channel oximetry for mass screening of obstructive sleep apnea. EClinicalMedicine 2019;11:81-88 [FREE Full text] [doi: 10.1016/j.eclinm.2019.05.015] [Medline: 31317133]
16. Tobin MJ, Laghi F, Jubran A. Why COVID-19 silent hypoxemia is baffling to physicians. Am J Respir Crit Care Med 2020 Aug 01;202(3):356-360 [FREE Full text] [doi: 10.1164/rccm.202006-2157CP] [Medline: 32539537]
17. Luks AM, Swenson ER. Pulse oximetry for monitoring patients with COVID-19 at home. Potential pitfalls and practical guidance. Ann Am Thorac Soc 2020 Sep;17(9):1040-1046 [FREE Full text] [doi: 10.1513/AnnalsATS.202005-418FR] [Medline: 32521167]
18. Reeder B, David A. Health at hand: A systematic review of smart watch uses for health and wellness. J Biomed Inform 2016 Oct;63:269-276 [FREE Full text] [doi: 10.1016/j.jbi.2016.09.001] [Medline: 27612974]
19. Guk K, Han G, Lim J, Jeong K, Kang T, Lim E, et al. Evolution of wearable devices with real-time disease monitoring for personalized healthcare. Nanomaterials (Basel) 2019 May 29;9(6) [FREE Full text] [doi: 10.3390/nano9060813] [Medline: 31146479]
20. Del Campo F, Crespo A, Cerezo-Hernández A, Gutiérrez-Tobal GC, Hornero R, Álvarez D. Oximetry use in obstructive sleep apnea. Expert Rev Respir Med 2018 Aug;12(8):665-681. [doi: 10.1080/17476348.2018.1495563] [Medline: 29972344]
21. Dehkordi P, Garde A, Molavi B, Ansermino JM, Dumont GA. Extracting instantaneous respiratory rate from multiple photoplethysmogram respiratory-induced variations. Front Physiol 2018;9:948 [FREE Full text] [doi: 10.3389/fphys.2018.00948] [Medline: 30072918]
22. Fietze I, Glos M, Zimmermann S, Penzel T. Long-term variability of the apnea-hypopnea index in a patient with mild to moderate obstructive sleep apnea. J Clin Sleep Med 2020 Feb 15;16(2):319-323 [FREE Full text] [doi: 10.5664/jcsm.8192] [Medline: 31992422]
23. Nitzan M, Romem A, Koppel R. Pulse oximetry: fundamentals and technology update. Med Devices (Auckl) 2014;7:231-239 [FREE Full text] [doi: 10.2147/MDER.S47319] [Medline: 25031547]
24. Feng S, Zeng FA, Chance B. Photon migration in the presence of a single defect: a perturbation analysis. Appl Opt 1995 Jul 01;34(19):3826-3837. [doi: 10.1364/AO.34.003826] [Medline: 21052207]
25. Niwayama M, Yamashita Y. Photon migration in tissue. In: Application of Near Infrared Spectroscopy in Biomedicine. Berlin, Germany: Springer; Feb 18, 2013:21-35.

26. Yang JB, Nguyen MN, San PP, Li XL, Krishnaswamy S. Deep convolutional neural networks on multichannel time series for human activity recognition. In: Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence. 2015 Presented at: Twenty-Fourth International Joint Conference on Artificial Intelligence (IJCAI 2015); July 25-31, 2015; Buenos Aires, Argentina URL: https://www.ijcai.org/Proceedings/15/Papers/561.pdf
27. ISO 80601-2-61:2017 Medical electrical equipment — Part 2-61: Particular requirements for basic safety and essential performance of pulse oximeter equipment. ISO. 2017 Dec. URL: https://www.iso.org/standard/67963.html [accessed 2021-04-19]
28. Pulse oximeters - premarket notification submissions [510(k)s]: guidance for industry and food and drug administration staff. US Food & Drug Administration. 2013 Mar. URL: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/pulse-oximeters-premarket-notification-submissions-510ks-guidance-industry-and-food-and-drug [accessed 2021-04-19]
29. Guber A, Epstein Shochet G, Kohn S, Shitrit D. Wrist-sensor pulse oximeter enables prolonged patient monitoring in chronic lung diseases. J Med Syst 2019 Jun 14;43(7):230. [doi: 10.1007/s10916-019-1317-2] [Medline: 31201528]

Abbreviations

AC: alternating current
COPD: chronic obstructive pulmonary disease
DC: direct current
FDA: US Food & Drug Administration
Hb: deoxygenated hemoglobin
HbO_2: oxygenated hemoglobin
LED: light-emitting diode
PD: photodiode
PPG: photoplethysmography
PSG: polysomnography
RMSE: root mean square error
SaO_2: arterial oxygen saturation
SAS: sleep apnea syndrome
SNR: signal-to-noise ratio
SpO_2: pulse oxygen saturation as measured by pulse oximetry

Edited by R Kukafka, G Eysenbach; submitted 01.02.21; peer-reviewed by D Salvi; comments to author 13.03.21; revised version received 17.03.21; accepted 11.04.21; published 26.04.21.

Please cite as:
Kirszenblat R, Edouard P
Validation of the Withings ScanWatch as a Wrist-Worn Reflective Pulse Oximeter: Prospective Interventional Clinical Study
J Med Internet Res 2021;23(4):e27503
URL: https://www.jmir.org/2021/4/e27503
doi:10.2196/27503
PMID:33857011

Original Paper

Potential Correlates of Internet Gaming Disorder Among Indonesian Medical Students: Cross-sectional Study

Kristiana Siste[1], MD, PhD; Enjeline Hanafi[1], BMedSci, MD; Lee Thung Sen[1], MRES, MD; Petra Octavian Perdana Wahjoepramono[2], MD, BMedSci; Andree Kurniawan[3], MD; Ryan Yudistiro[2], MD

[1]Department of Psychiatry, Faculty of Medicine, Universitas Indonesia - Dr Cipto Mangunkusumo General Hospital, Jakarta, Indonesia

[2]Faculty of Medicine, Universitas Pelita Harapan, Siloam Hospitals, Tangerang, Indonesia

[3]Department of Internal Medicine, Faculty of Medicine, Universitas Pelita Harapan Siloam Hospital, Tangerang, Indonesia

Corresponding Author:
Petra Octavian Perdana Wahjoepramono, MD, BMedSci
Faculty of Medicine
Universitas Pelita Harapan
Siloam Hospitals
Jl Siloam No 6
Tangerang, 15811
Indonesia
Phone: 62 811971169
Email: petra.wahjoepramono@uph.edu

Related Article:

This is a corrected version. See correction statement: https://www.jmir.org/2021/4/e29790

Abstract

Background: Internet gaming disorder has been a controversial topic for nearly a decade. Although internet addiction has been studied in medical students, there is a paucity of evidence regarding internet gaming disorder. Previous studies in Indonesia explored only the prevalence rate and characteristics.

Objective: This study aimed to determine the prevalence rate of internet gaming disorder and correlations between internet gaming disorder, temperament, and psychopathology among Indonesian medical students.

Methods: A cross-sectional study was performed from August 2019 to September 2019 using total and convenience sampling at a private university and a public university, respectively. The study variables were measured using the Indonesian version of the 10-item Internet Gaming Disorder Test, the Temperament and Character Inventory, and the Symptoms Checklist 90. Chi-square and logistic regression analyses were conducted to examine the relationships between demographic factors, temperament, psychopathology, and the presence of internet gaming disorder.

Results: Among the 639 respondents, the prevalence rate of internet gaming disorder was 2.03% (n=13), with a mean age of 20.23 (SD 0.13) years and an average gaming duration of 19.0 (SD 0.96) hours/week. Up to 71.2% respondents played using their mobile phones, and respondents with internet gaming disorder reported experiencing all psychopathologies assessed, except phobic anxiety. Bivariate analysis demonstrated that internet gaming disorder was associated with gender, gaming duration, gaming community affiliation, and 9 out of 10 domains of psychopathology. In a logistic regression model, internet gaming disorder was correlated with weekly gaming hours ≥20 hours (odds ratio [OR] 4.21, 95% CI 1.08-16.38, *P*=.04).

Conclusions: These findings suggest that the prevalence of internet gaming disorder among medical students in Jakarta, Indonesia is similar to that in other populations of Asian countries. The predisposing factor for internet gaming disorder was weekly gaming duration, while other demographic, temperament, and psychopathology variables acted as probable moderators. Strategies should, therefore, be developed and integrated into medical curriculum to screen and aid individuals with these predisposing factors.

(J Med Internet Res 2021;23(4):e25468) doi:10.2196/25468

KEYWORDS

internet gaming disorder; medical students; psychopathology; temperament; risk factors

Introduction

In this globalized era, the internet has become a necessity for people. It offers many benefits, especially for gamers, by connecting them digitally with others worldwide. Unfortunately, the internet bears its own negative effects when used excessively, which may lead to internet gaming disorder. Between 1995 and 2015, the percentage of internet users escalated (from 0%-14% to 45%-99% in 29 countries [1]). For internet gaming disorder, the prevalence globally varies from 0.7% to 15.6% among adults [2] and ranged from 0.6% to 19.9% among adolescents [3]. Specifically, a study in the US noted 8.5% of surveyed teenagers fulfilled criteria of internet gaming disorder [4], while German researchers revealed a prevalence of 1.2% [5]. Among Asian adolescents, a Taiwanese study discovered an internet gaming disorder rate of 3.1% [6], and a Japanese study demonstrated a prevalence 1.8% [7].

In Indonesia, internet usage rose from 0.9% in 2000 to 17.1% in 2014. Approximately 80% of the teenagers in Indonesia are on the internet daily [8]; consequently, the rate of internet gaming disorder among Indonesian teenagers is estimated at 10.15% [9] A systematic review [10] noted the high pooled prevalence of internet addiction (30.1%) among medical students [10], which is not surprising as this is a highly stressed group who are vulnerable to psychopathologies [11]. As such, they are at risk of adopting passive coping mechanisms. However, limited studies [10] have reported the more specific form of internet addiction (ie, internet gaming disorder) among medical students.

The varying prevalence of internet gaming disorder is due to the absence of consensus in diagnosing internet gaming disorder. Past studies [12] have looked at internet addiction as a generalized construct which encompasses all digital activities, and others, such as internet gaming, as a specific entity; however, the Diagnostic and Statistical Manual of Mental Disorders Fifth edition (DSM-5) and International Classification of Diseases 11 (ICD-11) have provided guidelines to identify affected individuals. According to the DSM-5, internet gaming disorder is defined as the persistent and recurrent use of the internet to play games that causes clinically significant impairment or distress. The patient must exhibit at least 5 out of 9 diagnostic criteria—preoccupation, withdrawal, tolerance, loss of control, giving up on other activities, continuation, deception, escape, and negative consequences—within a 12-month period [13]. Meanwhile, ICD-11 describes gaming disorder as a persistent or recurrent gaming behavior, either online or offline. The ICD-11 mentions 3 symptoms of gaming disorder, namely, impaired control over gaming, gaming being prioritized over daily activities and life interests, and its continuation or escalation, even after a negative impact has arisen [14]. Both internet gaming disorder (DSM-5) and internet gaming disorder (ICD-11) cover offline and online games [13], and accumulating evidence since the inclusion of internet gaming disorder within the section 3 of DSM-5 has led to international consensus for formal categorization of the disorder within ICD-11, which includes the internet as a specific identifier of the disorder [15]. As research and understanding of the disorders across populations grow, the nosology will require updates [15].

Researchers have developed numerous instruments to screen individuals afflicted with internet gaming disorder; however, there is no agreement regarding either screening tools or which clinical guidelines to adopt while diagnosing internet gaming disorder because there are limited studies on the diagnostic pathway, terminology definitions are arguable, and the diagnostic category is rather novel [16]. There are no definite indications yet for internet gaming disorder screening, and further research is required to define the risk and protective factors in order to narrow the specific population appropriate for screening [17].

Several studies in Indonesia have already described the prevalence, characteristics, and impacts of internet gaming disorder [9,18]. However, none of these studies focused on the relationship between internet gaming disorder and temperament, and internet gaming disorder and psychopathology. The main purpose of this study, hence, was to assess the correlations between internet gaming disorder, temperament, and psychopathology among medical students in Indonesia, given the prevalence of internet gaming disorder among Indonesian young adults and thus allowing for the ratification of preventive modules in medical schools.

Methods

Design

A cross-sectional study was conducted in private and public universities in Jakarta. The capital city of Indonesia was believed to be a representative sample base due to its dense and diverse population from across the archipelago. Total sampling in the private university and convenience sampling in the public university were performed from August to September 2019. This study was approved by Mochtar Riady Institute for Nanotechnology Ethics Committee (protocol number 1906010-02). Written consent was obtained from each respondent. Respondents who were screened with internet gaming disorder or other psychopathologies were invited for further psychiatric therapy sessions.

Respondents

The target population of this study was medical students in Indonesia, represented by the medical students in Jakarta. The sample consisted of medical students from private and public universities who had fulfilled the inclusion from respective and exclusion criteria (absence of severe psychotic disorders and substance use) by means of self-report and brief semistructured interviews with the research team. Criteria were chosen to ensure that the associations that generated were specific to internet gaming disorder. The respondents were briefed about the study in a specific session; respondents were required to give informed consent in order to participate. The minimal sample size was tabulated using the cross-sectional and hypothesis confirmation studies formula, with a type I error of 1.96, type II error of 0.84, and an absolute precision of .05 [19].

Measurements

The respondents completed a questionnaire providing demographic data and gaming-related characteristics (duration, affiliation to gaming community, and genre, whereby examples were provided). The game genres, which were adapted from prior studies [20-22], were divided into massive multiplayer role-playing games, multiplayer online battle arena, first-person shooting, multiplayer battle royale, real-time strategy, turn-based strategy, simulation, puzzle, or music/sports/platform. In addition, respondents answered 3 self-rated questionnaires.

The Indonesian version of the 10-item Internet Gaming Disorder Test (IGDT-10) consists of 10 statements about online gaming disorder similar to those of the original version. A 3-point Likert scale was used for response options (never, sometimes, and often). Scoring was then dichotomized (often=1; never or sometimes=0) to mirror the dichotomous nature of DSM-5 criteria. Items 9 and 10 represented the same construct and thus only a single score was used (ie, a response of "often" on either item scored only 1 point). The cut-off for internet gaming disorder, in accordance with DSM-5 criteria, specified a minimum of 5 out of 9 criteria [23]. In this study, we adopted the definition "persistent and recurrent use of the Internet to engage in games, often with other players, leading to clinically significant impairment or distress" for internet gaming disorder, and despite the terminology, internet gaming disorder encompassed individuals exhibiting problematic gaming symptoms in both offline and online settings [24]. For succinctness, respondents meeting the IGDT-10 cut-off were classified herein as having internet gaming disorder as IGDT-10 was found to cover clinical criteria of both DSM-5 and ICD-11 [25], though it should be cautioned that the instrument is a screening tool.

The Indonesian version of the modified Temperament and Character Inventory (TCI) was used to measure temperament. The questionnaire consists of 23 questions assessing the 3 domains of temperament (novelty seeking, reward dependence, and harm avoidance), and 16 questions that assess character (self-directedness and self-transcendence). However, this study only employed the former, and the partial instrument reliability was 52%. The answers were either "yes" (2 points) or "no" (1 point), and the points were summed up according to each domain. Each domain was divided into "low," if the total score was lower than average, and "high," if the total score was higher than average [26,27].

The Indonesian version of Symptoms Checklist 90 (SCL-90) has 82.9% sensitivity and 83.0% specificity. Respondents were asked if they had experienced any of the symptoms (90 statements) during the previous 7 days (1 week) using a 5-point Likert scale (0=never and 4=always). The points were summed based on somatization, obsessive-compulsiveness, interpersonal sensitivity, depression, anxiety, hostility, phobic anxiety, paranoid ideation, psychoticism, and additional symptoms. Respondents were then divided into a *No* group if the total score was <61, and a *Yes* group if the total T-score was ≥61 [28,29].

Statistical Analysis

The data were analyzed using statistical software (Statistical Package for Social Sciences for Windows version 25, IBM Corp). The characteristics of respondents are reported descriptively. The relationships of characteristics, temperament, and psychopathology with internet gaming disorder were determined using chi-square and Fisher exact tests. Furthermore, stepwise binomial logistic regression analysis was applied to examine the multivariate relationship between internet gaming disorder, the respondents' characteristics, their temperaments, and psychopathologies (mean raw score and standard deviation are presented). Nine variables were chosen between respondents' characteristics and SCL -90, results of the TCI were not included as none were significant; the criteria set was $P<.25$ [30] and relevance was based on prior studies [31,32] in the field.

Results

Respondent Characteristics

In total, there were 639 respondents (mean age 19.9 years), of whom 59.6% were late adolescents (381/639); 64.2% were female (410/639; mean age 19.7 years), and 35.8% were male (229/639; mean age 20.2 years). Respondents started gaming at a mean age of 11.31 years (SD 0.14) and had mean weekly gaming duration of 15.34 hours (SD 0.57; male: mean 19.00 hours, SD 0.96; female: mean 13.29 hours, SD 0.69); 55.2% of the respondents (353/639) lived independently, and 85.8% played games before 8 years of age (548/639). The age cut-off was based on a previous study [33]. Twice as many males (75/229, 32.8%) played more than 20 hours a week compared to females (68/410, 16.6%). Approximately 22.3% of males (51/229) joined a gaming community, compared to 12.0% of females (49/410). Device-wise, both genders preferred smartphones for playing games (455/639, 71.2%) (Table 1).

Table 1. Respondent demographics.

Characteristic	Male (n=229), n (%)	Female (n=410), n (%)	All (N=639), n (%)
Age			
Early adolescence	0 (0.0)	1 (0.2)	1 (0.2)
Mid-adolescence	10 (4.4)	39 (9.5)	49 (7.7)
Late adolescence	130 (56.8)	251 (61.2)	381 (59.6)
Adult	89 (38.9)	119 (29.0)	208 (32.6)
Residence			
Independent	114 (49.8)	239 (58.3)	353 (55.2)
Parent's house	115 (50.2)	171 (41.7)	286 (44.8)
Gaming onset age			
≤8 years	187 (81.7)	361 (88.0)	548 (85.8)
>8 years	42 (18.3)	49 (12.0)	91 (14.2)
Time spent gaming			
≤20 hours/week	154 (67.2)	342 (83.4)	496 (77.6)
>20 hours/week	75 (32.8)	68 (16.6)	143 (22.4)
Gaming community affiliation			
Yes	51 (22.3)	49 (12.0)	100 (15.7)
No	178 (77.7)	361 (88.0)	539 (84.4)
Device			
Smartphone	122 (53.3)	333 (81.2)	455 (71.2)
PC/desktop	41 (17.9)	25 (6.1)	66 (10.3)
Laptop	55 (24.0)	38 (9.3)	93 (14.6)
Tablet	11 (4.8)	14 (3.4)	25 (3.9)
Internet gaming disorder			
Yes	9 (3.9)	4 (1.0)	13 (2.0)
No	220 (96.1)	406 (99.0)	626 (98.0)

Respondents With Internet Gaming Disorder

Of the 639 respondents, 13 respondents were identified with internet gaming disorder (2.03%). The 13 respondents on average scored 5.56 (SD 0.53) on IGDT-10, compared to a mean score of 0.37 (SD 0.76) among the rest of the respondents. There was a statistically significant difference between gender, with males being 4 times more likely than females to have internet gaming disorder (odds ratio [OR] 4.15, 95% CI 1.26-13.64; *P*=.02). In addition, respondents playing >20 hours per week were 6 times more likely to be screened with internet gaming disorder (OR 5.82, 95% CI 1.87-18.08; *P*=.002). Attachment to a gaming community also displayed a statistical significance (OR 8.08, 95% CI 2.12-30.81; *P*=.04). No significant relationship was observed for age (*P*=.74), onset of gaming (*P*>.99), type of residence (*P*=.51), and device of choice (*P*=.57) with gaming disorder (Table 2). Among respondents with internet gaming disorder, 6 respondents (46.2%) mainly played multiplayer online battle royale, 4 respondents (30.8%) primarily enjoyed multiplayer online battle arena, 2 respondents (15.4%) focused on puzzle games, and 1 respondent (7.7%) chiefly played turn-based strategy games. Among respondents without internet gaming disorder, the most popular game genre was multiplayer online battle arena (190/626, 30.4%), second-most popular was multiplayer online battle royale (118/626, 18.9%), followed by real-time strategy (72/626, 11.5%), puzzle genre (69/626, 11.0%), and simulation (33/626, 5.3%).

Table 2. Comparison of demographics between respondents with and without internet gaming disorder.

Category	Internet gaming disorder		Comparison		
	Yes, n	No, n	Chi-square (*df*)	*P* value	OR[a] (95% CI)
Total	13	626			
Age			1.257 (3)	.74	—[b]
Early adolescence	0	1			
Mid-adolescence	0	49			
Late adolescence	9	372			
Adult	4	204			
Gender			6.435 (1)	.02	4.15 (0.13-13.64)
Male	9	220			
Female	4	406			
Gaming onset age			0.014 (1)	>.99	1.10 (0.24-5.03)
≤8 years	2	89			
>8 years	11	537			
Time spent gaming			11.715 (1)	.002	5.82 (1.87-18.08)
≤20 hours/week	5	491			
>20 hours/week	8	135			
Residence			0.443 (1)	.51	1.45 (0.48-4.37)
Independent	6	347			
Parent's house	7	279			
Community			5.231 (1)	.04	8.08 (2.12-30.81)
Yes	5	95			
No	8	531			
Device			2.037 (3)	.57	—
Smartphone	7	448			
PC/desktop	2	64			
Laptop	3	90			
Tablet	1	24			

[a]OR: odds ratio.

[b]Value missing as chi-square analysis would not produce risk estimate for *df* >1.

TCI and Internet Gaming Disorder

Of the 3 TCI domains, none was found to have a significant relationship with internet gaming disorder (novelty seeking: *P*=.13; reward dependence: *P*=.35; harm avoidance: *P*=.18). Eight out of 13 respondents with internet gaming disorder exhibited high novelty seeking behavior; similarly, 67% of respondents (10/13) also had high harm avoidance, and approximately only 33% of respondents (2/13) displayed high reward dependence (Table 3).

Table 3. Associations between Temperament and Character Inventory and internet gaming disorder.

Temperament and Character Inventory	Internet gaming disorder		Comparison		
	Yes	No	Chi-square (*df*)	*P* value	OR[a] (95% CI)
Novelty seeking			2.276 (1)	.13	2.33 (0.75-7.20)
High	8	255			
Low	5	371			
Reward dependence			0.56 (1)	.35	1.78 (0.39-8.23)
High	2	58			
Low	11	568			
Harm avoidance			1.71 (1)	.18	0.43 (0.12-1.59)
High	10	555			
Low	3	71			

[a]OR: odds ratio.

Association Between Internet Gaming Disorder and Psychopathology

The Symptoms Checklist 90 (SCL-90) has 9 domains with supplementary Global Symptoms Index (GSI) and an additional domain in the Indonesian version. Overall, respondents with internet gaming disorder scored 2 to 3 times higher on all domains of SCL-90 than respondents without internet gaming disorder. Out of all the domains, only phobic anxiety (*P*>.99) and the additional domain (*P*=.20) were not statistically significant (Table 4). GSI and interpersonal sensitivity were found to be markedly significant (OR 12.87, 95% CI 3.96-41.78, *P*<.001; OR 23.18, 95% CI 5.35-100.46, *P*=.001) respectively. Two out of 13 respondents (15%) who scored highly on IGDT-10 presented with positive somatization, obsessive-compulsive disorder, depression, anxiety, hostility, and paranoid symptoms. Interpersonal sensitivity and psychoticism were reported by 3 respondents. Those with low IGDT-10 scores also reported similar psychopathologies; for instance, 6 respondents without problematic gaming also had positive somatization symptoms. Phobic anxiety symptoms were exclusively reported by 7 respondents without internet gaming disorder.

Table 4. Symptoms Checklist 90 and internet gaming disorder.

Psychopathology	Internet gaming disorder				Comparison		
	Yes		No		Chi-square (*df*)	*P* value	OR[a] (95% CI)
	n (%)	Mean (SD)	n (%)	Mean (SD)			
Global Symptoms Index		150.1 (72.3)		54.9 (54.1)	28.93 (1)	<.001	12.87 (3.96-41.78)
Yes	5 (0.8)		29 (4.5)				
No	8 (1.3)		597 (93.4)				
Somatization		14.3 (11.0)		5.9 (6.9)	21.44 (1)	.01	18.79 (3.41-103.64)
Yes	2 (0.3)		6 (0.9)				
No	11 (1.7)		620 (97.0)				
Obsessive compulsiveness		23.1 (10.0)		8.7 (7.9)	21.44 (1)	.01	18.79 (3.41-103.64)
Yes	2 (0.3)		6 (0.9)				
No	11 (1.7)		620 (97.0)				
Interpersonal sensitivity		20.3 (10.3)		7.0 (7.0)	35.77 (1)	.001	23.18 (5.35-100.46)
Yes	3 (0.5)		8 (1.3)				
No	10 (1.6)		618 (96.7)				
Depression		26.4 (12.6)		9.3 (9.6)	21.44 (1)	.01	18.79 (3.41-103.64)
Yes	2 (0.3)		6 (0.9)				
No	11 (1.7)		620 (97.0)				
Anxiety		10.5 (8.0)		3.5 (4.6)	10.78 (1)	.03	9.30 (1.86-46.60)
Yes	2 (0.3)		12 (1.9)				
No	11 (1.7)		614 (96.1)				
Hostility		8.1 (4.5)		3.1 (3.7)	13.14 (1)	.02	11.20 (2.19-57.22)
Yes	2 (0.3)		10 (1.6)				
No	11 (1.7)		616 (96.4)				
Phobic anxiety		7.5 (5.1)		3.2 (4.0)	0.15 (1)	>.99	0.98 (0.97-0.99)
Yes	0 (0)		7 (1.1)				
No	13 (2.0)		619 (96.9)				
Paranoid ideation		11.2 (6.6)		4.1 (4.5)	10.78 (1)	.03	9.30 (1.86-46.60)
Yes	2 (0.3)		12 (1.9)				
No	11 (1.7)		614 (96.1)				
Psychoticism		10.3 (11.7)		5.2 (6.1)	35.77 (1)	.001	23.18 (5.35-100.46)
Yes	3 (0.5)		8 (1.3)				
No	10 (1.6)		618 (96.7)				
Additional		11.4 (5.2)		4.9 (4.9)	2.80 (1)	.20	5.13 (0.61-43.35)
Yes	1 (0.1)		10 (1.6)				
No	12 (1.9)		616 (96.4)				

[a]OR: odds ratio.

Multivariate Analysis of Internet Gaming Disorder, Respondent Characteristics, TCI, and SCL-90

Logistic regression analysis was performed to identify the factors that influenced risk of internet gaming disorder. The results are shown in Table 5. In the first model, when controlling for gender and community affiliation, weekly gaming duration of >20 hours per week remained statistically significant (OR 3.98, 95% CI 1.20-13.18, *P*=.02). The model explained 12.1% of the variance (R^2=0.121) and had acceptable goodness-of-fit with the Hosmer-Lemeshow test *P*=.74. When controlling for temperament and psychopathology, weekly gaming duration remained a significant factor associated with internet gaming disorder (OR 4.21, 95% CI 1.08-16.38, *P*=.04). The overall

model improved with pseudo R^2=0.301 and Hosmer-Lemeshow test P=.80.

Table 5. Logistic regression analysis of demographics, temperament, and psychopathologies with internet gaming disorder (dependent variable).

Variable	Model 1[a]			Model 2[b]		
	B	OR[c] (95% CI)	*P* value	B	OR (95% CI)	*P* value
Constant	–5.17	N/A[d]	<.001	–5.18	N/A	<.001
Gender	1.07	2.93 (0.86-9.95)	.08	1.38	3.99 (0.92-17.34)	.06
Community	0.67	1.95 (0.59-6.49)	.28	0.32	1.38 (0.34-5.50)	.65
Time spent gaming	1.38	3.98 (1.20-13.18)	.02	1.44	4.21 (1.08-16.38)	.04
Novelty seeking	N/A	N/A	N/A	0.60	1.82 (0.46-7.18)	.40
Harm avoidance	N/A	N/A	N/A	–1.14	0.32 (0.07-1.41)	.13
Global Symptoms Index	N/A	N/A	N/A	1.54	4.67 (0.55-39.40)	.16
Somatization	N/A	N/A	N/A	0.91	2.48 (0.15-42.29)	.53
Obsessive compulsiveness	N/A	N/A	N/A	2.05	7.77 (0.22-276.61)	.26
Interpersonal sensitivity	N/A	N/A	N/A	0.90	2.47 (0.034-180.71)	.68
Depression	N/A	N/A	N/A	0.49	1.62 (0.099-26.74)	.73
Anxiety	N/A	N/A	N/A	–1.05	0.35 (0.003-46.97)	.68
Hostility	N/A	N/A	N/A	0.54	1.71 (0.11-25.82)	.70
Paranoia	N/A	N/A	N/A	–1.52	0.22 (0.002-22.93)	.52
Psychoticism	N/A	N/A	N/A	2.47	11.85 (0.074- >999.999)	.34
Additional	N/A	N/A	N/A	–2.78	0.06 (0.001-2.98)	.16

[a]χ_3^2=14.13, P=.003; Nagelkerke R^2=0.121 Hosmer-Lemeshow test P=.74.

[b]χ_{15}^2=35.64, P=.002; Nagelkerke R^2=0.301; Hosmer-Lemeshow test P=.80.

[c]OR: odds ratio.

[d]N/A: not applicable.

Discussion

Prevalence and Demographics of Internet Gaming Disorder Among Medical Students

In a sample of Indonesian medical students, 2.03% (13/639) were suspected to suffer from internet gaming disorder, and a weekly gaming duration >20 hours was predictive of internet gaming disorder. Gender and participation in a gaming community did not increase the odds of internet gaming disorder. Subscales of the SCL-90, with the exception of phobic anxiety, were completed by respondents with internet gaming disorder, but none was found to increase susceptibility to internet gaming disorder in this study. The prevalence of internet gaming disorder has previously been found to range between 0.27% and 57.50% [34] and that of internet addiction has been found to range from 0.8% to 26.7% [35]. Notably, prior studies [5,36-41] indicated that internet addiction and video game addiction or problematic gaming prevalence dropped within the older age groups. Another study with young respondents (mean age 13 years), reported a higher prevalence of up to 4% [42] and up to 9% in primary school students [4]. Alternatively, a systematic review [43] estimated a pooled prevalence of 20.0% and 10.1% for internet addiction and internet gaming disorder, respectively, among the general population in Southeast Asia.

In addition, we found an association between gender and internet gaming disorder, with males constituting more than twice the number of respondents with internet gaming disorder than females, which was consistent with previous findings [36,39,40,44,45]. Furthermore, males played longer and were more likely to join a gaming community than females. In prior studies in pathological gambling, several theories have been suggested, such as differences in genetics and neurobiology between genders [46,47]. A functional magnetic resonance imaging study demonstrated that males exhibited amplified connectivity in their mesocorticolimbic pathway when playing video games compared to that exhibited by females [48]. This pathway is known for its pivotal role in reward assessment, motivated behavior, and cognitive regulation through dopaminergic modulation [49,50]. On a larger scope, gender differences have already been observed in multitudes of addictions, although many studies were culturally and politically biased [51,52]. A review [53] argued that although biological processes within the brain differ between males and females in certain areas, and there are variations in genetic expression, these are, nonetheless, further influenced at the phenotypic level by sociocultural factors and individual experiences. Gendered experiences, such as boys playing with robots and females with dolls, combined with sociocultural notions of gender-based activities, in which gambling and competitive activities are

perceived to be masculine, result in male and females tending to adopt opposite coping and escape mechanisms [53].

Internet Gaming Disorder and Gaming Characteristics Among Medical Students

Consistent with the findings of previous studies [5,45], the correlation between weekly hours spent gaming and internet gaming disorder was significant (*P*=.002), with approximately 22% of affected respondents spending more than 20 hours a week engaging in such activities. This relationship was maintained even after controlling for other significant sociodemographic factors, such as gender and community participation, temperament, and psychopathology (OR 4.21, 95% CI 1.08-13.68). A previous qualitative study [54] argued that although increased gaming time is associated with internet gaming disorder, it should not be considered a criterion for addiction when no negative consequence was observed. The study highlighted the essence of context when considering gaming time. This was revealed to be pertinent as the 2 cases identified demonstrated contrasting psychological and behavioral patterns, even though they both played over 14 hours a day [54]. Additionally, gamers may not accurately account for the duration spent on activities related to games, including strategizing, discussing, and fantasizing [55], as actual time spent gaming, thus camouflaged as seemingly short gaming duration. Another study [56] indicated that the effects of increased gaming time, particularly on weekdays, were more likely to develop into depressive, psychosomatic, and musculoskeletal symptoms. A probable explanation relates to the interaction with motive. Positive excessive gameplay might add to a person's life, in contrast to excessive play as a result of negative motivations, such as escapism. The time spent in gaming reduces the availability of time during which to perform other essential tasks, such as physical socializing, schoolwork, or work; thus, gameplay duration is a pivotal determinant when put together with gameplay motives [56-58]. Alternatively, serious gaming provided complementary avenue of training and education [59] and novel opportunities to socialize with individuals, from near and far. These generated social capitals, which are greatly influenced by physical and social proximity (familiarity) [60], more often than not spurred further indulgence in game activities [61] yet translate poorly to offline social support or provision of deep affective relationships [62,63].

Our study revealed a significant association between affiliation to a gaming community and internet gaming disorder (*P*=.04). Online gaming is an inherently social activity, particularly with the advent of massively massive multiplayer role-playing games, multiplayer online battle arena, and multiplayer online battle royale [20]. The majority of respondents with internet gaming disorder in our study favored multiplayer online battle royale and multiplayer online battle arena to other genres. Multiplayer online battle royale, as a survival game combined with scavenging and combat, presents gamers with unique scenarios and dynamic and competitive gameplay in each round [20,64]. Similarly, the game genre multiplayer online battle arena was also prevalent as it motivates gamers through nonrepetitive game style, array of in-game ranks and items, and emphasis on teamwork and clan [20]. The social factors in playing multiplayer online battle arena included making new friends, chatting, and working in a team. Social elements, such as developing reputation and admiration from the gaming community, were a central driving factor in obtaining enjoyment in playing these games [65]. Gamers with internet gaming disorder may have used online games as a substitute for establishing real-world relationships, as increasing internet gaming disorder symptoms are associated with social anxiety, and social motivations are associated with gaming addiction [66-69]. According to the social compensation hypothesis, the fleeting sense of social interaction in online games and the sense of escape from the physical world facilitates greater engagement in games [66,69]. This particular relationship was not explored in this study.

Temperament, Psychopathology, and Internet Gaming Disorder

Our study did not find an association between TCI domains and internet gaming disorder. This finding conformed with the results of a recent study [33], which argued that the 3 domains of TCI were neither protective nor risk factors for internet addiction. Individuals with high novelty-seeking supposedly modulate dopaminergic and heritable tendencies toward intense excitement, unpredictable and emotional behavior, and repeated exploratory activities in response to novelty [70]. The lack of association is partly explained as individuals with high novelty seeking are readily disinterested and lack persistence in doing a particular activity. The individuals then favor shifting their activities almost impulsively, and thus they are less likely to suffer internet addiction [33]. Our study showed a small distinction between high and low scores in all TCI domains. A previous study [71] in South Korea displayed a significant total score difference between problematic internet users and problematic drug users, although the subanalysis of the novelty seeking score difference was not statistically significant. The study [71] also showed an insignificant correlation between novelty seeking and internet gaming disorder, and a significant relationship with harm avoidance and reward dependence scores in the problematic internet user group, but not in the control group. This is in contrast with the findings of a similar study which proposed that novelty seeking and harm avoidance were strong predictors of internet gaming disorder compared to those for healthy individuals [72]. The contradictory evidence suggests that it is necessary to investigate personality factors more deeply, as novelty seeking, harm avoidance, and reward dependence have been found to be determined by a person's personality profile and their impulsivity [71].

Bivariate analysis revealed a correlation between respondents with internet gaming disorder and nearly all psychopathologies in the Indonesian version of SCL-90, except for phobia and the additional domains. This pattern may be an indication of the nonspecificity of psychological distress or a variety of clinical profiles presented by internet gaming disorder respondents. However, gender (*P*=.06), gaming community (*P*=.65), temperament (novelty seeking: *P*=.40; harm avoidance: *P*=.13), and psychopathological factors (GSI: *P*=.16; somatization: *P*=.53; obsessive compulsiveness: *P*=.26; interpersonal sensitivity: *P*=.68; depression: *P*=.73; anxiety: *P*=.68; hostility: *P*=.70; paranoia: *P*=.52; psychoticism: *P*=.34; additional: *P*=.16) lost significance in multivariate analysis. Nonetheless,

incorporating temperament and psychopathology enhanced the overall model pseudo R^2 from 0.12 to 0.30, suggesting that they may be important moderators. The complex interaction between online video game overuse and associated psychopathologies has persisted and is apparent in the ambiguous evidence available. Several cross-sectional [67,73,74] and cohort studies [4] described significant effects associated with social phobia, gaming, and addiction, while other studies failed to establish any association [75-77].

Our study had limited identification of the dysfunctions experienced by respondents at a single point in time; however, the directional causality between internet gaming disorder and psychopathology is of high importance. An elaborate cohort study [4] demonstrated that individuals who engaged in chronic persistent gaming developed depression, anxiety, and social phobia after 2 years. In addition, they had significantly lower grades and social functioning compared to those who recovered from internet gaming disorder or who did not develop internet gaming disorder [4]. Based on the proposed Escape Theory [78], a number of studies have depicted gaming as an escape strategy for patients dealing with depression [68,79]. Nonetheless, it has been argued that gaming addiction should not be viewed as merely an escape or defective coping mechanism because symptoms of other disorders are also evident [4]. Internet gaming disorder might worsen psychopathological distress, while the management of internet gaming disorder could alleviate these complaints [4]. It is uncertain, however, as to whether the psychopathologies were preexisting and then reinforced by internet gaming disorder, as previous longitudinal research has failed to establish significant causality between internet gaming disorder and symptoms of depression and anxiety [77].

Strengths and Limitations

This study is the first, to the knowledge of the authors, to investigate associations between temperament, psychopathologies, and internet gaming disorder in medical students in Indonesia. Additionally, our study was able to recruit a considerable number of respondents. Our study also distinguished online gaming by platform, from computers to mobile phones, given that an expanding number of people are occupied by digital games. Data from our study suggest that incorporating selective preventive measures among medical students targeted to proactively shift and shape gaming as positive and nurturing experiences within the medical education field is required. Prior research has demonstrated that serious games can be developed for the purpose of medical education and provided moderate effects in aiding transfer of knowledge or skills, with the additional benefit of generating a motivating and recreational learning experience [80]. Some have noted the lack of standard in pedagogical approaches of these game develops, but efforts are on-going to generate a structured framework [81] and advocate the use of serious games, which are disparate from commercial games, as prevention for gaming addiction [82].

One specific limitation of our study, apart from being cross-sectional, was that the samples were taken from medical students alone. Interpretations of correlations, or absence thereof, should be made with the sample scope in mind. Although the prevalence of internet gaming disorder among medical students fell within the range observed in the general population, the characteristics of the respondents differed largely from those of previous studies. For example, in our study, females comprised more than half of the respondents, whereas internet gaming disorder is more common in males, and past studies [35,39,75,83] have largely examined a male-dominant sample, This cross-sectional study utilized a convenience sample rather than randomized respondents, which introduced selection bias. Our study enrolled medical students to focus on an understudied population with respect to internet gaming disorder and which is presumed to have better health care access yet in reality most neglect to follow up on existing symptoms [10]; concurrently, the resources available limited how widespread sample recruitment could be. Further research should employ longitudinal design, wider population, randomized sampling, comparison to clinical psychiatric diagnoses, and investigate the development of potential interventions.

Conclusions

Our study demonstrated that the point-prevalence of medical students screened with internet gaming disorder in Indonesia is within the estimated global range. The weekly duration of gaming was the strongest determinant of internet gaming disorder. Additional research should explore the diversity of motives and negative life consequences in engaged gamers to fully contextualize the effect of gaming duration and internet gaming disorder. Therefore, strategies should be developed and incorporated into medical curriculum to screen and aid students with prolonged gaming duration and psychopathologies.

Acknowledgments

This work was supported by Hibah Insinas Dikti 2019 (KS, PW, AK, and RY). This funding source had no role in the study conception, design, analysis, interpretation, or decision to submit for publication.

Authors' Contributions

KS, PW, and AK designed the study. KS, EH, PW, AK, and RY collected the data. KS, EH, and LTS performed data analysis and wrote the first draft. All authors critically reviewed the manuscript and approved the final version.

Conflicts of Interest

None declared.

References

1. McDougall J, Duncan M. Children, video games and physical activity: an exploratory study. Int J Disabil Hum Dev 2008;7(1):89-94. [doi: 10.1515/ijdhd.2008.7.1.89]
2. Cole H, Griffiths MD. Social interactions in massively multiplayer online role-playing gamers. Cyberpsychol Behav 2007 Aug;10(4):575-583. [doi: 10.1089/cpb.2007.9988] [Medline: 17711367]
3. Brilliant T, Nouchi R, Kawashima R. Does video gaming have impacts on the brain: Evidence from a systematic review. Brain Sci 2019 Sep 25;9(10):251 [FREE Full text] [doi: 10.3390/brainsci9100251] [Medline: 31557907]
4. Gentile DA, Choo H, Liau A, Sim T, Li D, Fung D, et al. Pathological video game use among youths: a two-year longitudinal study. Pediatrics 2011 Feb;127(2):e319-e329. [doi: 10.1542/peds.2010-1353] [Medline: 21242221]
5. Rehbein F, Kliem S, Baier D, Mößle T, Petry NM. Prevalence of internet gaming disorder in German adolescents: diagnostic contribution of the nine DSM-5 criteria in a state-wide representative sample. Addiction 2015 May 10;110(5):842-851. [doi: 10.1111/add.12849] [Medline: 25598040]
6. Chiu Y, Pan Y, Lin Y. Chinese adaptation of the 10-item Internet Gaming Disorder Test and prevalence estimate of internet gaming disorder among adolescents in Taiwan. J Behav Addict 2018 Sep 01;7(3):719-726 [FREE Full text] [doi: 10.1556/2006.7.2018.92] [Medline: 30264599]
7. Nakayama H, Matsuzaki T, Mihara S, Kitayuguchi T, Higuchi S. Relationship between problematic gaming and age at the onset of habitual gaming. Pediatr Int 2020 Nov 26;62(11):1275-1281. [doi: 10.1111/ped.14290] [Medline: 32379947]
8. Kurniasanti KS, Assandi P, Ismail RI, Nasrun MWS, Wiguna T. Internet addiction: a new addiction? Med J Indones 2019 May 08;28(1):82-91. [doi: 10.13181/mji.v28i1.2752]
9. Jap T, Tiatri S, Jaya ES, Suteja MS. The development of Indonesian online game addiction questionnaire. PLoS One 2013;8(4):e61098 [FREE Full text] [doi: 10.1371/journal.pone.0061098] [Medline: 23560113]
10. Zhang MWB, Lim RBC, Lee C, Ho RCM. Prevalence of internet addiction in medical students: a meta-analysis. Acad Psychiatry 2018 Feb;42(1):88-93. [doi: 10.1007/s40596-017-0794-1] [Medline: 28849574]
11. Gentile JP, Roman B. Medical student mental health services: psychiatrists treating medical students. Psychiatry (Edgmont) 2009 May;6(5):38-45 [FREE Full text] [Medline: 19724734]
12. Brand M, Laier C, Young KS. Internet addiction: coping styles, expectancies, and treatment implications. Front Psychol 2014 Nov 11;5:1256 [FREE Full text] [doi: 10.3389/fpsyg.2014.01256] [Medline: 25426088]
13. American Psychiatric Association. Section III: emerging measures and models. In: Diagnostic and Statistical Manual of Mental Disorders Fifth Edition. Washington, DC: American Psychiatric Publishing; 2013.
14. International classification of diseases for mortality and morbidity statistics 11th revision. World Health Organization. 2018. URL: https://icd.who.int/browse11/l-m/en [accessed 2020-08-09]
15. Stein DJ, Szatmari P, Gaebel W, Berk M, Vieta E, Maj M, et al. Mental, behavioral and neurodevelopmental disorders in the ICD-11: an international perspective on key changes and controversies. BMC Med 2020 Jan 27;18(1):21 [FREE Full text] [doi: 10.1186/s12916-020-1495-2] [Medline: 31983345]
16. Costa S, Kuss DJ. Current diagnostic procedures and interventions for gaming disorders: a systematic review. Front Psychol 2019 Mar 27;10:578 [FREE Full text] [doi: 10.3389/fpsyg.2019.00578] [Medline: 30971971]
17. King DL, Adair C, Saunders JB, Delfabbro PH. Clinical predictors of gaming abstinence in help-seeking adult problematic gamers. Psychiatry Res 2018 Mar;261:581-588. [doi: 10.1016/j.psychres.2018.01.008] [Medline: 29407726]
18. Kurnianingsih N, Ratnawati R, Surya Yudhantara D, Bagus Setyo Prawiro R, Permatasari M, Rachma H, et al. Association between time spent for internet gaming, grade point average and internet gaming disorder risk among medical students. Res J Life Sci 2018 Dec 01;5(3):140-148. [doi: 10.21776/ub.rjls.2018.005.03.1]
19. Charan J, Biswas T. How to calculate sample size for different study designs in medical research? Indian J Psychol Med 2013 Apr;35(2):121-126 [FREE Full text] [doi: 10.4103/0253-7176.116232] [Medline: 24049221]
20. King DL, Delfabbro PH, Perales JC, Deleuze J, Király O, Krossbakken E, et al. Maladaptive player-game relationships in problematic gaming and gaming disorder: a systematic review. Clin Psychol Rev 2019 Nov;73:101777. [doi: 10.1016/j.cpr.2019.101777] [Medline: 31707185]
21. Elliott L, Golub A, Ream G, Dunlap E. Video game genre as a predictor of problem use. Cyberpsychol Behav Soc Netw 2012 Mar;15(3):155-161 [FREE Full text] [doi: 10.1089/cyber.2011.0387] [Medline: 22242785]
22. Bilgihan A, Cobanoglu C, Nusair K, Okumus F, Bujisic M. A quantitative study exploring the difference between gaming genre preferences. Comput Game J 2013 Apr 15;2(1):19-40. [doi: 10.1007/bf03392334]
23. Király O, Sleczka P, Pontes HM, Urbán R, Griffiths MD, Demetrovics Z. Validation of the 10-item Internet Gaming Disorder Test (IGDT-10) and evaluation of the nine DSM-5 Internet Gaming Disorder criteria. Addict Behav 2017 Jan;64:253-260. [doi: 10.1016/j.addbeh.2015.11.005] [Medline: 26632194]
24. Gentile DA, Bailey K, Bavelier D, Brockmyer JF, Cash H, Coyne SM, et al. Internet gaming disorder in children and adolescents. Pediatrics 2017 Nov 01;140(Supplement 2):S81-S85. [doi: 10.1542/peds.2016-1758h]
25. King DL, Chamberlain SR, Carragher N, Billieux J, Stein D, Mueller K, et al. Screening and assessment tools for gaming disorder: a comprehensive systematic review. Clin Psychol Rev 2020 Apr;77:101831 [FREE Full text] [doi: 10.1016/j.cpr.2020.101831] [Medline: 32143109]

26. Damayanti R. Peran biopsikososial terhadap perilaku berisiko tertular HIV pada remaja SLTA di DKI. Universitas Indonesia Library. 2006. URL: http://www.lib.ui.ac.id/detail.jsp?id=20425428&lokasi=lokal [accessed 2021-04-08]
27. Cloninger CR, Svrakic DM, Przybeck TR. A psychobiological model of temperament and character. Arch Gen Psychiatry 1993 Dec 01;50(12):975-990. [doi: 10.1001/archpsyc.1993.01820240059008] [Medline: 8250684]
28. Herianto M. Penentuan T score standar normal instrument psikometrik SCL-90, dan uji coba pada pasien rawat jalan Poliklinik Jiwa Rumah Sakit Dr Cipto Mangunkusumo Jakarta. Indonesia: Universitas Indonesia; 1994.
29. Holi M. Assessment of psychiatric symptoms using the SCL-90. University of Helsinki. 2003. URL: https://tinyurl.com/6n7xam4s [accessed 2020-08-09]
30. Bursac Z, Gauss CH, Williams DK, Hosmer DW. Purposeful selection of variables in logistic regression. Source Code Biol Med 2008 Dec 16;3(1):17 [FREE Full text] [doi: 10.1186/1751-0473-3-17] [Medline: 19087314]
31. González-Bueso V, Santamaría JJ, Fernández D, Merino L, Montero E, Ribas J. Association between internet gaming disorder or pathological video-game use and comorbid psychopathology: a comprehensive review. Int J Environ Res Public Health 2018 Apr 03;15(4):668 [FREE Full text] [doi: 10.3390/ijerph15040668] [Medline: 29614059]
32. Gervasi A, La Marca L, Costanzo A, Pace U, Guglielmucci F, Schimmenti A. Personality and internet gaming disorder: a systematic review of recent literature. Curr Addict Rep 2017 Jun 22;4(3):293-307. [doi: 10.1007/s40429-017-0159-6]
33. Siste K. Development of kuesioner diagnostik adiksi Internet for adolescents: Brain functional connectivity through fMRI BOLD, study of prevalence, risk factors, and protective factors. Indonesia: Universitas Indonesia; 2019.
34. Darvesh N, Radhakrishnan A, Lachance CC, Nincic V, Sharpe JP, Ghassemi M, et al. Exploring the prevalence of gaming disorder and internet gaming disorder: a rapid scoping review. Syst Rev 2020 Apr 02;9(1):68 [FREE Full text] [doi: 10.1186/s13643-020-01329-2] [Medline: 32241295]
35. Kuss D, Griffiths M, Karila L, Billieux J. Internet addiction: a systematic review of epidemiological research for the last decade. Curr Pharm Des 2014 Jun 31;20(25):4026-4052. [doi: 10.2174/13816128113199990617] [Medline: 24001297]
36. Mentzoni RA, Brunborg GS, Molde H, Myrseth H, Skouverøe KJM, Hetland J, et al. Problematic video game use: estimated prevalence and associations with mental and physical health. Cyberpsychol Behav Soc Netw 2011 Oct;14(10):591-596. [doi: 10.1089/cyber.2010.0260] [Medline: 21342010]
37. Wittek CT, Finserås TR, Pallesen S, Mentzoni RA, Hanss D, Griffiths MD, et al. Prevalence and predictors of video game addiction: a study based on a national representative sample of gamers. Int J Ment Health Addict 2016 Sep 23;14(5):672-686 [FREE Full text] [doi: 10.1007/s11469-015-9592-8] [Medline: 27688739]
38. Petry NM, Rehbein F, Gentile DA, Lemmens JS, Rumpf H, Mößle T, et al. An international consensus for assessing internet gaming disorder using the new DSM-5 approach. Addiction 2014 Sep;109(9):1399-1406. [doi: 10.1111/add.12457] [Medline: 24456155]
39. Király O, Bőthe B, Ramos-Diaz J, Rahimi-Movaghar A, Lukavska K, Hrabec O, et al. Ten-item Internet Gaming Disorder Test (IGDT-10): measurement invariance and cross-cultural validation across seven language-based samples. Psychol Addict Behav 2019 Feb;33(1):91-103. [doi: 10.1037/adb0000433] [Medline: 30589307]
40. Mihara S, Higuchi S. Cross-sectional and longitudinal epidemiological studies of internet gaming disorder: a systematic review of the literature. Psychiatry Clin Neurosci 2017 Jul 31;71(7):425-444 [FREE Full text] [doi: 10.1111/pcn.12532] [Medline: 28436212]
41. Ferguson CJ, Coulson M, Barnett J. A meta-analysis of pathological gaming prevalence and comorbidity with mental health, academic and social problems. J Psychiatr Res 2011 Dec;45(12):1573-1578. [doi: 10.1016/j.jpsychires.2011.09.005] [Medline: 21925683]
42. Brunborg GS, Mentzoni RA, Melkevik OR, Torsheim T, Samdal O, Hetland J, et al. Gaming addiction, gaming engagement, and psychological health complaints among Norwegian adolescents. Media Psychology 2013 Jan;16(1):115-128. [doi: 10.1080/15213269.2012.756374]
43. Chia DXY, Ng CWL, Kandasami G, Seow MYL, Choo CC, Chew PKH, et al. Prevalence of internet addiction and gaming disorders in Southeast Asia: a meta-analysis. Int J Environ Res Public Health 2020 Apr 09;17(7):2582 [FREE Full text] [doi: 10.3390/ijerph17072582] [Medline: 32283803]
44. Durkee T, Kaess M, Carli V, Parzer P, Wasserman C, Floderus B, et al. Prevalence of pathological internet use among adolescents in Europe: demographic and social factors. Addiction 2012 Dec;107(12):2210-2222. [doi: 10.1111/j.1360-0443.2012.03946.x] [Medline: 22621402]
45. Lemmens JS, Valkenburg PM, Peter J. Development and validation of a game addiction scale for adolescents. Media Psychology 2009 Mar 05;12(1):77-95. [doi: 10.1080/15213260802669458]
46. Ibáñez A, Blanco C, Perez de Castro I, Fernandez-Piqueras J, Sáiz-Ruiz J. Genetics of pathological gambling. J Gambl Stud 2003;19(1):11-22. [doi: 10.1023/a:1021271029163] [Medline: 12635538]
47. Potenza MN. The neurobiology of pathological gambling. Semin Clin Neuropsychiatry 2001 Jul;6(3):217-226. [Medline: 11447573]
48. Hoeft F, Watson CL, Kesler SR, Bettinger KE, Reiss AL. Gender differences in the mesocorticolimbic system during computer game-play. J Psychiatr Res 2008 Mar;42(4):253-258. [doi: 10.1016/j.jpsychires.2007.11.010] [Medline: 18194807]

49. Alcaro A, Huber R, Panksepp J. Behavioral functions of the mesolimbic dopaminergic system: an affective neuroethological perspective. Brain Res Rev 2007 Dec;56(2):283-321 [FREE Full text] [doi: 10.1016/j.brainresrev.2007.07.014] [Medline: 17905440]
50. Adinoff B. Neurobiologic processes in drug reward and addiction. Harv Rev Psychiatry 2004;12(6):305-320 [FREE Full text] [doi: 10.1080/10673220490910844] [Medline: 15764467]
51. Campbell ND. Using Women: Gender, Drug policy, and Social Justice. Great Britain: Routledge; 2000:19-67.
52. Campbell ND, Ettorre E. Gendering Addiction: The Politics of Drug Treatment in a Neurochemical World. UK: Palgrave Macmillan; 2011.
53. Becker JB, McClellan ML, Reed BG. Sex differences, gender and addiction. J Neurosci Res 2017 Jan 02;95(1-2):136-147 [FREE Full text] [doi: 10.1002/jnr.23963] [Medline: 27870394]
54. Griffiths MD. The role of context in online gaming excess and addiction: some case study evidence. Int J Ment Health Addiction 2009 Jul 7;8(1):119-125. [doi: 10.1007/s11469-009-9229-x]
55. Kuss DJ, Griffiths MD, Pontes HM. Chaos and confusion in DSM-5 diagnosis of Internet Gaming Disorder: issues, concerns, and recommendations for clarity in the field. J Behav Addict 2017 Jun 01;6(2):103-109 [FREE Full text] [doi: 10.1556/2006.5.2016.062] [Medline: 27599673]
56. Hellström C, Nilsson KW, Leppert J, Åslund C. Effects of adolescent online gaming time and motives on depressive, musculoskeletal, and psychosomatic symptoms. Ups J Med Sci 2015 Jun 14;120(4):263-275 [FREE Full text] [doi: 10.3109/03009734.2015.1049724] [Medline: 26072677]
57. Hellström C, Nilsson KW, Leppert J, Åslund C. Influences of motives to play and time spent gaming on the negative consequences of adolescent online computer gaming. Comput Hum Behav 2012 Jul;28(4):1379-1387. [doi: 10.1016/j.chb.2012.02.023]
58. Billieux J, Van der Linden M, Achab S, Khazaal Y, Paraskevopoulos L, Zullino D, et al. Why do you play World of Warcraft? an in-depth exploration of self-reported motivations to play online and in-game behaviours in the virtual world of Azeroth. Computers in Human Behavior 2013 Jan;29(1):103-109. [doi: 10.1016/j.chb.2012.07.021]
59. Bartolome N, Zorrilla A, Zapirain B. Can game-based therapies be trusted? is game-based education effective? a systematic review of the serious games for health and education. 2011 Presented at: 16th International Conference on Computer Games; July 27-30; Louisville, KY, USA URL: https://ieeexplore.ieee.org/document/6000353 [doi: 10.1109/cgames.2011.6000353]
60. Trepte S, Reinecke L, Juechems K. The social side of gaming: How playing online computer games creates online and offline social support. Comput Hum Behav 2012 May;28(3):832-839. [doi: 10.1016/j.chb.2011.12.003]
61. Park S, Chung N. Mediating roles of self-presentation desire in online game community commitment and trust behavior of Massive Multiplayer Online Role-Playing Games. Comput Hum Behav 2011 Nov;27(6):2372-2379. [doi: 10.1016/j.chb.2011.07.016]
62. Steinkuehler CA, Williams D. Where everybody knows your (screen) name: online games as "third places". J Comp Mediated Comm 2006 Jul;11(4):885-909. [doi: 10.1111/j.1083-6101.2006.00300.x]
63. Huvila I, Holmberg K, Ek S, Widén‐Wulff G. Social capital in Second Life. Online Inf Rev 2010 Apr 20;34(2):295-316. [doi: 10.1108/14684521011037007]
64. Choi GH, Kim MJ. Gameplay of battle royale game by rules and actions of play. 2018 Presented at: 7th Global Conference on Consumer Electronics; October 9-12; Japan URL: https://ieeexplore.ieee.org/abstract/document/8574781 [doi: 10.1109/gcce.2018.8574781]
65. Yee N. Motivations for play in online games. Cyberpsychol Behav 2006 Dec;9(6):772-775. [doi: 10.1089/cpb.2006.9.772] [Medline: 17201605]
66. Lo S, Wang C, Fang W. Physical interpersonal relationships and social anxiety among online game players. Cyberpsychol Behav 2005 Feb;8(1):15-20. [doi: 10.1089/cpb.2005.8.15] [Medline: 15738689]
67. Sioni SR, Burleson MH, Bekerian DA. Internet gaming disorder: social phobia and identifying with your virtual self. Comput Hum Behav 2017 Jun;71:11-15. [doi: 10.1016/j.chb.2017.01.044]
68. Wang J, Sheng J, Wang H. The association between mobile game addiction and depression, social anxiety, and loneliness. Front Public Health 2019 Sep 6;7:247 [FREE Full text] [doi: 10.3389/fpubh.2019.00247] [Medline: 31552213]
69. Blinka L, Mikuška J. The role of social motivation and sociability of gamers in online game addiction. Cyberpsychology 2014 Jul 01;8(2):6. [doi: 10.5817/cp2014-2-6]
70. Moss AC, Dyer KR. Psychology of Addictive Behaviour. UK: Red Globe Press; Aug 20, 2010.
71. Lee M, Jung I. Comparisons of temperament and character between problematic internet users and problematic drug users in Korean adolescents. OJPsych 2012;02(03):228-234. [doi: 10.4236/ojpsych.2012.23030]
72. Lee YS, Son JH, Park JH, Kim SM, Kee BS, Han DH. The comparison of temperament and character between patients with internet gaming disorder and those with alcohol dependence. J Ment Health 2017 Jun 28;26(3):242-247. [doi: 10.1080/09638237.2016.1276530] [Medline: 28132570]
73. Cole SH, Hooley JM. Clinical and Personality Correlates of MMO Gaming. Soc Sci Comput Rev 2013 Feb 17;31(4):424-436. [doi: 10.1177/0894439312475280]
74. Wei H, Chen M, Huang P, Bai Y. The association between online gaming, social phobia, and depression: an internet survey. BMC Psychiatry 2012 Jul 28;12(1):92. [doi: 10.1186/1471-244x-12-92]

75. Schou Andreassen C, Billieux J, Griffiths MD, Kuss DJ, Demetrovics Z, Mazzoni E, et al. The relationship between addictive use of social media and video games and symptoms of psychiatric disorders: a large-scale cross-sectional study. Psychol Addict Behav 2016 Mar;30(2):252-262. [doi: 10.1037/adb0000160] [Medline: 26999354]
76. Kircaburun K, Griffiths MD, Billieux J. Psychosocial factors mediating the relationship between childhood emotional trauma and internet gaming disorder: a pilot study. Eur J Psychotraumatol 2019 Jan 14;10(1):1565031 [FREE Full text] [doi: 10.1080/20008198.2018.1565031] [Medline: 30693081]
77. Van Rooij AJ, Schoenmakers T, Vermulst A, Van den Eijnden RJJM, Van de Mheen D. Online video game addiction: identification of addicted adolescent gamers. Addiction 2011 Jan;106(1):205-212. [doi: 10.1111/j.1360-0443.2010.03104.x] [Medline: 20840209]
78. Baumeister R. Escaping the Self: Alcoholism, Spirituality, Masochism, and Other Flights From the Burden of Selfhood. New York, NY: BasicBooks; 1991.
79. Kim DJ, Kim K, Lee H, Hong J, Cho MJ, Fava M, et al. Internet game addiction, depression, and escape from negative emotions in adulthood. J Nerv Ment Dis 2017;205(7):568-573. [doi: 10.1097/Nmd.0000000000000698]
80. Gorbanev I, Agudelo-Londoño S, González RA, Cortes A, Pomares A, Delgadillo V, Muñoz. A systematic review of serious games in medical education: quality of evidence and pedagogical strategy. Med Educ Online 2018 Dec;23(1):1438718 [FREE Full text] [doi: 10.1080/10872981.2018.1438718] [Medline: 29457760]
81. Olszewski AE, Wolbrink TA. Serious gaming in medical education: a proposed structured framework for game development. Simul Healthc 2017 Aug;12(4):240-253. [doi: 10.1097/SIH.0000000000000212] [Medline: 28027076]
82. Kim JY, Bae JH. Analysis of serious games for preventing internet gaming addiction. Int J Inf Technol Manag 2018;17(1/2):62. [doi: 10.1504/ijitm.2018.10010508]
83. Billieux J, Thorens G, Khazaal Y, Zullino D, Achab S, Van der Linden M. Problematic involvement in online games: a cluster analytic approach. Comput Hum Behav 2015 Feb;43:242-250. [doi: 10.1016/j.chb.2014.10.055]

Abbreviations

DSM-5: Diagnostics and Statistical Manual for Mental Disorders, fifth edition
GSI: Global Symptoms Index
ICD-11: International Classification of Diseases, 11th revision
IGDT-10: 10-item Internet Gaming Disorder Test
OR: odds ratio
SCL-90: Symptoms Checklist 90
TCI: Temperament and Character Inventory

Edited by R Kukafka; submitted 03.11.20; peer-reviewed by B Sindelar, D McDonnell, S Steingrimsson; comments to author 13.12.20; revised version received 15.01.21; accepted 16.03.21; published 19.04.21.

Please cite as:
Siste K, Hanafi E, Sen LT, Wahjoepramono POP, Kurniawan A, Yudistiro R
Potential Correlates of Internet Gaming Disorder Among Indonesian Medical Students: Cross-sectional Study
J Med Internet Res 2021;23(4):e25468
URL: https://www.jmir.org/2021/4/e25468
doi:10.2196/25468
PMID:33871379

Original Paper

Development of a Resource Guide to Support the Engagement of Mental Health Providers and Patients With Digital Health Tools: Multimethod Study

Gillian Strudwick[1,2], RN, FAMIA, PhD; David McLay[1], MHI, PhD; Brian Lo[1,2], MHI; Hwayeon Danielle Shin[3], BScN, RN; Leanne Currie[4], RN, PhD; Nicole Thomson[1], PhD; Éric Maillet[5], RN, PhD; Vanessa Strong[6], MSc; Alanna Miller[7], BSc, MScPH; Nelson Shen[1,2], PhD; Janis Campbell[6], MEd

[1]Centre for Addiction and Mental Health, Toronto, ON, Canada
[2]University of Toronto, Toronto, ON, Canada
[3]Dalhousie University, Halifax, NS, Canada
[4]University of British Columbia, Vancouver, ON, Canada
[5]University of Sherbrooke, Longueuil, QC, Canada
[6]Memorial University, St. John's, NL, Canada
[7]McGill University, Montreal, QC, Canada

Corresponding Author:
Gillian Strudwick, RN, FAMIA, PhD
Centre for Addiction and Mental Health
1001 Queen St W
Toronto, ON, M6J 1H4
Canada
Phone: 1 416 535 8501 ext 39333
Email: gillian.strudwick@camh.ca

Abstract

Background: As mental illness continues to affect 1 in 5 individuals, and the need for support has increased during the COVID-19 pandemic, the promise of digital mental health tools remains largely unrealized due to a lack of uptake by patients and providers. Currently, most efforts on supporting the uptake of digital mental health tools remain fragmented across organizations and geography. There is a critical need to synthesize these efforts in order to provide a coordinated strategy of supporting the adoption of digital mental health tools.

Objective: The specific aim of this project is to develop a web-based resource document to support the engagement of mental health providers and patients in the use of digital mental health tools.

Methods: The web-based resource was developed using a multimethod approach. A grey literature review was conducted in 2019 to identify relevant toolkits that are available in the public domain. This was supplemented with an environmental scan where individuals with expertise in the development, acquisition, implementation, and evaluation of digital mental health tools were invited to contribute additional tools or documents not identified in the grey literature search. An engagement workshop was held with stakeholders to explore how the resource document should be developed and delivered. These findings were collectively used to develop the final iteration of the resource document.

Results: Based on a gray literature review and environmental scan with 27 experts, 25 resources were identified and included in the resource guide. These resources were developed for patients and providers by organizations from 5 countries. An engagement workshop was held with 14 stakeholders, and barriers related to cultural sensitivity, sustainability, and accessibility of the toolkit were identified. The final iteration of the resource document was developed by the research team using findings from the gray literature review, environmental scan, and engagement workshop. The contents of the 45-page resource guide are directed at mental health care providers, administrators, and patients (inclusive of families and caregivers).

Conclusions: The use of a multimethod approach led to the development of a resource guide that builds on existing evidence on digital mental health tools and was co-designed with stakeholders and end-users. The resource guide is now publicly available online for free and is being promoted through digital health and mental health websites. Future work should explore how this document can be integrated into clinical care delivery and pathways.

(J Med Internet Res 2021;23(4):e25773) doi:10.2196/25773

KEYWORDS

digital health; mental health; psychiatry; COVID-19; nursing informatics; health informatics

Introduction

Mental illness continues to be a global challenge, particularly during the COVID-19 pandemic [1]. Even prior to the pandemic, mental illness affected 1 in 5 individuals in the United States [2]. Unfortunately, with fewer than 50% of individuals receiving treatment for their mental health issues, mental illness remains a top contributor of disability in many countries including the United States [3] and Canada [4]. This increasing and unprecedented demand has spurred great interest in the use of telehealth or other digital tools during the current pandemic, highlighting new opportunities for improving access to health care through digital technologies [5].

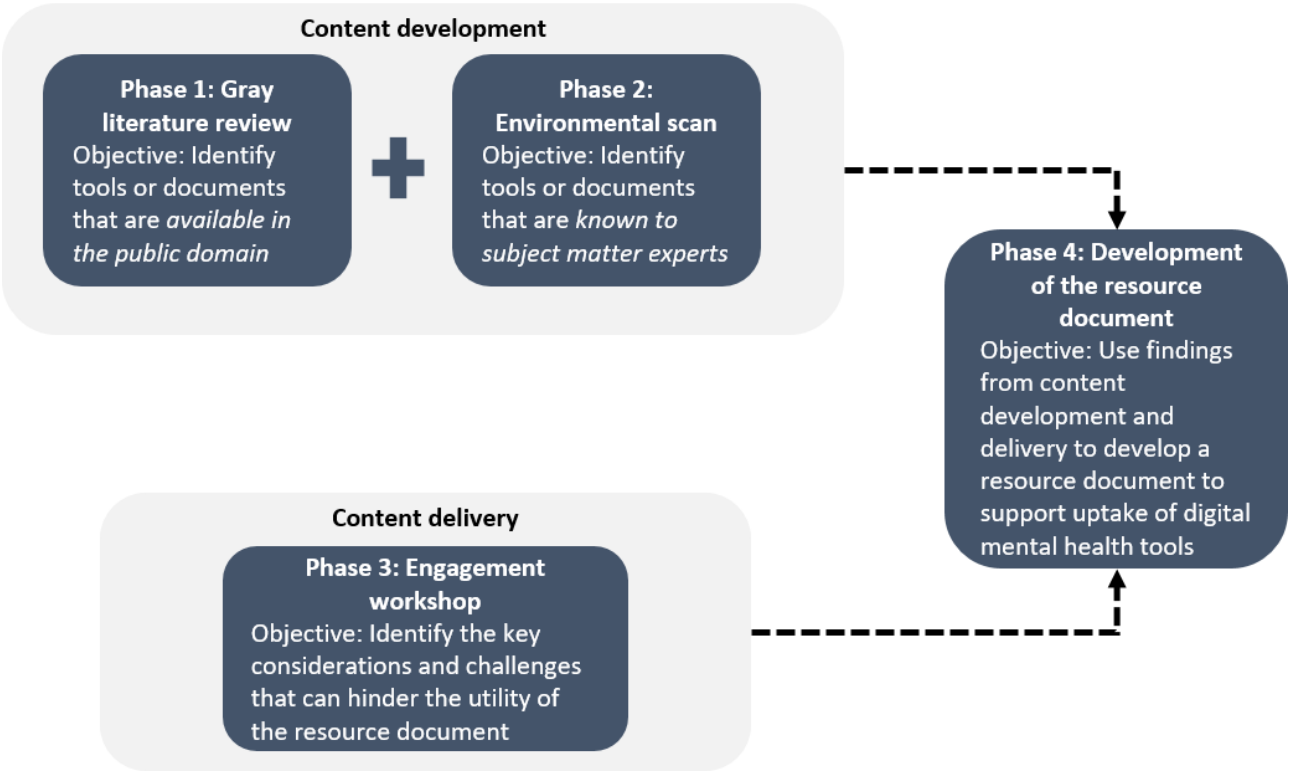
Digital mental health interventions, such as mobile apps, have been advocated by organizations, including the American Psychiatric Association (APA) [6] and the Health and Aging Department of Australia [7], as promising tools to support the current challenges in mental health service delivery. However, uptake of these tools by patients and providers remains poor [8]. In response, tools and resources for supporting the uptake of these tools into practice have recently been developed [9], yet these resources remain largely underused. Currently, most of the efforts and resources that are developed to support the adoption of digital mental health tools have been done in a piecewise effort across government organizations [10], hospitals, and mental health associations [11]. To our knowledge, there is no single source or repository where users can seek guidance on identifying relevant eHealth technologies for their needs. Given this, there is a critical need to collaborate with stakeholders and end-users in synthesizing a strong body of guidance and evidence to support digital health activities (eg, usability, user needs) and accelerate the adoption of digital tools for mental health contexts, especially during a global pandemic.

The objective of this project is to develop a comprehensive web-based resource guide to support mental health providers and patients in the selection and adoption of digital health tools through consideration of relevant factors (eg, demographics, clinical needs). The intended audiences of the guide are mental health care providers (eg, psychologists) and administrators (eg, implementation specialists) who are interested in integrating digital health tools into clinical practice, as well as people with lived experience and families or caregivers looking to use the resource to select helpful tools for their own needs. In this article, we share our approach and methodology for developing the resource document and outline how the main findings from each phase of the method informed the final development of the resource document. In addition, the implications of the resource document and challenges identified throughout the development process are discussed.

Methods

Following guidelines of the Agency for Healthcare Research and Quality [12], we used a multimethod approach (Figure 1) to develop the resource guide. The methods include a gray literature review, an environmental scan and engagement of experts in the field, and an engagement workshop with relevant stakeholders from a variety of backgrounds and interests. The findings from these 3 sources were then used to inform the development of the final version of the resource document (Multimedia Appendix 1). A multimethod approach [13] was selected to maximize the consolidation efforts of this work and deliver the findings in a meaningful and useful resources for a diverse audience. This work spanned across all provinces in Canada.

Figure 1. Overview of the multimethod approach for the development of the resource document.



Phase 1: Gray Literature Review

The objective of the gray literature review [14] was to identify relevant toolkits that are publicly available in the public domain. Following best practices on gray literature review [15], one of the authors (DM) conducted Google searches using key terms related to digital mental health tools, resources, and toolkits (Textbox 1). The first 10-20 pages of the Google Search results were reviewed, and tools or documents that met the inclusion criteria were identified and included. The inclusion criteria were as follows: available in English; technologies used or referred available in Canada; practical guidance for the use of digital mental health tools in clinical care provided; a target audience of providers, clients and caregivers, or both; and having relevance or being easily adaptable to the Canadian context (eg, health care system structure, processes).

Preference was given to Canadian sources and bilingual (French and English) resources. Relevant websites from mental health organizations (eg, Canadian Mental Health Association), medical organizations or hospitals (eg, British Medical Association), patient organizations (eg, The Mental Elf), and governmental organizations (eg, US Department of Health and Human Services) were also included. Tools of documents were excluded for one or more of the following reasons: they were more than 3 years old; they were digital mental health tools (eg, mental health apps, telemedicine portals); information was intended for policymakers, industry, or other audiences outside of providers, clients, or caregivers; they were academic or research articles; the tools had significant contextual information (eg, legal context or policy context) that rendered the information irrelevant for the Canadian context.

Blogposts or other lists (usually of apps) were also excluded due to a concern for the information being outdated. Included tools or documents were then catalogued using a Microsoft Excel spreadsheet for analysis. Relevant information related to the scope and utility was extracted from each tool or document. A content analysis [16] was used to understand the characteristics of the included tools and documents.

Textbox 1. Search strategy for gray literature review.

Gray literature search strings
• (electronic OR digital OR mobile) AND “mental health” AND (tool OR resource or e-tool OR e-resource OR toolkit OR app OR web)
• (electronic OR digital OR mobile) AND patient AND (tool OR resource OR e-tool OR e-resource OR toolkit OR app OR web)
• digital mental health tool
• digital tools to help my mental health

Phase 2: Environmental Scan

The purpose of the environmental scan was to identify relevant documents and tools that currently exist and are used in the field. To maximize the impact of the environmental scan and the number of documents and tools found, experts in Canada and the United States were identified using a snowball sampling approach through the professional networks of the project team and those who published, conducted research, or worked in the

field. Individuals who were knowledgeable across various digital health activities (eg, implementation, evaluation, design) were eligible to participate. Experts were contacted via email and telephone by the project team. Each expert was asked if they were aware of any tools or documents relevant to guide the uptake of mental health tools in the delivery of mental health care. These tools or documents were added to the list from the gray literature review (phase 1) and screened using the same inclusion and exclusion criteria. Content analysis [16] was also used to characterize the identified documents.

Phase 3: Engagement Workshop

We conducted an engagement workshop to increase the relevance and use of our research findings in practice [17]. Based on the environmental scan and literature review, a 1-day engagement workshop was held in January 2020 to gather information on the design and use of the resource guide. The engagement workshop was based on the Theory of Inventive Problem Solving (TRIZ) framework, which was developed in the 1960s to solve problems through the development of innovative and creative solutions [18]. Stemming from the principle of identifying the conceptual issue, it has become the basis for developing a conceptual solution that guides the arrival of the actual solution [18]. In contrast to traditional approaches, the TRIZ is said to guide the development of solutions by focusing on the contradictions that arise between the ideal and real system, and identifying solutions that close this gap [18]. Stakeholders, with diversity across roles, perspectives, geographies, and gender, were invited to participate in the engagement symposium. Individuals were eligible to participate if they had interest or experience with digital mental health tools. The inclusion criteria were kept broad to ensure we maximized the diversity among participating stakeholders. The structure of the workshop was completed using a World Café style [19], which is a structured approach to gather feedback from a large audience. The engagement workshop begins with asking participants to consider elements that would make the worst resource guide ever, and what approaches can be leveraged to prevent this from happening. Following this discussion, participants were invited to comment on what would be the critical elements that are relevant for an excellent resource. After each participant discussed these questions within the smaller group, a larger discussion was conducted before the engagement workshop was concluded. In the afternoon, participants were asked about current approaches for seeking information on digital mental health tools and how this resource document may be implemented to address current unmet needs. Challenges related to the uptake of the resource document were also discussed. Notes were taken by a member of the research team (JC) and were consolidated using a content analysis [16] approach.

Phase 4: Development of the Resource Guide

In order to develop the resource guide, the aforementioned efforts were consolidated by the research team. Foremost, after screening of the identified toolkits from the gray literature scan and environmental scan, a member of the research team (DM) consolidated and organized the list of resources based on the purpose and description provided by each toolkit. Each toolkit was also characterized by the intended audience, format, scope, language, and country of origin. The findings from the engagement workshop were then used by the research team to refine the draft of the resource document to a format and delivery that aligned with the need of stakeholders.

Results

The gray literature review and environmental scan led to the identification of 25 resources that were deemed relevant for the resource document. The engagement workshop, which was conducted with 14 participants, was used to inform the development of the resource document (Multimedia Appendix 1) [20]. A detailed description of the findings from each phase of the development (Figure 1) can be found in the next section.

Phase 1: Gray Literature Review

The gray literature review was conducted in September 2019, and a total of 19 resources were identified. Most of the identified resources from this phase of the project were websites or blog posts that contained a collection of apps (n=9). Other types of resources included app rating frameworks (n=3), implementation guides for clinicians (n=2), evaluation tools (n=1), electronic health record–related comic strips (n=2), and social media and info guides (n=2).

Of the 19 resources that were identified from the gray literature review, only 11 (60%) of the tools or documents met the inclusion criteria and were summarized for the final iteration of the toolkit. These included the HITEQ (Health Information Technology Evaluation, and Quality Center) Health App Decision Tree [21], the “e-Mental Health in Practice” document [22] from the Black Dog Institute, and patient information guides [23] from the National Health Service in the United Kingdom.

Phase 2: Environmental Scan

A total of 27 experts from Canada and the United States participated in the environmental scan. The demographics of the experts are outlined in Table 1. These experts have administrative, clinical, or research roles in mental health and either actively participate or have experience in activities (eg, development, implementation) related to digital mental health tools. There was representation across many provinces in Canada, with most participants being from Ontario (n=10). In terms of organization, most participants had affiliations with academic and government organizations. From the 42 tools or documents that were suggested by participants of the environmental scan, a total of 14 tools or documents met the inclusion criteria and were included in the final resource document.

Table 1. Demographic characteristics of participants in the environmental scan.

Characteristic	Number of participants (N=27), n (%)
Province/location	
British Columbia	1 (4%)
New Brunswick	2 (8%)
Nova Scotia	8 (30%)
Ontario	10 (37%)
Prince Edward Island	2 (8%)
Quebec	1 (4%)
Saskatchewan	1 (4%)
Outside of Canada	2 (8%)
Organizational affiliation	
Academic institution	10 (37%)
Government	9 (33%)
Hospital	3 (11%)
Nonprofit organization	5 (19%)

Characteristics of Resources

The gray literature review (phase 1) and environmental scan (phase 2) led to the identification of 25 resources in our web-based resource guide. Among the 25 resources, 9 resources (36%) provided ratings or reviews of digital mental health tools, and 3 resources (12%) provided guidance on the implementation of these technologies. Additionally, there were 4 resources (16%) that were tools for patients, 6 resources (24%) for clinicians, and 3 (12%) resources designed for both patients and clinicians.

In terms of the resources identified, most were developed in Canada (n= 10) and the United States (n=10), with other resources being from the United Kingdom (n=2), Australia (n=2), and New Zealand (n=1). Only resources developed in Canada were found to be available in French. In addition, only 60% of the resources (15/25) were developed or updated in 2018 and later. The latter resources do not have an updated date or were last updated before 2018. Most resources were developed collaboratively with private or not-for-profit organizations (n=15), academic groups (n=6), provider associations like Canadian Medical Association (n=5), health care organizations (n=2), and governments (n = 2). Likewise, funding for the development of the resource originated from not-for-profit organizations (eg, One Mind), health service organizations (eg, Ministry of Health of New Zealand), government-funded organizations (eg, Ontario Telemedicine Network), provider organizations (eg, British Medical Association), and academic institutions (eg, University of Chicago).

In terms of the audience, identified resources included content relevant to patients (n=10) and clinicians (n=20). In particular, 15 resources had clinician-specific resources, 5 resources contained patient-specific resources, and 5 resources had content for both populations. Some resources indicated a specific audience, such as primary care or general providers (n=2), frontline workers (n=2), physicians (n=1), medical school students (n=2), and researchers or app developers (n=1). Some resources targeted a specific mental health condition (eg, depression), while 13 resources focused broadly on mental health and e-mental health technologies.

As per the typology outlined by the Mental Health Commission of Canada (MHCC) [24], 5 resources reviewed and rated "computerized treatments, resources & apps," and 4 resources provided frameworks to conduct evaluations and reviews of these technologies. In addition, 3 resources provided implementation frameworks and guidance on integrating these technologies into clinical environments. These implementation frameworks were not limited to a single category within the typology.

The report from the MHCC [24] suggested that apps should disclose information related to the evidence base, cultural appropriateness, and gender responsiveness. For the 5 resources that provided app reviews, 4 of the sites included supporting evidence and 3 of the sites outlined privacy information about the apps. Although all tools are free to access, some tools reference apps that have a cost requirement, and these requirements are indicated in some tools (eg, Psyberguide). At last, only 2 of the sites provided data (eg, Practical Apps) about usability and user experience.

Phase 3: Engagement Workshop

A total of 14 participants from mental health organizations across Canada took part in the engagement workshop that was facilitated by 6 members (GS, DM, LC, HDS, AM, and JC) of the research team. The demographics of participants and facilitators in the engagement workshop are outlined in Table 2.

Table 2. Demographic characteristics of participants in the engagement workshop.

Category	Number of participants including facilitators (N=20), n (%)
Gender	
Female	16 (80%)
Male	4 (20%)
Province	
British Columbia	1 (5%)
New Brunswick	2 (10%)
Newfoundland	1 (5%)
Nova Scotia	2 (10%)
Ontario	13 (65%)
Quebec	1 (5%)
Role/contribution (multiselection)	
Clinician (eg, nurse, psychologist)	13 (65%)
Graduate trainee	4 (20%)
Indigenous perspective	1 (5%)
Person with lived experience	2 (10%)
Research personnel/expert	4 (20%)

In the first part of the exercise, the research team asked the participants, "What would make the worst toolkit ever?" Respondents suggested that accessibility barriers were an important consideration. Examples of accessibility barriers included a lack of "searchability" and poor user-friendliness of the resource guide itself. A document with too much text and use of jargon would make it difficult for the end-user to effectively integrate it into practice. In addition, respondents highlighted the need for making the resource a "living document" that does not contain outdated information and broken links. It was further noted that resources that are not culturally sensitive and not trauma-informed may also be dangerous for the end-user and can impede the value of the resource. Other factors discussed included a lack of a dissemination plan, discussion on privacy issues, and the absence of patients, families, or the community in the development of the resource guide.

The subsequent discussion aimed to address the challenges identified by exploring the questions "How could we prevent this from happening?" and "What would the best possible toolkit look like?" Participants made several suggestions including a focus on evidence-based development of the document, co-design with the audience, and development of a postdevelopment sustainability plan. It was indicated that the evidence-based development should be inclusive of the views and perspectives of intended end-users and include open-source links for readers to explore if they are interested. It was further suggested that the methodology of the resource guide be transparent in the resource document. Participants explained that the content of the resource document should also be inclusive of the different learning styles of individuals and manage the expectations of the reader (ie, relatively new field). With regard to the postdevelopment sustainability plan, participants suggested a "review cycle" where the materials would be revisited after a certain period of time to ensure up-to-date content. In addition, a follow-up/feedback loop with participants was encouraged to allow for continuous improvement of the resource document.

The afternoon session of the workshop focused on dissemination of the resource document. When participants were asked where they seek information on digital mental health tools, a variety of academic (eg, school) and professional (eg, regulatory college) organizations were listed. Other approaches included conferences, word of mouth, and the intranet of their employer. With regard to the challenges of implementing the resource document for uptake of digital mental health tools, there were concerns on the definition of the "toolkit" and who the target audience is. There was also discussion on how the scope of the toolkit may not be compatible with current structure of care systems. For example, some clinicians may not have a choice in deciding which tools would be made available to the patient, and the process would require engaging stakeholders across project management, clinical services, and privacy domains. It was also unclear if this resource document would be based on current principles of mental health care, such as the stepped care model [25]. Finally, some participants gave suggestions on delivering the contents of this document in other formats, such as a webpage as opposed to a static PDF document [20].

Phase 4: Development of the Resource Document

The findings from phases 1-3 of this project were used to develop the final iteration of the resource document (Multimedia Appendix 1) by the research team. The 45-page resource document [20] begins with background information on digital mental health tools and how various tools could be used to support the mental health needs of an individual. A set of questions were also developed to help guide a client's decision

on whether digital mental health tools are appropriate for their needs. The list of resources identified from phase 1 and 2 was then summarized in a chart by audience, format, language, country of origin, and whether the toolkit is specific to mental health. A summary of each of tool or document is subsequently provided. This summary includes additional information such as whether internet connection is required, if data is collected on the user, and the suitability of the resource for use during interaction with a client. The resource document concludes with a high-level overview of the project methodology.

Moreover, many suggestions and concerns from the engagement workshop (phase 3) were incorporated in the development of the final version of the resource guide. For example, the language used throughout the document was reflective of the suggestions of the stakeholders (eg, neutral, welcoming, and free of jargon) and brief instructions were provided at the beginning of the resource document to orient the end-user on usage of the document. The guide was also optimized for the search functionalities of the application.

Discussion

Principal Findings

Although digital mental health tools have gained significant traction and interest from patients, caregivers, family members, providers, and mental health organizations [26], uptake and integration of these technologies remain fairly poor across mental health care [8]. From the gray literature review (phase 1) and environmental scan (phase 2), a total of 25 resource guides that were relevant in supporting the uptake of digital mental health tools were identified. These resource guides were developed by various health care and mental health care organizations and targeted both patients and clinicians. Feedback on the delivery of these findings were identified from the engagement workshop with 14 participants. These findings collectively informed the development of the final resource document [20], which can be found in Multimedia Appendix 1.

In our experience, the use of a multimethod approach [13] provided a solid foundation for developing a document that aligns with the needs of the clients, caregivers, and end-users. Of the 25 resources that were identified, there was a relatively even spread of resources from the gray literature review (phase 1) and the environmental scan (phase 2). This demonstrates the importance and value of both sources as part of a comprehensive synthesis of tools or documents relevant to digital mental health tools. However, the absence of guidance on the delivery of the content can jeopardize the success of the project and lead to products that are not compatible with the needs of end-users [27]. Our engagement workshop (phase 3) was instrumental in engaging patients, clinicians and other relevant stakeholders in addressing this gap in guidance [27]. The application of the World Café [19] approach facilitated the consensus of opinions and perceptions throughout the workshop and provided great insight into how the resource document should be best delivered. Thus, the multimethod approach [13] is a valuable approach in consolidating the knowledge from each source to develop a relevant and timely resource document that is applicable for a variety of audiences.

This paper introduces a web-based resource guide [20] that our research team designed to foster the engagement of mental health patients and providers with digital mental health tools. This resource guide [20] is now publicly available for free and is expected to be used by both patients and practitioners in supporting the uptake of digital mental health tools. Patients and their caregivers may use this document to choose appropriate resources to guide the selection of a suitable app to meet their needs and requirements. In addition, using the questions that are listed on pages 9-11 of the resource document [20], individuals can also examine if digital mental health tools are appropriate and suitable for their needs, or if other (eg, in-person) interventions are necessary. Similarly, this resource guide will help providers and clinicians become acquainted and knowledgeable about the use of digital mental health technologies. In particular, providers may consider this resource document in speaking to patients and family members about the use of digital mental health tools as part of care [28]. Providers may also consider using this document as a means to guide the conversation and planning of which tools should be used and in what manner [29]. At a broader level, health care administrators may also use this newly developed resource to develop proper training and support for providers interested in using digital health tools in their practice.

During the development of the document, a number of evidence gaps were identified. Chiefly, the identified resources included in the resource guide fail to cover many of the technologies outlined in the MHCC typology [24]. As most of the published resources focused on mobile health apps, there is a lack of resources for other technologies such as virtual reality technology and robots [30-32]. Future work that focuses on the uptake of these emerging technologies would be useful [8]. Additionally, most of the resources are available in English only. Making the identified resources available in other languages, such as French, would be helpful [33], particularly given the abundance of French-speaking people in Canada. Similarly, none of the identified tools encompass other cultures, including indigenous perspectives, or address cultural appropriateness. Furthermore, given the emerging crisis of caregivers, future tools should explore the role of e-mental health technologies to support the needs of caregivers [34].

Although this resource document was developed prior to the COVID-19 pandemic, we expect that it will continue to be of value for supporting the ongoing mental health needs and demands during and beyond the pandemic. However, it is important to note that the pandemic has greatly accelerated the uptake of some digital mental health tools [35]. For example, many organizations have converted their delivery of outpatient or ambulatory care to telemental health visits (eg, using Zoom or Microsoft Teams) [28]. Some of these changes in practice have also led to the use of digital mental health tools, such as patient portals and mobile apps [36]. As part of the postdevelopment sustainability plan, it would be of value to synthesize the recent outputs from the use of digital mental health tools during the COVID-19 pandemic [37] to the current resource document.

Limitations

Although this resource guide has been developed with extensive input from gray literature and experts in the field, it has yet to be integrated into clinical workflows and refined by providers with experience using this resource with patients [8]. Additionally, the identification of relevant resources was limited to those available in the English language. Experts who participated in the engagement workshop and environmental scan were also all from North America. Exploring resources of other languages and consulting experts from other countries (eg, the United Kingdom) may provide insight into novel resources not identified here. Although, the findings from the engagement workshop were derived from a technique used to support consensus (World Café) [19], further validation of the findings (eg, member checking) [38] was not conducted with the participants of the group or with other participants. Moreover, the participants were not engaged during the review of the final iteration of the resource document. Thus, there is a need to evaluate the efficacy of the toolkit. Currently, it is unclear how the document may impact the uptake of digital mental health tools [8,26,39] (eg, patient–clinician relationship [40-44]), and examining the analytics and usage of these tools [45] may provide insights into this evidence gap.

Future Directions

This document is the product of a careful and meaningful synthesis of resources that encourage uptake of digital mental health tools, and future work should explore how these tools can be or have been integrated into clinical care pathways for mental health conditions (eg, depression). This may involve promoting and sharing the resource document across organizations that may be interested in the uptake of digital mental health tools. This may include the identified recommendations from the engagement workshop on expanding the delivery of the resource document to web-based approaches (eg, website, mobile app). It would also be useful to validate the findings from the resource document (eg, with other similar toolkits or documents) and to examine the efficacy of this resource document in addressing barriers and opportunities of digital mental health tools [38]. This can be conducted using a mixed methods approach [46] to incorporate both measurable outcomes and user experience. Moreover, at the solution level, identifying strategies to enhance uptake of emerging digital mental health tools is warranted. There remains a need to examine factors (eg, gamification [47]) that may relate to the engagement with digital mental health tools [8]. Finally, with regard to the recent events of the COVID-19 pandemic [35,37], it would also be useful to explore additional work and guidelines that are being developed during the pandemic and their impact on engagement with digital mental health tools, particularly concerning virtual care and telemental health [48,49].

Conclusions

This paper describes the development of a web-based resource guide that we designed to guide the uptake of digital mental health tools into the clinical environment through a multimethod approach. The document, which is available online for public use, includes a number of resources to guide the selection, implementation, and evaluation of digital mental health tools. Although these resources cover many objectives and audiences, there are disproportionately fewer resources available for emerging technologies like virtual reality. Moreover, the lack of resources designed for caregivers warrants further research. There is also a critical need to ensure that resources are inclusive of the needs of diverse cultures, including the First Nations, Inuit, and Métis people of Canada. Finally, future work should explore how this resource guide can be adopted and integrated into clinical environments.

Acknowledgments

Authors would like to sincerely thank all the stakeholders, participants, and topic experts who participated in the workshop and environmental scan during this project. This project was funded by Canada Health Infoway.

Authors' Contributions

GS contributed to the conceptualization of the study, methodology, formal analysis, investigation, supervision, project administration, and reviewing and editing of the manuscript. DM contributed to the conceptualization of the study, methodology, formal analysis, investigation, project administration, and the reviewing and editing of the manuscript. NT contributed to the conceptualization of the study and methodology. EM contributed to the conceptualization of the study and methodology. HDS contributed to the investigation, visualization, literature review, and the reviewing and editing of the manuscript. NS contributed to the conceptualization of the study and methodology. VS contributed to the literature review. JC contributed to the stakeholder workshop. BL contributed to methodology and the writing of the manuscript. LC contributed to the conceptualization of the study, methodology, and stakeholder workshop. AC contributed to the gap analysis, methodology, and stakeholder workshop. All authors gave final approval of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Final version of the resource document.
[PDF File (Adobe PDF File), 499 KB - jmir_v23i4e25773_app1.pdf]

References

1. COVID-19 national survey dashboard. Centre for Addiction and Mental Health. URL: https://www.camh.ca/en/health-info/mental-health-and-covid-19/covid-19-national-survey [accessed 2020-11-12]
2. Mental illness. National Institute of Mental Health. 2019. URL: https://www.nimh.nih.gov/health/statistics/mental-illness.shtml [accessed 2020-11-03]
3. U.S. Burden of Disease Collaborators. The state of US health, 1990-2010: burden of diseases, injuries, and risk factors. JAMA 2013 Aug 14;310(6):591-608 [FREE Full text] [doi: 10.1001/jama.2013.13805] [Medline: 23842577]
4. Making the case for investing in mental health in Canada. Mental Health Commission of Canada. 2016. URL: https://www.mentalhealthcommission.ca/sites/default/files/2016-06/Investing_in_Mental_Health_FINAL_Version_ENG.pdf [accessed 2020-11-03]
5. Pfefferbaum B, North CS. Mental health and the Covid-19 pandemic. N Engl J Med 2020 Aug 06;383(6):510-512. [doi: 10.1056/NEJMp2008017] [Medline: 32283003]
6. APA app advisor. American Psychiatric Association. URL: https://www.psychiatry.org/psychiatrists/practice/mental-health-apps [accessed 2020-11-03]
7. Department of Health and Aging. E-Mental health strategy for Australia. Australian Government. 2012. URL: https://www1.health.gov.au/internet/publications/publishing.nsf/Content/mental-pubs-e-emstrat-toc [accessed 2020-11-03]
8. Torous J, Nicholas J, Larsen ME, Firth J, Christensen H. Clinical review of user engagement with mental health smartphone apps: evidence, theory and improvements. Evid Based Ment Health 2018 Aug;21(3):116-119. [doi: 10.1136/eb-2018-102891] [Medline: 29871870]
9. Toolkit for e-Mental health implementation. Mental Health Commission of Canada. 2018. URL: https://www.mentalhealthcommission.ca/sites/default/files/2018-09/E_Mental_Health_Implementation_Toolkit_2018_eng.pdf [accessed 2021-02-08]
10. NHS apps library. National Health Service. URL: https://www.nhs.uk/apps-library/ [accessed 2021-02-08]
11. Mental health apps: how to make an informed choice. Mental Health Commission of Canada. URL: https://www.mentalhealthcommission.ca/sites/default/files/2018-01/eMH_app_eng.pdf [accessed 2021-02-08]
12. Section 6: toolkit guidance. Agency for Healthcare Research and Quality. 2013. URL: https://www.ahrq.gov/sites/default/files/publications/files/pcguide6.pdf [accessed 2021-02-08]
13. Brewer J, Hunter A. Multimethod Research: A Synthesis of Styles. Newbury Park, CA: Sage Publications; 1989.
14. Pappas C, Williams I. Grey literature: its emerging importance. Journal of Hospital Librarianship 2011 Jul;11(3):228-234. [doi: 10.1080/15323269.2011.587100]
15. Grey Matters: a practical tool for searching health-related grey literature. Canadian Agency for Drugs and Technologies in Health. 2019. URL: https://www.cadth.ca/resources/finding-evidence/grey-matters [accessed 2021-02-09]
16. Hsieh H, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res 2005 Nov;15(9):1277-1288. [doi: 10.1177/1049732305276687] [Medline: 16204405]
17. Van De Ven AH, Johnson PE. Knowledge for theory and practice. AMR 2006 Oct;31(4):802-821. [doi: 10.5465/amr.2006.22527385]
18. Ilevbare IM, Probert D, Phaal R. A review of TRIZ, and its benefits and challenges in practice. Technovation 2013 Feb;33(2-3):30-37. [doi: 10.1016/j.technovation.2012.11.003]
19. Brown J, Isaacs D. The World Café – Shaping Our Futures through Conversations that Matter. San Francisco, CA: Berrett-Koehler Publishers; 2005.
20. Strudwick G, McLay DW, Currie LM, Thomson N, Maillet E, Campbell J, et al. Digital mental health tools: resources to support mental health clinical practice. Centre for Addiction and Mental Health. Toronto; 2020. URL: https://www.camh.ca/-/media/images/all-other-images/covid-19-professionals/final-digital-mh-resource-document-april-2020-pdf.pdf?la=en&hash=78EC69BE5AF6F92C866E8FDCF58D24449C6AB048 [accessed 2021-02-09]
21. HITEQ health app decision tree. The Health Information and Technology, Evaluation and Quality Center. URL: https://hiteqcenter.org/Resources/HITEQ-Resources/hiteq-health-app-decision-tree-1 [accessed 2021-02-09]
22. e-Mental health in practice. Black Dog Institute. URL: https://www.blackdoginstitute.org.au/education-training/health-professionals/emental-health-in-practice [accessed 2021-02-08]
23. Patient information guides. National Health Service. URL: https://www.england.nhs.uk/gp-online-services/support/supporting-material/patient-information-guides/ [accessed 2021-02-08]
24. E-mental health in Canada: transforming the mental health system using technology. Ottawa, ON: Mental Health Commission of Canada; 2014. URL: https://www.mentalhealthcommission.ca/sites/default/files/MHCC_E-Mental_Health-Briefing_Document_ENG_0.pdf [accessed 2021-02-08]
25. Bower P, Gilbody S. Stepped care in psychological therapies: access, effectiveness and efficiency. Narrative literature review. Br J Psychiatry 2005 Jan;186:11-17. [doi: 10.1192/bjp.186.1.11] [Medline: 15630118]
26. Torous J, Wisniewski H, Liu G, Keshavan M. Mental health mobile phone app usage, concerns, and benefits among psychiatric outpatients: comparative survey study. JMIR Ment Health 2018 Nov 16;5(4):e11715 [FREE Full text] [doi: 10.2196/11715] [Medline: 30446484]

27. Strudwick G, Leung K, McLean D, Booth R, Borycki E, Rossetti S, et al. Patient and family engagement in health information technology initiatives: findings of a literature review, focus groups and symposium. Centre for Addiction and Mental Health. Toronto; 2019. URL: https://www.camh.ca/-/media/files/pdfs---reports-and-books---research/camh-health-it-initiative-28mar2018-6002.pdf [accessed 2021-02-08]
28. Schwamm LH, Estrada J, Erskine A, Licurse A. Virtual care: new models of caring for our patients and workforce. Lancet Digit Health 2020 Jun;2(6):e282-e285. [doi: 10.1016/S2589-7500(20)30104-7] [Medline: 32382724]
29. Noel VA, Carpenter-Song E, Acquilano SC, Torous J, Drake RE. The technology specialist: a 21st century support role in clinical care. NPJ Digit Med 2019;2:61 [FREE Full text] [doi: 10.1038/s41746-019-0137-6] [Medline: 31388565]
30. Freeman D, Reeve S, Robinson A, Ehlers A, Clark D, Spanlang B, et al. Virtual reality in the assessment, understanding, and treatment of mental health disorders. Psychol Med 2017 Oct;47(14):2393-2400 [FREE Full text] [doi: 10.1017/S003329171700040X] [Medline: 28325167]
31. Lau HM, Smit JH, Fleming TM, Riper H. Serious games for mental health: are they accessible, feasible, and effective? a systematic review and meta-analysis. Front Psychiatry 2016;7:209 [FREE Full text] [doi: 10.3389/fpsyt.2016.00209] [Medline: 28149281]
32. Huschilt J, Clune L. The use of socially assistive robots for dementia care. J Gerontol Nurs 2012 Oct;38(10):15-19. [doi: 10.3928/00989134-20120911-02] [Medline: 22998095]
33. Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. Spine (Phila Pa 1976) 2000 Dec 15;25(24):3186-3191. [doi: 10.1097/00007632-200012150-00014] [Medline: 11124735]
34. Tal A, Torous J. The digital mental health revolution: Opportunities and risks. Psychiatr Rehabil J 2017 Sep;40(3):263-265. [doi: 10.1037/prj0000285] [Medline: 28891658]
35. Taylor CB, Fitzsimmons-Craft EE, Graham AK. Digital technology can revolutionize mental health services delivery: The COVID-19 crisis as a catalyst for change. Int J Eat Disord 2020 Jul;53(7):1155-1157 [FREE Full text] [doi: 10.1002/eat.23300] [Medline: 32449523]
36. Barsom EZ, Feenstra TM, Bemelman WA, Bonjer JH, Schijven MP. Coping with COVID-19: scaling up virtual care to standard practice. Nat Med 2020 May;26(5):632-634. [doi: 10.1038/s41591-020-0845-0] [Medline: 32405054]
37. Torous J, Jän Myrick K, Rauseo-Ricupero N, Firth J. Digital mental health and COVID-19: using technology today to accelerate the curve on access and quality tomorrow. JMIR Ment Health 2020 Mar 26;7(3):e18848 [FREE Full text] [doi: 10.2196/18848] [Medline: 32213476]
38. Lincoln YS, Guba EG. But is it rigorous? Trustworthiness and authenticity in naturalistic evaluation. New Directions for Program Evaluation 1986 Jun;1986(30):73-84. [doi: 10.1002/ev.1427]
39. Lipschitz J, Miller CJ, Hogan TP, Burdick KE, Lippin-Foster R, Simon SR, et al. Adoption of mobile apps for depression and anxiety: cross-sectional survey study on patient interest and barriers to engagement. JMIR Ment Health 2019 Jan 25;6(1):e11334 [FREE Full text] [doi: 10.2196/11334] [Medline: 30681968]
40. Berry N, Bucci S, Lobban F. Use of the internet and mobile phones for self-management of severe mental health problems: qualitative study of staff views. JMIR Ment Health 2017 Nov 01;4(4):e52 [FREE Full text] [doi: 10.2196/mental.8311] [Medline: 29092809]
41. Ignatowicz A, Slowther A, Elder P, Bryce C, Hamilton K, Huxley C, et al. Ethical implications of digital communication for the patient-clinician relationship: analysis of interviews with clinicians and young adults with long term conditions (the LYNC study). BMC Med Ethics 2018 Feb 23;19(1):11 [FREE Full text] [doi: 10.1186/s12910-018-0250-0] [Medline: 29475437]
42. Lattie EG, Nicholas J, Knapp AA, Skerl JJ, Kaiser SM, Mohr DC. Opportunities for and tensions surrounding the use of technology-enabled mental health services in community mental health care. Adm Policy Ment Health 2020 Jan;47(1):138-149 [FREE Full text] [doi: 10.1007/s10488-019-00979-2] [Medline: 31535235]
43. Orlowski S, Lawn S, Matthews B, Venning A, Wyld K, Jones G, et al. The promise and the reality: a mental health workforce perspective on technology-enhanced youth mental health service delivery. BMC Health Serv Res 2016 Oct 10;16(1):562 [FREE Full text] [doi: 10.1186/s12913-016-1790-y] [Medline: 27724951]
44. Schueller SM, Washburn JJ, Price M. Exploring mental health providers' interest in using web and mobile-based tools in their practices. Internet Interv 2016 May;4(2):145-151 [FREE Full text] [doi: 10.1016/j.invent.2016.06.004] [Medline: 28090438]
45. Pham Q, Shaw J, Morita PP, Seto E, Stinson JN, Cafazzo JA. The service of research analytics to optimize digital health evidence generation: multilevel case study. J Med Internet Res 2019 Nov 11;21(11):e14849 [FREE Full text] [doi: 10.2196/14849] [Medline: 31710296]
46. Creswell J, Plano Clark VL. Designing and Conducting Mixed Methods Research. Los Angeles: Sage Publications; 2011.
47. El-Hilly AA, Iqbal SS, Ahmed M, Sherwani Y, Muntasir M, Siddiqui S, et al. Game on? Smoking cessation through the gamification of mHealth: a longitudinal qualitative study. JMIR Serious Games 2016 Oct 24;4(2):e18 [FREE Full text] [doi: 10.2196/games.5678] [Medline: 27777216]

48. Yellowlees P, Nakagawa K, Pakyurek M, Hanson A, Elder J, Kales HC. Rapid conversion of an outpatient psychiatric clinic to a 100% virtual telepsychiatry clinic in response to COVID-19. Psychiatr Serv 2020 Jul 01;71(7):749-752. [doi: 10.1176/appi.ps.202000230] [Medline: 32460683]
49. Shore JH, Schneck CD, Mishkind MC. Telepsychiatry and the coronavirus disease 2019 pandemic-current and future outcomes of the rapid virtualization of psychiatric care. JAMA Psychiatry 2020 May 11;77(12):1211-1212. [doi: 10.1001/jamapsychiatry.2020.1643] [Medline: 32391861]

Abbreviations

APA: American Psychiatric Association
HITEQ: Health Information Technology Evaluation, and Quality Center
MHCC: Mental Health Commission of Canada
TRIZ: Theory of Inventive Problem Solving

Edited by R Kukafka; submitted 14.11.20; peer-reviewed by J Bierbooms, N Kamyari, MA Bahrami, M Sobolev, A Bhattaharya, Z Aghaei, A Khaleghi; comments to author 30.12.20; revised version received 17.02.21; accepted 18.03.21; published 22.04.21.

Please cite as:
Strudwick G, McLay D, Lo B, Shin HD, Currie L, Thomson N, Maillet É, Strong V, Miller A, Shen N, Campbell J
Development of a Resource Guide to Support the Engagement of Mental Health Providers and Patients With Digital Health Tools: Multimethod Study
J Med Internet Res 2021;23(4):e25773
URL: https://www.jmir.org/2021/4/e25773
doi:10.2196/25773
PMID:33885374

Original Paper

Investigating the Use of Electronic Well-being Diaries Completed Within a Psychoeducation Program for University Students: Longitudinal Text Analysis Study

Myles-Jay Anthony Linton[1,2], BSc, PhD; Sarah Jelbert[3], BA, MSc, PhD; Judi Kidger[1], MA, MSc, PhD; Richard Morris[1], BSc, MSc, PhD; Lucy Biddle[1], BA, MA, PhD; Bruce Hood[3], MA, MPhil, PhD

[1]Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, United Kingdom

[2]School of Education, University of Bristol, Bristol, United Kingdom

[3]School of Experimental Psychology, Bristol Cognitive Development Centre, University of Bristol, Bristol, United Kingdom

Corresponding Author:
Myles-Jay Anthony Linton, BSc, PhD
Population Health Sciences
Bristol Medical School
University of Bristol
Canynge Hall
39 Whatley Road
Bristol, BS8 2PS
United Kingdom
Phone: 44 1173314380
Email: mj.linton@bristol.ac.uk

Abstract

Background: Psychoeducation has the potential to support students experiencing distress and help meet the demand for support; however, there is a need to understand how these programs are experienced. Web-based diaries are a useful activity for psychoeducation because of their therapeutic benefits, ability to capture naturalistic data relevant to well-being, and appropriateness for text analysis methods.

Objective: This study aims to examine how university students use electronic diaries within a psychoeducation program designed to enhance mental well-being.

Methods: The Science of Happiness course was administered to 154 undergraduate students in a university setting (the United Kingdom). Diaries were collected from the students for 9 weeks. Baseline well-being data were collected using the Short Warwick-Edinburgh Mental Wellbeing Scale (SWEMWBS). The percentage of negative and positive emotion words used in diaries (emotional tone) and use of words from five life domains (social, work, money, health, and leisure) were calculated using the Linguistic Inquiry and Word Count 2015 software. Random effects (generalized least squares) regression models were estimated to examine whether time, diary characteristics, demographics, and baseline well-being predict the emotional tone of diaries.

Results: A total of 149 students participated in the diary study, producing 1124 individual diary entries. Compliance with the diary task peaked in week 1 (n=1041, 92.62%) and was at its lowest in week 3 (n=807, 71.81%). Compared with week 1, diaries were significantly more positive in their emotional tone during week 5 (mean difference 23.90, 95% CI 16.89-30.90) and week 6 (mean difference 26.62, 95% CI 19.35-33.88) when students were tasked with writing about gratitude and their strengths. Across weeks, moderate and high baseline SWEMWBS scores were associated with a higher percentage of positive emotion words used in diaries (increases compared with students scoring low in SWEMWBS were 5.03, 95% CI 0.08-9.98 and 7.48, 95% CI 1.84-13.12, respectively). At week 1, the diaries of students with the highest levels of baseline well-being (82.92, 95% CI 73.08-92.76) were more emotionally positive on average than the diaries of students with the lowest levels of baseline well-being (59.38, 95% CI 51.02-67.73). Diaries largely focused on the use of social words. The emotional tone of diary entries was positively related to the use of leisure (3.56, 95% CI 2.28-4.85) and social words (0.74, 95% CI 0.21-1.27), and inversely related to the use of health words (−1.96, 95% CI −3.70 to −0.22).

Conclusions: We found evidence for short-term task-specific spikes in the emotional positivity of web-based diary entries and recommend future studies examine the possibility of long-term impacts on the writing and well-being of students. With student well-being strategies in mind, universities should value and encourage leisure and social activities.

(J Med Internet Res 2021;23(4):e25279) doi:10.2196/25279

KEYWORDS

psychoeducation; diary; students; text analysis; wellbeing; science of happiness; university; emotional tone; e-mental health; mobile phone

Introduction

Background

There is mounting concern regarding increases in the number of young people (aged 16-24 years) reporting long-standing mental health difficulties [1]. Rising university attendance among school-leavers approaching the end of adolescence has been observed globally [2], making universities an increasingly advantageous setting for providing mental health and well-being interventions [3]. Alongside treatment-based approaches for managing mental ill-health, there is a need to better understand the use of more accessible preventative approaches that equip students with resources to protect and promote their well-being [4].

Student Mental Health and Well-being

The mental health of university students has been a priority in the higher education sector for at least a century [5]. In the last decade, however, academics have described a growing crisis in university students' mental health [6,7]. For many students, university education is undertaken at a crucial period of transition into adulthood, associated with social, psychological, and developmental challenges across multiple life domains [8,9]. The World Health Organization's World Mental Health International College Student project in 8 countries found that 31% of students had experienced a mental health disorder (such as anxiety or depression) within the previous 12 months [10]. As demand from students for counseling support continues to exceed supply, there is growing pressure to invest in strategies with the potential to prevent students from reaching the point where high-intensity one-on-one support is required [11].

Embedding Psychoeducation in University

Psychoeducation, which involves the provision of information, tips for self-management, and guidance for staying well [12], may be one method for embedding well-being–enhancing interventions within university life. In a framework adapted from World Health Organization guidelines, experts have highlighted that alongside formal specialized treatment for severe mental ill-health problems, there is a need for structured and unstructured support for students experiencing varying levels of distress [13]. In line with this model, psychoeducation interventions can be used to facilitate social support networks and provide space for students to engage in self-care strategies. Research into university-based psychoeducation continues to expand [14,15], and there is a need to understand how students, with varying levels of well-being, experience and engage with these interventions.

Practical and Methodological Use of Diaries

Diaries are a promising tool for integration into psychoeducation programs because of their potential for both therapeutic benefit and data collection [16]. Diaries offer writers a naturalistic space to reflect on experiences, and provide researchers with a depth of detail that is difficult to accurately achieve when relying on tools that require participants to retrospectively recall events from their past [17].

One of the earliest diary studies examining well-being in university found that common worries included academic stress and common sources of happiness included friendships [18]. Previous research with students has also indicated that participation in writing tasks about positive life events may lead to improvements in mood [19]. In a study where students were asked to write about their thoughts and feelings in relation to starting university, students who screened positive for depression (Beck Depression Index scores above 14) used significantly more negative emotion words than students without depression (scores below 7) [20]. These studies highlight a link between the emotional content of writing and how students report feeling about their mental health and well-being. Less is known about the balance between negative and positive feelings [21] and the emotional tone (balance of positive and negative words) of writing [22].

Technological Developments in Diary Methodology

The availability and progression of technology has been critical in the ongoing advancement of diary-based research. In the field of diary analysis, researchers have used technology to identify prospective participants on the web [23], automatically prompt participants with poor compliance to diary completion tasks [24], and automatically monitor the time of day when respondents choose to complete their diaries [25].

Significant advancement has also been made in the analysis of open-text data contained within diaries. Natural language processing methods are being increasingly used to automate the analysis of diaries, as seen in a web-based eating disorder intervention [26]. Statistical models have been developed to predict depression severity based on the use of emotional words within social media posts [27]. The use of technology to enhance aspects of the research process, however, should not be undertaken without careful consideration. For example, where tasks are automated, it remains crucial to understand how the process is undertaken and be mindful of any trade-offs involved in generating these efficiencies.

This Study and Research Questions

This study aims to examine how university students use electronic diaries within a psychoeducation program designed to develop mental well-being skills (Science of Happiness course). Longitudinal diary data enabled us to examine how the program was experienced and how the proportion of positive and negative emotion words (emotional tone) within diaries fluctuated over time. Automated text analysis methods were used to analyze the written content of the diaries. This study has 5 key research questions:

1. How compliant are students to the web-based diary task across weeks?
2. How does the proportion of positive and negative emotion words (emotional tone) within the diaries develop over time?
3. How are time, sociodemographics, diary characteristics, and baseline self-reported well-being related to the emotional tone of diaries? (model 1)
4. Is the trajectory of emotional tone within diaries dependent on baseline levels of well-being? (model 2)
5. Which life domains do students discuss most in their diaries, and how do these topics relate to the emotional tone of diary entries? (model 3)

Methods

Participants

The sample consisted of 154 university undergraduate students at a university in the UK. The *Science of Happiness* course was offered to students on 14 undergraduate courses in their first year of study (including *study abroad programs*) in exchange for academic credits. The course began in the first semester of the study (September 2019) and involved weekly lectures, weekly group tasks, and weekly diary entries (the focus of this research). The main evaluation examining the impact of the intervention is presented in a publication submitted separately [28].

Data Collection

Sociodemographic characteristics and well-being data were collected at baseline. Parallel to the intervention, linked anonymized diary entries were collected digitally on a weekly basis at 9 time points.

Each student was assigned a study participant number, and no personally identifiable data were included in the research data set. Diary data were extracted from the web platform they were submitted to (Blackboard) and deposited in a Microsoft Excel spreadsheet stored on a secure server. Within the software, students were only able to submit one diary entry per week. The study obtained ethical approval from the University of Bristol Faculty Ethics Review Committee (reference: 27061987862).

Electronic Diaries

The diaries we collected were *solicited*, given that they were requested by us for this study, rather than being spontaneous [29]. Students were able to submit their diary entries on the web via a desktop or laptop or smartphone. Students were prompted to write in their diaries about events in the week that had influenced their well-being or they were given guidance to write about their well-being in relation to specific topics such as their goals for the future. These tasks are described in Multimedia Appendix 1. Although this imposed some structure on the task, care was taken to encourage open responses and space for respondents to reflect without excessive constraints [29].

Measures

A brief sociodemographic survey was used to capture data on gender, age, nationality, and ethnicity. Mental well-being was measured using the 7-item Short Warwick-Edinburgh Mental Wellbeing Scale (SWEMWBS) [30]. The original full-length Warwick-Edinburgh Mental Wellbeing Scale has demonstrable content validity, structural validity, criterion validity, internal consistency, and test-retest reliability across a UK nationally representative sample of adults and a specific sample of students in 2 UK universities [31]. The shortened 7-item version of the scale has been found to be internally consistent [30], and data on UK national norms have since been published to aid in interpreting scores [32]. Responses to the SWEMWBS are scored on a 5-point scale ranging from 1 (none of the time) to 5 (all of the time). Total scores were transformed, and higher scores indicated greater levels of mental well-being. In the regression analysis, well-being was recoded into a 3-level variable representing scores in the first (low), second (moderate), and third (high) tertiles of scores.

Text Analysis

This study used a *dictionary method* approach to text analysis. This deductive approach involved automating the analysis of diary content based on a predetermined dictionary-based coding scheme [33].

Data Processing and Cleaning

All diaries were completed in the English language. Text data were stored in Microsoft Excel in a wide format (one participant per row and 1 week of diary entries per column). Typos were managed using guidance developed by the Language Use and Social Interaction lab [34]. A manual rather than an automated process was selected to preserve as much of the text as possible in its original form.

All diaries were manually scanned for typos, and words such as *roominate* were changed to *ruminate* to ensure the text analysis software was able to correctly identify the words used. Words such as *kinda* were not corrected because of the ability of text analysis software to recognize and correctly classify *slang text-speech*. In total, 892 words were manually corrected during the data cleaning process.

Linguistic Inquiry and Word Count Text Analysis Variables

Linguistic Inquiry and Word Count (LIWC) 2015 is a stand-alone piece of software that has been empirically validated to analyze the linguistic, social, and psychological content of text data [35]. LIWC 2015 uses an internal dictionary of almost 6400 words coded into different categories and counts the frequency of words used within any target text data. The software is now in its fourth major revision [22]. In a sample of 117,779 pieces of text data (novels, tweets, natural speech,

expressive writing, and Twitter), LIWC 2015 correctly classified an average of 85.18% of words (SD 5.36%) [22]. In this study, we generated 7 LIWC text variables: positive emotions (eg, love, nice, and sweet), negative emotions (eg, hurt, ugly, and nasty), social (eg, mate, talk, and they), work (eg, job, class, and boss), health (eg, clinic, flu, and pill), leisure (eg, cook, TV, and movie), and money (eg, audit, cash, and owe). Each variable reflects the total number of words within each diary entry that falls into these prespecified categories of the LIWC's internal dictionary. We also used LIWC to automatically calculate the word count of each diary. Using these variables builds on the use of sentiment analysis approaches in dictionary-based text analysis [33].

Emotional Tone

We also used the emotional tone variable generated by the LIWC software. This reflects the ratio of positive and negative emotion words. If a diary entry contained 100 words and 5 of its words were found in the *negative emotion* section of LIWC's internal dictionary and 5 of its words were found in the *positive emotion* section of LIWC's internal dictionary, the *positive emotion* score would be 5, the *negative emotion* score would be 5, and the *emotional tone* score would be 50. Higher scores on the emotional tone variable indicate a greater ratio of positivity within the text of diary entries, with scores above 50 indicating a greater proportion of positive words and scores below 50, indicating a greater proportion of negative words.

Statistical Analysis

Data were analyzed using STATA 16 (StataCorp LLC) [36]. Sociodemographic characteristics (age, gender, ethnicity, and nationality) and diary characteristics (word count, diary entries, and emotional tone) were summarized descriptively for the whole sample. Participants' characteristics were compared descriptively according to levels of compliance to the diary task. Diaries were also subgrouped based on word count to compare the characteristics of students who wrote the most and least.

The trajectory of emotional tone across weeks is displayed in line graphs. The trajectory of emotional tone seen in the whole data set (available-case analysis) was compared with the trajectory seen for the subsample with full compliance to diary entry (complete-case analysis).

We applied 4 models using random effects generalized least squares with an autoregressive disturbances regression approach (Table 1). This method was selected because the diaries in our data set are clustered around repeated measures from the same individuals and are thus not independent and because of the need for a method that is robust to variations in the number of repeated measures collected across individuals. Model 1 addressed research question 3, model 2 addressed research question 4, and model 3 addressed research question 5. Model 0 (a basic model with only week effects) was only estimated to establish whether our main model (model 1) had an improved statistical model fit (Wald chi-square) following the addition of sociodemographics, diary characteristics, and baseline well-being variables.

Table 1. Random effects generalized least squares with autoregressive disturbances regression models estimated.

Model	Sample (n=154), n (%)	Diaries (n=1124), n (%)	Dependent variable	Independent variables
Model 0: investigating the role of week effects only on the emotional tone of diary entries	110 (71.43)[a]	855 (76.07)	Emotional tone	Time (weeks)
Model 1: investigating the role of week effects, sociodemographics, diary characteristics, and baseline well-being on the emotional tone of diary entries	110 (71.43)	855 (76.07)	Emotional tone	Time (weeks), gender, age, diary word count, total diary entries, and baseline well-being
Model 2: investigating interaction effects between baseline well-being and week (Multimedia Appendix 2)	110 (71.43)	855 (76.07)	Emotional tone	Time (weeks), gender, age, diary word count, total diary entries, baseline well-being, and interactions between baseline well-being and time
Model 3: investigating the relationship between word use in 5 life domains on the emotional tone of diary entries	120 (77.92)	927 (82.47)	Emotional tone	Time (weeks), social words, work words, health words, money words, leisure words, gender, age, diary word count, and total diary entries

[a]Sample of respondents with complete data on variables in model 1 to enable Wald chi-square model fit comparisons (Multimedia Appendix 3).

Results

Sociodemographic and Diary Characteristics

A summary of the sample characteristics is provided in Table 2. Of a total sample of 154 participants, 149 (96.8%) participated in the diary completion activity, resulting in a total of 1124 diary entries. Participants had a mean age of 19.29 years (SD 1.47 years) and were predominantly female (93/124, 75.0%). The sample mainly consisted of White (100/123, 81.3%) and UK nationality students (96/121, 79.3%). Participants, on average, completed a mean of 7.54 diary entries (SD 1.47), and diary entries had a mean word count of 209.63 (SD 165.79). The text analysis software LIWC correctly captured 92% of the 235,621 words analyzed in this study, which is in line with expectations from previous LIWC analyses. The mean emotional tone of diary entries was 75.15 (SD 28.60).

Table 2. Total sample sociodemographics.

Sample characteristics	Values
Age (years; n=124), mean (SD)	19.29 (1.47)
Gender (n=124), n (%)	
Female	93 (75.0)
Male	29 (23.4)
Prefer not to say	2 (1.6)
Ethnicity (n=123), n (%)	
Asian or Asian British	12 (9.8)
Black, African, Caribbean, or Black British	3 (2.4)
Mixed or multiple ethnic groups	5 (4.1)
Other ethnic group	3 (2.4)
White	100 (81.3)
Nationality (n=121), n (%)	
British	96 (79.3)
All other nationalities	25 (20.7)
Diary characteristics (n=149), mean (SD)	
Word count	209.63 (165.79)
Diary entries	7.54 (1.47)
Emotional tone	75.15 (28.60)

From 855 diaries, the SWEMWBS scores were recoded by tertiles for the diaries of students with low well-being (range 14.08-19.98; n=296), moderate well-being (range 20.73-22.35; n=345), and high well-being (range 23.21-25.03; n=214).

Compliance with the diary task by week is displayed in Figure 1, which indicates that week 1 (138/149, 92.6%) and week 5 (136/149, 91.3%) had the highest compliance rates, whereas week 3 (107/149, 71.8%) and week 9 (108/149, 72.5%) had the lowest compliance rates.

Figure 1. Compliance with diary completion activity.

Diary Entry Compliance—Subgroup Analysis

Participants were grouped by those who completed all 9 entries (44/149, 29.5%), 8 entries (45/149, 30.2%), 7 entries (32/149, 21.5%), or fewer than 6 entries (28/149, 18.8%). Descriptive statistics for the characteristics of participants across the 4 levels of diary compliance are presented in Table 3. Participants who were fully compliant with the diary activity (9 entries) had higher diary word counts (mean 242.96, SD 201.88) than participants who completed 6 or fewer diary entries (mean 175.00, SD 160.59). Furthermore, participants who were fully compliant had the highest levels of baseline well-being (mean 22.16, SD 2.57) compared with students with 6 of fewer diary entries (mean 20.78, SD 2.69). The emotional tone and percentage of females was highest for participants in the most compliant group; however, the pattern across levels of compliance was less clear.

Table 3. Key sample and diary characteristics by diary-compliance subsamples.

Characteristics	Diary compliance			
	9 entries	8 entries	7 entries	≤6 entries
Diaries (n=149), n	396	360	224	396
Word count (n=149), mean (SD)	242.96 (201.88)	213.00 (142.24)	167.53 (112.14)	175.00 (160.59)
Age (years; n=124), mean (SD)	19.26 (1.38)	19.26 (1.54)	19.04 (0.74)	19.74 (2.10)
Gender, female (n=124), n (%)	35 (28.2)	25 (20.2)	20 (16.1)	13 (10.5)
Baseline well-being (n=119), mean (SD)	22.16 (2.57)	20.91 (2.45)	21.46 (3.33)	20.78 (2.69)
Emotional tone (n=149), mean (SD)	77.39 (26.27)	75.33 (28.95)	70.68 (30.88)	75.53 (29.67)

Word Count—Subgroup Analysis

There were no clear patterns in the age of respondents or emotional tone of diary entries based on the 4 word count categories (Table 4). Longer diary entries had the highest proportion of female respondents (155/281, 55.2% for diaries of *0-103.5* word count length and 195/280, 69.6% for diaries of *265+* word count).

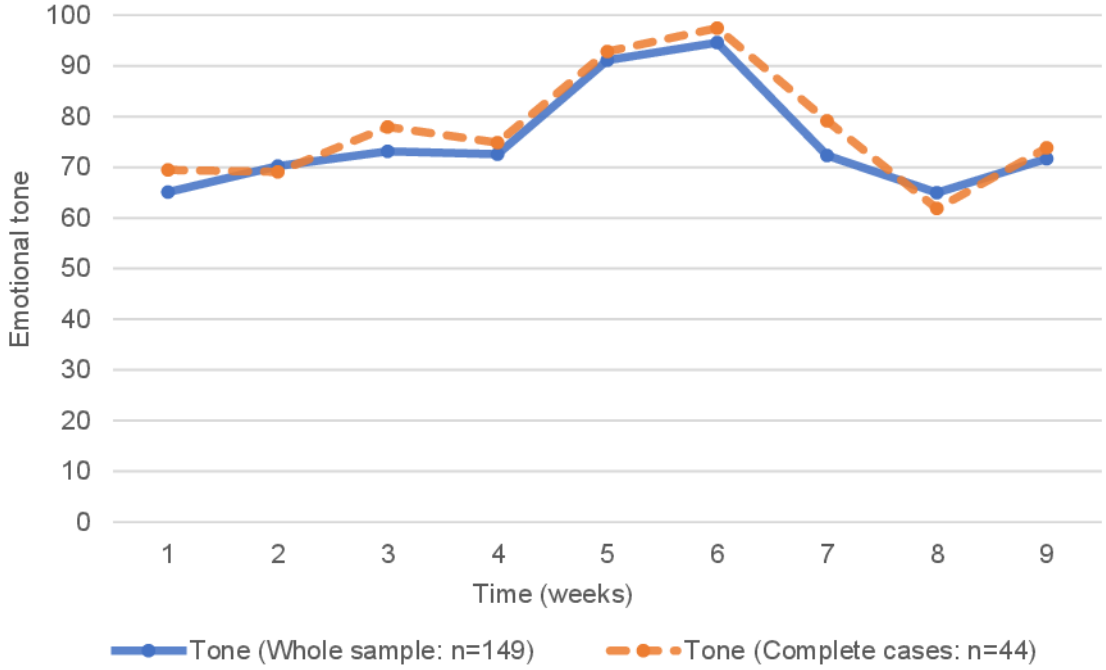
Table 4. Key sample and diary characteristics by diary-compliance subsamples.

Characteristics	Word count (percentile range)			
	265+ (75-100)	163-265 (50-75)	103.5-163 (25-50)	0-103.5 (0-25)
Diary entries, n (%)	280 (24.9)	281 (25.0)	282 (25.1)	281 (25.0)
Age (years), mean (SD)	19.56 (1.56)	19.09 (1.21)	19.07 (1.20)	19.36 (1.64)
Gender (female), n (%)	195 (69.6)	182 (64.8)	178 (63.1)	155 (55.2)
Emotional tone, mean (SD)	72.48 (26.97)	79.33 (26.91)	74.13 (28.50)	74.67 (31.47)

Trajectory of Diary Emotional Tone

The pattern of scores for the available-case analysis and the complete-case analysis followed an inverted U shape (Figure 2), with a peak between weeks 5 and 6 when students are completing the gratitude and signature strengths diaries, respectively. Given the similarity in the trajectory pattern of emotional tone, whether complete case data or all available data were used, subsequent analyses in this paper are presented using the whole data set diary data (available-case analysis).

Figure 2. Trajectory of mean emotional tone within diaries for whole sample (and participants with complete data for all 9 weeks).



Factors Determining the Emotional Tone of Diaries (Model 1)

The results from the random effects generalized least squares regression examining the role of week effects, sociodemographics, diary characteristics, and baseline levels of well-being in determining the emotional tone of diary entries are presented in Table 5. The emotional tone of diary entries in weeks 5 and 6 were, on average, higher by 23.90 (95% CI 16.89-30.90; *P*<.001) and 26.62 (95% CI 19.35-33.88; *P*<.001), respectively, in comparison with the emotional tone of diary entries in week 1. The mean emotional tone did not significantly differ from week 1 for the remaining 6 weeks. As the age of students increased by 1 year, the emotional tone of diary entries was higher, on average, by 1.98 (95% CI 0.51-3.46; *P*=.008). Students with both moderate (5.03, 95% CI 0.08-9.98; *P*=.046) and high (7.48, 95% CI 1.84-13.12; *P*=.009) levels of baseline well-being had diaries with significantly higher emotional tone compared with students with low baseline well-being. Gender and diary characteristics had no clear effects.

Table 5. Model 1: random effects generalized least squares regression examining the role of time (week effects), sociodemographics (age and gender), diary characteristics (word count and entries), and baseline well-being (moderate and high well-being, compared with low well-being) in determining the emotional tone of diary entries.[a]

Covariates	Coefficient	*P* value	95% CI
Time			
Week 1	Reference[b]	N/A[c]	N/A
Week 2	4.57	.19	−2.24 to 11.38
Week 3	4.72	.20	−2.57 to 12.02
Week 4	3.78	.30	−3.33 to 10.89
Week 5	23.90[d]	<.001	16.89 to 30.90
Week 6	26.62[d]	<.001	19.35 to 33.88
Week 7	4.73	.19	−2.38 to 11.84
Week 8	−3.41	.35	−10.62 to 3.79
Week 9	4.56	.23	−2.91 to 12.03
Sociodemographics			
Gender			
Male	Reference	N/A	N/A
Female	−2.44	.34	−7.48 to 2.59
Age	1.98[e]	.008	0.51 to 3.46
Diary characteristics			
Word count	−0.01	.10	−0.02 to 0.00
Total diary entries	0.43	.66	−1.47 to 2.33
Baseline well-being			
Low well-being	Reference	—	—
Moderate well-being	5.03[f]	.046	0.08 to 9.98
High well-being	7.48[e]	.009	1.84 to 13.12

[a]Wald chi-square, X^2_{15}=137.0 (N=855); *P*<.001. This model provided an improved model fit, compared with model 0, which only included time (week effects) presented in Multimedia Appendix 3.

[b]Reference category for factor variables.

[c]N/A: not applicable.

[d]*P*<.001.

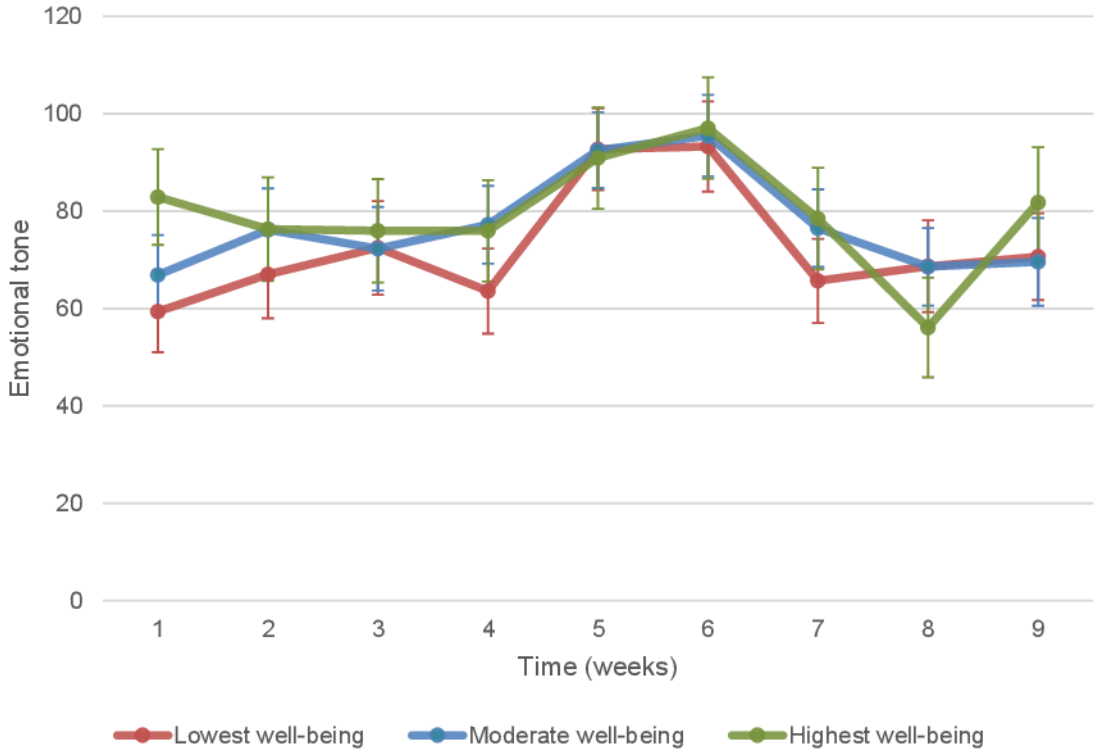
[e]*P*<.01.

[f]*P*<.05.

Differences in Emotional Tone Across Weeks for Students With the Highest and Lowest Levels of Baseline Well-being (Model 2)

As a secondary analysis, we tested whether adding an interaction effect between baseline well-being (tertiles for the lowest, moderate, and highest scores) and time (weeks) improved the fit of the model (Multimedia Appendix 2). Adding this interaction to the model produced a significantly higher Wald chi-squared value (*P*=.03), indicating a better fitting model. To examine how the pattern of emotional tone differed between students with the lowest and highest well-being across the weeks, we plotted the mean emotional tone for students with the highest and lowest levels of well-being (adjusted for age, gender, word count, and total diary entries; Figure 3). The difference in the pattern was most substantial in the first week, where students with the lowest levels of baseline well-being started out with markedly lower average emotional tone of their diary entries. Furthermore, students with the highest levels of baseline well-being demonstrated a marked drop in the emotional tone of their diaries during week 8 when they were asked to write about their goals for the future.

Figure 3. Trajectory of emotional tone for students with low, moderate, and high baseline well-being, adjusted for age, gender, word count, and total diary entries.



Life Domain Analysis (Social, Health, Leisure, Money, and Work; Model 3)

The use of words from 5 life domains (mean percentage of words per diary entry) across the 9 weeks is presented in Figure 4. Social topics were the most discussed topic apart from in week 8 (when participants were asked to diary about their goals) when work was the most dominant life domain discussed. A peak in the discussion of social topics was observed for week 5 (when participants were asked to diary about gratitude). Money was the least discussed topic of diaries for all 9 weeks.

Figure 4. Use of life domain–specific terms in diary entries across weeks.

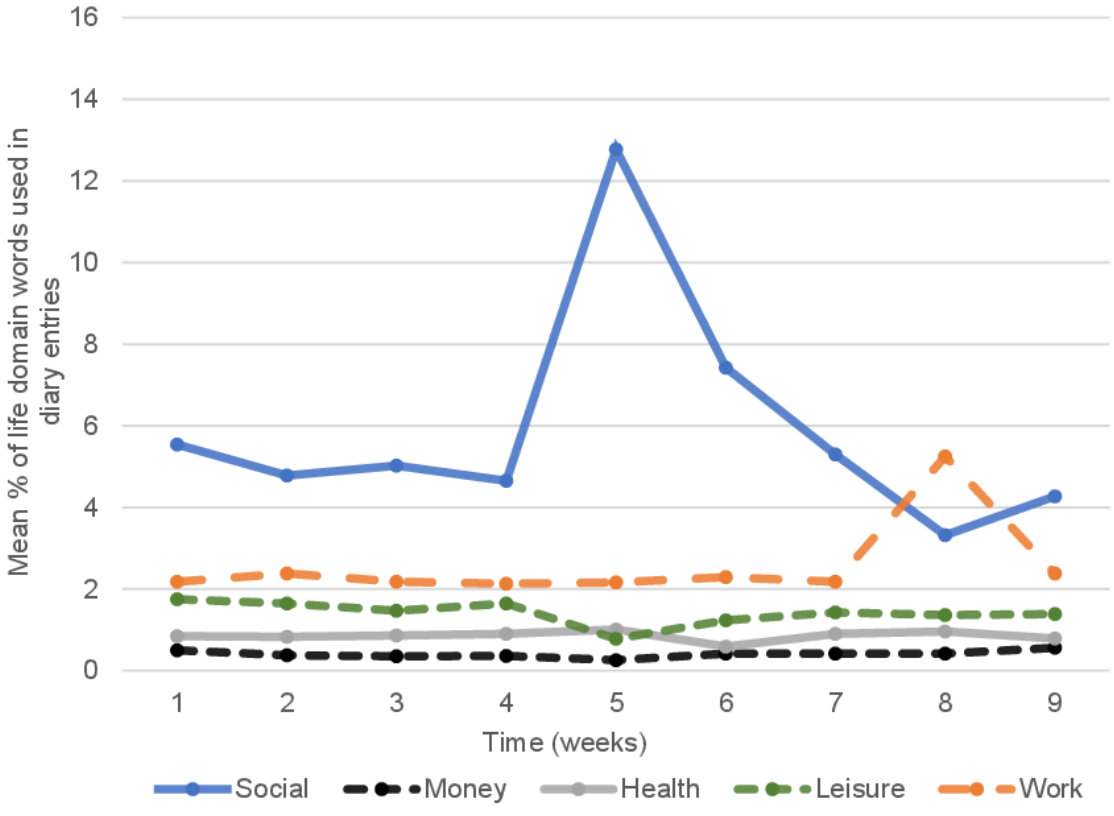


Table 6 presents the results of the random effects generalized least squares regression examining how the use of words in the 5 life domains relates to the emotional tone of diary entries. As the percentage of *social* words in diaries increased by 1, the emotional tone of diaries increased by an average of 0.74 (95% CI 0.21-1.27; *P*=.006). As the percentage of *leisure* words in diaries increased by 1, the emotional tone of diaries increased on average by 3.56 (95% CI 2.28-4.85; *P*<.001). Finally, as the percentage of *health* words in diaries increased by 1, the emotional tone of diaries was more negative by an average of 1.96 (95% CI −3.70 to −0.22; *P*=.03). The relationships between *work*- and *money*-related words and the emotional tone of diary entries were nonsignificant.

Table 6. Model 3: random effects generalized least squares regression examining how the use of words from 5 life domains (social, health, leisure, work, and money) relates to the emotional tone of diary entries, controlling for week effects (time), sociodemographics (age and gender), and diary characteristics.[a]

Covariates	Coefficient	*P* value	95% CI
Time			
Week 1	Reference[b]	N/A[c]	N/A
Week 2	5.61	.08	−0.70 to 11.91
Week 3	6.38	.07	−0.54 to 13.30
Week 4	5.69	.10	−1.05 to 12.44
Week 5	22.32[d]	<.001	14.51 to 30.13
Week 6	27.01[d]	<.001	20.08 to 33.93
Week 7	6.86	.05	0.13 to 13.59
Week 8	1.28	.73	−5.99 to 8.54
Week 9	6.45	.08	−0.66 to 13.57
Sociodemographics			
Gender			
Male	Reference	N/A	N/A
Female	−4.68	.08	−9.91 to 0.54
Age	2.00[e]	.01	0.46 to 3.55
Diary characteristics			
Word count	−0.01	.13	−0.02 to 0.00
Total diary entries	0.32	.74	−1.58 to 2.22
Life domains			
Social	0.74[f]	.006	0.21 to 1.27
Health	−1.96[e]	.027	−3.70 to −0.22
Leisure	3.56[d]	<.001	2.28 to 4.85
Work	−0.16	.73	−1.03 to 0.71
Money	−0.80	.49	−3.07 to 1.46

[a]Wald chi-square, X^2_{18}=189.7 (N=927); *P*<.001.

[b]Reference category for factor variables.

[c]N/A: not applicable.

[d]*P*<.001.

[e]*P*<.05.

[f]*P*<.01.

Discussion

Principal Findings

Compliance with the diary task peaked in week 1 (1041/1124, 92.62%) and the fewest diaries were completed in week 3 (807/1124, 71.81%). Students with the most completed diaries had the highest diary word counts, highest levels of baseline well-being, and on average had diaries with the highest emotional tone. Compared with week 1, diaries were significantly more positive in their emotional tone during weeks

5 and 6 when diary tasks involved writing about gratitude and strengths, respectively. This improvement in emotional tone was not observed at the end of the course, indicating a short-term rather than a lasting improvement in the emotional tone of writing. Higher levels of baseline well-being were associated with more emotionally positive diary entries, and the pattern of emotional tone within diaries across the weeks seen for students with low, moderate, and high baseline well-being was distinct. Diaries predominantly focused on social topics throughout the weeks, and the emotional tone of diaries was positively related to the use of leisure and social words and negatively related to the use of health words.

Existing Literature

The dominant focus on social topics within well-being–oriented diaries reinforces findings from previous research, indicating that a lack of social connectedness predicted higher levels of related mental distress (anxiety and depression) in a cross-sectional study of UK university students [37]. In this study, students with the lowest levels of baseline well-being used on balance more negative words in their diaries (emotional tone), in line with findings linking experiences of depression with the use of negative words [20]. Our findings indicate a strong link between writing about leisure and the use of positive emotion words complements a recent study in China that reported when students were asked to draw their happiest moments, they often depicted leisure activities [38]. Despite links in the literature between financial circumstances and well-being among university students [39,40], our work indicated that in some student samples, money worries may not necessarily factor into the weekly well-being experiences of students.

Implications

The analysis of word use across life domains highlights the importance of social factors and leisure in the lives of university students. As such, attempts to tackle student well-being concerns should continue to experiment with the utility of peer support networks, engaging familial support and responsible use of social media. In terms of the important role leisure plays in student well-being, universities are encouraged to ensure multiple options for leisure are readily available, and students are encouraged to select activities that are personally satisfying [41]. The promotion of leisure among students also has the potential to encourage participation in health-promoting behaviors, such as physical activity [42].

These findings also have implications for the ongoing development of the Science of Happiness course. We present evidence for students' acceptance and willingness to engage in a reasonably rigorous schedule of weekly web-based written diary tasks. It is also noteworthy that one of the weeks with lower compliance to the diary entry task occurred when students were on a break from study (week 3), which provides us important insight into the level of engagement to expect when students are disengaged from university academic activities. This work also provides data-driven guidance as to which students may be more prone to disengage from the task (ie, males and students with lower levels of baseline well-being).

There are wider implications for how student data are used as technology develops. The text analysis methods described in this paper could readily be applied to routinely collected information from students, for example, in written requests for well-being support, to estimate levels of distress based on the use of emotional words. Any developments in this area should reflect on the ethical questions about privacy and student preferences raised by experts working in the area of learning analytics [43]. Students should be involved in these discussions, and work should be undertaken to determine the risks, benefits, and opportunities provided by increased analytic involvement in how students are supported.

Limitations

This study has several limitations. Participants in this study were self-selected; therefore, the respondents in this study may not be representative of the wider student population. Available national data indicate that the UK university population is slightly more female and two-thirds of White ethnicity [44]; however, in our sample, both demographics were overrepresented. This means that the observed gender differences need to be interpreted with caution. Furthermore, the study may not have attracted students who did not believe they have any difficulties with their well-being. As the course was only offered to first-year undergraduate students, these results may not be generalizable to other undergraduate years, postgraduates, and PhD graduate students. Separately, as diary data were collected without the presence of researchers, we were unable to follow up with participants to explore any specific points raised in depth. Finally, without a control group also completing weekly diaries, in this study, we were unable to test whether specifically the diary task had a positive impact on the self-reported well-being of participants. However, on balance students demonstrated a high level of compliance with the diary task, the course generated novel data, and we were able to flexibly apply the LIWC software to analyze the content and underlying emotional tone of the available text data.

Future Research

In this study of undergraduate students, diaries predominantly focused on social topics; however, future research could examine whether different patterns, such as a focus on money or health, are observed in nonstudent samples of young adults. This study could build on attempts to investigate how distinct the challenges experienced by students are to university populations [45,46]. Separately, it would be valuable to investigate whether money was more of a focus within the diaries of subgroups with different financial circumstances [47]. Although leisure was discussed less frequently than social topics, the use of leisure words was related to the emotional positivity of diary entries. Building on work that has theorized about the many dimensions of leisure in university settings [48], it would be valuable to conduct more in-depth research with students to examine how and when different forms of leisure contribute to their well-being.

Given the noticeable spike in emotional tone observed when students were tasked with writing diaries about gratitude, a future iteration of the course could focus solely on this subject [49]. Future research could examine the sources of gratitude for

university students and explore whether repeated engagement in this task has a positive and sustained impact on the emotional positivity of writing and self-reported well-being. The more we understand about the relationship between what students write and how they feel, the better informed we will be to decide how far the use of these methods should be extended. We also recognize that a future analysis of unsolicited diaries would enrich our understanding of how diaries are used in nonexperimental settings.

One explanation for why compliance with the diary task fell to 72.5% (108/149) in week 9 is that this time point marked the conclusion of the course. Exploring this phenomenon and broader motivations for participation in the course would expand our understanding of how the course is interpreted and experienced. In future research, it would be informative to examine the acceptability of automated digital prompts designed to encourage the use of the web-based diaries, especially for participants who are willing yet have simply forgotten to complete the task.

Conclusions

This study demonstrated the informative power of web-based diaries, the flexibility of computerized text analysis methods, and differential experiences of students with varying levels of baseline well-being engaged in psychoeducation. Students used their diaries with a high level of compliance and wrote with the highest proportion of positive emotion words during weeks where diaries focused on gratitude and strengths. Further research is needed to explore the importance of leisure to well-being the longer-term impact of diaries on well-being, and suggestions are provided for how the science of happiness could be adapted in the future. We present support for previous studies highlighting the importance of social factors and leisure for student well-being, and echo recommendations that universities should ensure these activities are facilitated and encouraged.

Acknowledgments

The authors would like to thank Prof Laurie Santos for developing the original course on which the Science of Happiness course was based. This work was supported by the Elizabeth Blackwell Institute, University of Bristol, and the Development and Alumni Relations Office, University of Bristol.

Authors' Contributions

MAL designed the study with the support of all the coauthors. SJ and BH collected data. MAL performed the data analysis using RM. MAL drafted the manuscript, and all coauthors helped edit the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Diary tasks for each week of the study.
[DOCX File , 15 KB - jmir_v23i4e25279_app1.docx]

Multimedia Appendix 2
Model 2: random effects generalized least squares regression examining interaction effects between time and baseline well-being levels, controlling for sociodemographics and diary characteristics.
[DOCX File , 22 KB - jmir_v23i4e25279_app2.docx]

Multimedia Appendix 3
Model 0: random effects generalized least squares regression examining only week effects among samples with complete-case data in model 1.
[DOCX File , 17 KB - jmir_v23i4e25279_app3.docx]

References

1. Pitchforth J, Fahy K, Ford T, Wolpert M, Viner RM, Hargreaves DS. Mental health and well-being trends among children and young people in the UK, 1995–2014: analysis of repeated cross-sectional national health surveys. Psychol Med 2018 Sep 11;49(08):1275-1285. [doi: 10.1017/s0033291718001757]
2. Marginson S. The worldwide trend to high participation higher education: dynamics of social stratification in inclusive systems. High Educ 2016 Jun 2;72(4):413-434. [doi: 10.1007/s10734-016-0016-x]
3. Fernandez A, Howse E, Rubio-Valera M, Thorncraft K, Noone J, Luu X, et al. Setting-based interventions to promote mental health at the university: a systematic review. Int J Public Health 2016 Sep 30;61(7):797-807. [doi: 10.1007/s00038-016-0846-4] [Medline: 27364779]

4. Worsley J, Pennington A, Corcoran J. What interventions improve college and university students' mental health and wellbeing?: a review of review-level evidence. National Grey Literature Collection. 2020. URL: http://allcatsrgrey.org.uk/wp/wpfb-file/student-mental-health-full-review-pdf/ [accessed 2020-06-19]
5. Kraft D. One hundred years of college mental health. J Am Coll Health 2011;59(6):477-481. [doi: 10.1080/07448481.2011.569964] [Medline: 21660801]
6. Regehr C, Glancy D, Pitts A. Interventions to reduce stress in university students: a review and meta-analysis. J Affect Disord 2013 May 15;148(1):1-11. [doi: 10.1016/j.jad.2012.11.026] [Medline: 23246209]
7. Xiao H, Carney DM, Youn SJ, Janis RA, Castonguay LG, Hayes JA, et al. Are we in crisis? National mental health and treatment trends in college counseling centers. Psychol Serv 2017 Nov;14(4):407-415. [doi: 10.1037/ser0000130] [Medline: 29120199]
8. Arnett JJ, Žukauskienė R, Sugimura K. The new life stage of emerging adulthood at ages 18–29 years: implications for mental health. Lancet Psychiatry 2014 Dec;1(7):569-576. [doi: 10.1016/s2215-0366(14)00080-7]
9. Howard AL, Galambos NL, Krahn HJ. Paths to success in young adulthood from mental health and life transitions in emerging adulthood. Int J Behav Dev 2010 Jul 05;34(6):538-546. [doi: 10.1177/0165025410365803]
10. Auerbach RP, Mortier P, Bruffaerts R, Alonso J, Benjet C, Cuijpers P, WHO WMH-ICS Collaborators. WHO World Mental Health Surveys International College Student Project: Prevalence and distribution of mental disorders. J Abnorm Psychol 2018 Oct 13;127(7):623-638 [FREE Full text] [doi: 10.1037/abn0000362] [Medline: 30211576]
11. Broglia E, Millings A, Barkham M. Challenges to addressing student mental health in embedded counselling services: a survey of UK higher and further education institutions. Br J Guid Counsell 2017 Sep 29;46(4):441-455. [doi: 10.1080/03069885.2017.1370695]
12. Jones RB, Thapar A, Rice F, Beeching H, Cichosz R, Mars B, et al. A web-based psychoeducational intervention for adolescent depression: design and development of MoodHwb. JMIR Ment Health 2018 Feb 15;5(1):e13 [FREE Full text] [doi: 10.2196/mental.8894] [Medline: 29449202]
13. Brown JS. Student mental health: some answers and more questions. J Ment Health 2018 Jun 16;27(3):193-196. [doi: 10.1080/09638237.2018.1470319] [Medline: 29768071]
14. Lambert L, Passmore H, Joshanloo M. A positive psychology intervention program in a culturally-diverse university: boosting happiness and reducing fear. J Happiness Stud 2018 May 11;20(4):1141-1162. [doi: 10.1007/s10902-018-9993-z]
15. Smith BW, Ford CG, Erickson K, Guzman A. The effects of a character strength focused positive psychology course on undergraduate happiness and well-being. J Happiness Stud 2020 Feb 18;22(1):343-362. [doi: 10.1007/s10902-020-00233-9]
16. Thiele C, Laireiter A, Baumann U. Diaries in clinical psychology and psychotherapy: a selective review. Clin Psychol Psychother 2002 Jan;9(1):1-37. [doi: 10.1002/cpp.302]
17. Bolger N, Davis A, Rafaeli E. Diary methods: capturing life as it is lived. Annu Rev Psychol 2003 Feb;54(1):579-616. [doi: 10.1146/annurev.psych.54.101601.145030] [Medline: 12499517]
18. Waldron I, Eyer J. Socioeconomic causes of the recent rise in death rates for 15–24-yr-olds. Soc Sci Med (1967) 1975 Jul;9(7):383-396. [doi: 10.1016/0037-7856(75)90139-0]
19. Burton CM, King LA. The health benefits of writing about intensely positive experiences. J Res Pers 2004 Apr;38(2):150-163. [doi: 10.1016/s0092-6566(03)00058-8]
20. Rude S, Gortner E, Pennebaker J. Language use of depressed and depression-vulnerable college students. Cogn Emot 2004 Dec;18(8):1121-1133. [doi: 10.1080/02699930441000030]
21. Veilleux JC, Lankford NM, Hill MA, Skinner KD, Chamberlain KD, Baker DE, et al. Affect balance predicts daily emotional experience. Pers Individ Differ 2020 Feb;154. [doi: 10.1016/j.paid.2019.109683]
22. Pennebaker JW, Boyd RL, Jordan KN, Blackburn K. The development and psychometric properties of LIWC2015. The University of Texas at Austin 2015 [FREE Full text] [doi: 10.15781/T29G6Z]
23. Stone AA, Wen CKF, Schneider S, Junghaenel DU. Evaluating the effect of daily diary instructional phrases on respondents' recall time frames: survey experiment. J Med Internet Res 2020 Feb 21;22(2):16105 [FREE Full text] [doi: 10.2196/16105] [Medline: 32130129]
24. Fraia MD, Tripodi S, Arasi S, Dramburg S, Castelli S, Villalta D, et al. Adherence to prescribed e-diary recording by patients with seasonal allergic rhinitis: observational study. J Med Internet Res 2020 Mar 16;22(3):16642 [FREE Full text] [doi: 10.2196/16642] [Medline: 32175909]
25. Metsäranta K, Kurki M, Valimaki M, Anttila M. How do adolescents use electronic diaries? A mixed-methods study among adolescents with depressive symptoms. J Med Internet Res 2019 Feb 20;21(2):11711 [FREE Full text] [doi: 10.2196/11711] [Medline: 30785408]
26. Funk B, Sadeh-Sharvit S, Fitzsimmons-Craft EE, Trockel MT, Monterubio GE, Goel NJ, et al. A framework for applying natural language processing in digital health interventions. J Med Internet Res 2020 Feb 19;22(2):13855 [FREE Full text] [doi: 10.2196/13855] [Medline: 32130118]
27. Seabrook EM, Kern ML, Fulcher BD, Rickard NS. Predicting depression from language-based emotion dynamics: longitudinal analysis of facebook and twitter status updates. J Med Internet Res 2018 May 08;20(5):e168 [FREE Full text] [doi: 10.2196/jmir.9267] [Medline: 29739736]

28. Hood B, Jelbert S, Santos LR. Benefits of a psychoeducational happiness course on university student mental well-being both before and during a COVID-19 lockdown. Health Psychol Open 2021;8(1) [FREE Full text] [doi: 10.1177/2055102921999291] [Medline: 33796324]
29. Alaszewski A. Using diaries for social research. Sage Research Methods 2006. [doi: 10.4135/9780857020215]
30. Stewart-Brown S, Tennant A, Tennant R, Platt S, Parkinson J, Weich S. Internal construct validity of the Warwick-Edinburgh Mental Well-being Scale (WEMWBS): a Rasch analysis using data from the Scottish Health Education Population Survey. Health Qual Life Outcomes 2009 Feb 19;7(1):15 [FREE Full text] [doi: 10.1186/1477-7525-7-15] [Medline: 19228398]
31. Tennant R, Hiller L, Fishwick R, Platt S, Joseph S, Weich S, et al. The Warwick-Edinburgh Mental Well-being Scale (WEMWBS): development and UK validation. Health Qual Life Outcomes 2007 Nov 27;5(1):63 [FREE Full text] [doi: 10.1186/1477-7525-5-63] [Medline: 18042300]
32. Fat LN, Scholes S, Boniface S, Mindell J, Stewart-Brown S. Evaluating and establishing national norms for mental wellbeing using the short Warwick-Edinburgh Mental Well-being Scale (SWEMWBS): findings from the Health Survey for England. Qual Life Res 2017 May 16;26(5):1129-1144 [FREE Full text] [doi: 10.1007/s11136-016-1454-8] [Medline: 27853963]
33. Boumans JW, Trilling D. Taking stock of the toolkit. Digit Journal 2015 Nov 03;4(1):8-23. [doi: 10.1080/21670811.2015.1096598]
34. Resources. Language Use and Social Interaction. 2020. URL: http://www.depts.ttu.edu/psy/lusi/resources.php [accessed 2020-01-21]
35. Tausczik YR, Pennebaker JW. The psychological meaning of words: LIWC and computerized text analysis methods. J Lang Soc Psychol 2009 Dec 08;29(1):24-54. [doi: 10.1177/0261927x09351676]
36. StataCorp. Stata Statistical Software: Release 16. 2019. Texas: StataCorp LLC; 2019. URL: https://www.stata.com/company/ [accessed 2021-03-19]
37. McIntyre JC, Worsley J, Corcoran R, Woods PH, Bentall RP. Academic and non-academic predictors of student psychological distress: the role of social identity and loneliness. J Ment Health 2018 Jun 13;27(3):230-239. [doi: 10.1080/09638237.2018.1437608] [Medline: 29436883]
38. Liu H, Da S. The relationships between leisure and happiness-a graphic elicitation method. Leis Studies 2019 Feb 08;39(1):111-130. [doi: 10.1080/02614367.2019.1575459]
39. Archuleta KL, Dale A, Spann SM. College students and financial distressxploring debt, financial satisfaction, and financial anxiety. J Financial Couns Plan 2013;24(2):50-62. [doi: 10.1037/t13109-000]
40. Montalto CP, Phillips EL, McDaniel A, Baker AR. College student financial wellness: student loans and beyond. J Fam Econ Iss 2018 Sep 24;40(1):3-21. [doi: 10.1007/s10834-018-9593-4]
41. Kuykendall L, Tay L, Ng V. Leisure engagement and subjective well-being: A meta-analysis. Psychol Bull 2015 Mar;141(2):364-403. [doi: 10.1037/a0038508] [Medline: 25602273]
42. Hartman CL, Barcelona RJ, Trauntvein NE, Hall SL. Well-being and leisure-time physical activity psychosocial factors predict physical activity among university students. Leis Studies 2019 Sep 26;39(1):156-164. [doi: 10.1080/02614367.2019.1670722]
43. Rubel A, Jones KM. Student privacy in learning analytics: an information ethics perspective. The Information Society 2016 Feb 05;32(2):143-159. [doi: 10.1080/01972243.2016.1130502]
44. HESA. Higher Education Student Statistics: UK, 2018/19. 2018. URL: https://www.hesa.ac.uk/news/16-01-2020/sb255-higher-education-student-statistics [accessed 2020-04-20]
45. Kovess-Masfety V, Leray E, Denis L, Husky M, Pitrou I, Bodeau-Livinec F. Mental health of college students and their non-college-attending peers: results from a large French cross-sectional survey. BMC Psychol 2016 Apr 21;4(1):20 [FREE Full text] [doi: 10.1186/s40359-016-0124-5] [Medline: 27098291]
46. Stallman HM. Psychological distress in university students: a comparison with general population data. Aust Psychol 2010 Nov 06;45(4):249-257. [doi: 10.1080/00050067.2010.482109]
47. Benson-Egglenton J. The financial circumstances associated with high and low wellbeing in undergraduate students: a case study of an English Russell Group institution. J Furth High Educ 2018 Jan 16;43(7):901-913. [doi: 10.1080/0309877x.2017.1421621]
48. Iwasaki Y, Mannell RC. Hierarchical dimensions of leisure stress coping. Leis Sci 2010 Oct 29;22(3):163-181. [doi: 10.1080/01490409950121843]
49. Toepfer S, Walker K. Letters of gratitude: improving well-being through expressive writing. J Writing Res 2009 Nov;1(3):181-198. [doi: 10.17239/jowr-2009.01.03.1]

Abbreviations

LIWC: Linguistic Inquiry and Word Count
SWEMWBS: Short Warwick-Edinburgh Mental Wellbeing Scale

Edited by R Kukafka; submitted 26.10.20; peer-reviewed by M Afzal, M Mulvenna; comments to author 28.12.20; revised version received 21.02.21; accepted 16.03.21; published 22.04.21.

Please cite as:
Linton MJA, Jelbert S, Kidger J, Morris R, Biddle L, Hood B
Investigating the Use of Electronic Well-being Diaries Completed Within a Psychoeducation Program for University Students: Longitudinal Text Analysis Study
J Med Internet Res 2021;23(4):e25279
URL: https://www.jmir.org/2021/4/e25279
doi:10.2196/25279
PMID:33885373

Original Paper

Rural Telemedicine Use Before and During the COVID-19 Pandemic: Repeated Cross-sectional Study

Cherry Chu[1], MSc; Peter Cram[2,3], MD, MBA; Andrea Pang[4], MPH; Vess Stamenova[1], PhD; Mina Tadrous[1,5], PharmD, PhD; R Sacha Bhatia[1,4], MD, MBA

[1]Women's College Hospital Institute for Health System Solutions and Virtual Care, Women's College Hospital, Toronto, ON, Canada

[2]Division of General Internal Medicine and Geriatrics, Sinai Health System and University Health Network, Toronto, ON, Canada

[3]Department of Medicine, University of Toronto, Toronto, ON, Canada

[4]Institute for Clinical Evaluative Sciences, Toronto, ON, Canada

[5]Leslie Dan Faculty of Pharmacy, University of Toronto, Toronto, ON, Canada

Corresponding Author:
R Sacha Bhatia, MD, MBA
Women's College Hospital Institute for Health System Solutions and Virtual Care
Women's College Hospital
76 Grenville St
Toronto, ON, M5S 1B2
Canada
Phone: 1 416 312 9147
Email: sacha.bhatia@wchospital.ca

Abstract

Background: The COVID-19 pandemic has led to a notable increase in telemedicine adoption. However, the impact of the pandemic on telemedicine use at a population level in rural and remote settings remains unclear.

Objective: This study aimed to evaluate changes in the rate of telemedicine use among rural populations and identify patient characteristics associated with telemedicine use prior to and during the pandemic.

Methods: We conducted a repeated cross-sectional study on all monthly and quarterly rural telemedicine visits from January 2012 to June 2020, using administrative data from Ontario, Canada. We compared the changes in telemedicine use among residents of rural and urban regions of Ontario prior to and during the pandemic.

Results: Before the pandemic, telemedicine use was steadily low in 2012-2019 for both rural and urban populations but slightly higher overall for rural patients (11 visits per 1000 patients vs 7 visits per 1000 patients in December 2019, *P*<.001). The rate of telemedicine visits among rural patients significantly increased to 147 visits per 1000 patients in June 2020. A similar but steeper increase (*P*=.15) was observed among urban patients (220 visits per 1000 urban patients). Telemedicine use increased across all age groups, with the highest rates reported among older adults aged ≥65 years (77 visits per 100 patients in 2020). The proportions of patients with at least 1 telemedicine visit were similar across the adult age groups (n=82,246/290,401, 28.3% for patients aged 18-49 years, n=79,339/290,401, 27.3% for patients aged 50-64 years, and n=80,833/290,401, 27.8% for patients aged 65-79 years), but lower among younger patients <18 years (n=23,699/290,401, 8.2%) and older patients ≥80 years (n=24,284/290,401, 8.4%) in 2020 (*P*<.001). There were more female users than male users of telemedicine (n=158,643/290,401, 54.6% vs n=131,758/290,401, 45.4%, respectively, in 2020; *P*<.001). There was a significantly higher proportion of telemedicine users residing in relatively less rural than in more rural regions (n=261,814/290,401, 90.2% vs n=28,587/290,401, 9.8%, respectively, in 2020; *P*<.001).

Conclusions: Telemedicine adoption increased in rural and remote areas during the COVID-19 pandemic, but its use increased in urban and less rural populations. Future studies should investigate the potential barriers to telemedicine use among rural patients and the impact of rural telemedicine on patient health care utilization and outcomes.

(J Med Internet Res 2021;23(4):e26960) doi:10.2196/26960

KEYWORDS

chronic disease; chronic illness; COVID-19; health care; health services; older adults; remote; rural; pandemic; population; telemedicine; virtual care

Introduction

Access to health care services in rural communities has proven challenging among individuals for many reasons. There is a scarcity of health care professionals serving rural populations, and consequently, rural residents face barriers to health care access, such as long travel times, limited public transportation and associated costs, and time away from work. Ontario is Canada's most populous province with over 14,000,000 people, and, at >1,000,000 km^2, it has a larger area than Texas and Montana combined. While most of Ontario's population is situated in the southern region of the province in urban areas, approximately 10% of the population resides in rural areas [1], including the northern regions of the province. The Ontario Telemedicine Network (OTN) is an organization funded by the provincial government, which aimed to initially increase health care access among the rural population by facilitating 2-way videoconferencing between patients and care providers [2]. However, prior to the COVID-19 pandemic, telemedicine adoption was modest in Ontario [3], in part owing to strict requirements for both the patient and the physician, such as attending a virtual visit at an OTN facility, and using an OTN-approved communication platform [4]. A prepandemic study from Ontario revealed that Northern Ontario, a predominantly rural region in the province, saw markedly higher annual rates of telemedicine use than the mostly urban Southern Ontario, and even within Northern Ontario, rates of telemedicine use were higher in the "very rural" than in the "less rural" populations [5].

The COVID-19 pandemic has posed a unique challenge for rural patients—particularly those who are more susceptible to the disease—such as older adults and those with chronic diseases. In particular, there has been a need to balance chronic disease management with the risk of virus transmission, either during health care consultations or when traveling from rural to urban areas [6]. In order to reduce virus transmission in the health care setting, the Ontario Health Insurance Plan (OHIP)—the government health insurer across the province—introduced temporary billing codes that broadened the range of modalities eligible for reimbursement beyond OTN-approved videoconferencing platforms, including telephone calls and many other types of technology [7]. While a limited number of studies have reported substantial increases in the use of telemedicine during the pandemic, there are currently no published data on the use of telemedicine in rural communities or on the potential differences in telemedicine uptake between rural and urban populations during the pandemic. An understanding of rural telemedicine uptake and how it has been affected by the COVID-19 pandemic can provide insight into the utility of telemedicine in rural areas, particularly among vulnerable and at-risk populations; moreover, this information can contribute to the development of targeted approaches to increase telemedicine access where uptake was limited.

This study aimed to measure the utilization of telemedicine in rural and urban populations and among at-risk patient groups in rural Ontario prior to and during the COVID-19 pandemic, and to describe the characteristics of rural patients who have used telemedicine.

Methods

Study Design and Data Sources

We conducted a population-based, repeated cross-sectional study of monthly and quarterly rural ambulatory telemedicine visits in Ontario, Canada, beginning prior to the COVID-19 pandemic (January 1, 2012) and extending to June 30, 2020, using data from the following administrative databases: (1) OHIP, which records all health services delivered by physicians to Ontario residents and (2) Registered Persons Database, which contains demographic information of all patients covered under OHIP. We linked each patient's postal code of residence to census-based neighborhood income measures and then stratified patients across the province into neighborhood income quintiles. To understand telemedicine uptake among patients with various chronic conditions, we defined individual patient cohorts on the basis of select chronic ambulatory care conditions, using the Discharge Abstract Database, which records all in-patient hospital admissions, the National Ambulatory Care Reporting System, which contains data on all hospital- and community-based ambulatory care (including emergency department visits), and various Institute for Clinical Evaluative Sciences (ICES)–validated disease-specific registries. ICES is an independent, nonprofit research institute whose legal status under Ontario's health information privacy law allows it to collect and analyze health care and demographic data without consent for health system evaluation and improvement. Databases were linked using unique encoded identifiers and analyzed at the ICES. The use of these databases for the purpose of this study was authorized under clause 45 of Ontario's Personal Health Information Protection Act [8], which does not require review by a research ethics board.

Population

We identified all patients residing in the rural regions of Ontario by using the rurality index for Ontario (RIO) and their ambulatory visits through telemedicine through relevant physician billing codes (Multimedia Appendix 1). The RIO score has been established by the Ministry of Health and Long-Term Care as a method to fairly and consistently measure a community's degree of rurality based on its postal code in order to allocate governmental funding for rural health programs. The RIO score is derived from three measures: population density, distance from a basic referral center, and distance from an advanced referral center [9]. A higher RIO score indicates a higher degree of rurality, with a maximum score of 100. Rural residents have been defined as patients with a RIO score of ≥40, which is a standard cut-off used by the provincial government [10]. We compared the rates of rural telemedicine visits to those of urban visits, the latter defined with a RIO score of <40. We excluded claims for any patient who was a nonresident of Ontario or had an invalid or missing health card number.

We further identified patients diagnosed with chronic disease including chronic obstructive pulmonary disease, congestive heart failure, asthma, hypertension, and diabetes mellitus from the existing established ICES registries and algorithms [11].

Patients with serious mental illness were identified from at least 2 outpatient or 1 inpatient claims with the corresponding International Classification of Diseases, 9th revision or 10th revision codes for schizophrenia, psychotic disorders, or bipolar disorder in the past 12 months. Patients with angina were identified by at least one emergency department visit with the aforementioned relevant disease codes in the past 12 months (Multimedia Appendix 1).

Statistical Analysis

For each month in our study period (January 1, 2012, to June 30, 2020), we reported the rate of telemedicine visits per 1000 patients in rural Ontario. Given the markedly high population of older adults in rural Ontario [12], we also calculated quarterly rates separately among patients aged ≥65 years and compared them with those of other age groups (visits per 100 patients). Chi-square tests were conducted to assess the distribution of various characteristics of patients who have used telemedicine (age category, sex, region of residence, neighborhood income quintile, level of rurality, and chronic disease diagnosis) across 2012, 2016, and 2020. We selected 2012 as the first year of the study period, 2016 as the middle period, and 2020 as the pandemic period. We also compared quarterly telemedicine utilization across chronic disease subgroups (visits per 100 patients). A *P* value less than .05 was considered significant. All analyses were performed using SAS (version 9.4, The SAS Institute).

Results

We compared the characteristics of rural patients who used telemedicine across the calendar years of 2012 and 2016 and the first half of 2020. The number of rural patients who received at least one telemedicine visit increased from 2012 (n=14,666/1,017,546, 1.4%) to 2020 (n=290,401/1,033,271, 28.1%). Prior to the pandemic, a greater proportion of adults aged 18-49 years used telemedicine (n=4965/14,666, 33.9% in 2012) compared to the other age groups (n=4140/14,666, 28.2% for patients aged 50-64 years and n=3716/14,666, 25.3% for patients aged 65-79 years), but after the onset of the pandemic, we observed similar proportions of telemedicine users among the older age groups as well (n=82,246/290,401, 28.3% for patients aged 18-49 years; n=79,339/290,401, 27.3% for patients aged 50-64 years, and n=80,833/290,401, 27.8% for patients aged 65-79 years). In 2012 and 2016, there was an approximately equal proportion of male and female telemedicine users; however, in 2020, there were more female users than male users (n=158,643/290,401, 54.6% vs n=131,758, 45.4%, respectively). Prior to the COVID-19 pandemic, most users were based in Northern Ontario (n=10,431/14,666, 71.1% in 2012 and n=15,702/27,145, 57.8% in 2016), but after the onset of the pandemic, there was a more balanced distribution of users across all regions, with the highest proportion of users in Eastern Ontario (n=98,008/290,401, 33.7% in 2020). These results are summarized in Table 1. Similarly, in 2012, the highest rate of telemedicine visits was observed among patients residing in Northern Ontario; however, the Central and Eastern regions recorded the highest rates in 2020 (Multimedia Appendix 2). Smaller shifts were reported in telemedicine use across income quintiles, with more patients in the lower-income quintiles using telemedicine in 2012 (n=3270/14,666, 22.3% for quintile 1 and n=3013/14,666, 20.5% for quintile 2) and 2016 (n=6012/27,145, 22.1% for quintile 1 and n=6770/27,145, 24.9% for quintile 2) but more users in the middle-income quintiles in 2020 (n=67,172/290,401, 23.1% in quintile 2 and n=65,128/290,401, 22.4% in quintile 3). In 2012, we observed a higher proportion of telemedicine users living in less rural areas (lower RIO score) than in more rural areas (n=8709/14,666, 59.4% vs n=5957/14,666, 40.6%, respectively). This difference further increased across the years, with the largest difference observed in 2020 (n=261,814/290,401, 90.2% vs n=28,587/290,401, 9.8%, respectively). While the absolute number of telemedicine users with chronic disease has increased significantly over the years, for most chronic disease groups, the proportion of users with the disease appeared to decrease slightly (eg, 1041 of 14,666 [7.1%] patients had heart failure in 2012 as opposed to 12,580 of 290,401 [4.3%] in 2020; *P*<.001). Among the health conditions of interest in this study, hypertension was the most frequently diagnosed condition (n=112,297/290,401, 38.7% in 2020), followed by diabetes (n=52,830/290,401, 18.2% in 2020).

We observed a significant increase in the rate of telemedicine visits among both rural and urban patients after the start of the COVID-19 pandemic. Prior to the pandemic, this rate was consistently higher among rural patients than among urban patients (11 visits per 1000 patients vs 7 visits per 1000 patients in December 2019, respectively; *P*<.001). During the pandemic, both urban and rural telemedicine usage rates increased significantly, but utilization rates among rural patients were lower than those among urban patients (147 visits per 1000 patients vs 220 visits per 1000 patients in June 2020, respectively; *P*=.15). These findings are illustrated in Figure 1.

Among physicians who provided telemedicine consultations to rural patients, there was a large shift from mostly in-person visits in 2012 and 2016 to telemedicine visits in 2020. The highest proportion of telemedicine providers in 2020 were from family or general practice (n=9044/17,601, 51.4%) (Multimedia Appendix 3).

The rate of telemedicine visits increased significantly from before to during the COVID-19 pandemic among patients with various chronic disease conditions, with the highest rates observed among patients in the mental illness subgroup (126 visits per 100 patients in in second quarter of 2020), followed by those in the congestive heart failure (116 visits per 100 patients), chronic obstructive pulmonary disease (110 visits per 100 patients), angina (96 visits per 100 patients), diabetes (92 visits per 100 patients), hypertension (80 visits per 100 patients), and asthma (65 visits per 100 patients) subgroups. These trends are shown in Figure 2.

Prior to the pandemic, the rates of telemedicine visits were low and stable; however, the onset of the pandemic led to a surge in telemedicine use across all age groups. The highest rates were reported among adults aged ≥65 years, approaching 77 visits per 100 patients in the second quarter of 2020 as opposed to 4 visits per 100 patients in the fourth quarter of 2019. The next highest telemedicine usage rates during the pandemic was observed among patients aged 51-64 years (53 visits per 100

patients), followed by those aged 31-50 years (42 visits per 100 patients), and those aged 18-30 years (32 visits per 100 patients). The lowest rates of telemedicine use were observed among the youngest group of patients (aged 0-17 years, 16 visits per 100 patients). These findings are illustrated in Figure 3.

Table 1. Characteristics of rural patients who received at least 1 telemedicine visit in 2012, 2016, and 2020.

Variables	Number of patients, n (%)			*P* value
	2012 (n=14,666)	2016 (n=27,145)	2020 (n=290,401)	
Age group (years)				<.001
<18	565 (3.9)	1011 (3.7)	23,699 (8.2)	
18-49	4965 (33.9)	9146 (33.7)	82,246 (28.3)	
50-64	4140 (28.2)	7917 (29.2)	79,339 (27.3)	
65-79	3716 (25.3)	6841 (25.2)	80,833 (27.8)	
≥80	1280 (8.7)	2230 (8.2)	24,284 (8.4)	
Sex				<.001
Female	7345 (50.1)	13,343 (49.2)	158,643 (54.6)	
Male	7321 (49.9)	13,802 (50.8)	131,758 (45.4)	
Region of Ontario				<.001
Central	775 (5.3)	1870 (6.9)	48,526 (16.7)	
East	2886 (19.7)	7042 (25.9)	98,008 (33.7)	
North	10,431 (71.1)	15,702 (57.8)	69,552 (24.0)	
West	574 (3.9)	2531 (9.3)	74,315 (25.6)	
Income quintile				<.001
1 (lowest)	3270 (22.3)	6012 (22.1)	53,448 (18.4)	
2	3013 (20.5)	6770 (24.9)	67,172 (23.1)	
3	2868 (19.6)	5698 (21.0)	65,128 (22.4)	
4	2482 (16.9)	4713 (17.4)	59,767 (20.6)	
5 (Highest)	2965 (20.2)	3952 (14.6)	44,885 (15.5)	
Level of rurality (rurality index of Ontario score)				<.001
40-75 (less rural)	8709 (59.4)	19,212 (70.8)	261,814 (90.2)	
76-100 (more rural)	5957 (40.6)	7933 (29.2)	28,587 (9.8)	
Chronic disease				
Hypertension	6239 (42.5)	11,300 (41.6)	112,297 (38.7)	<.001
Diabetes	3451 (23.5)	5920 (21.8)	52,830 (18.2)	<.001
Chronic obstructive pulmonary disease	1371 (9.3)	2600 (9.6)	16,941 (5.8)	<.001
Congestive heart failure	1041 (7.1)	1957 (7.2)	12,580 (4.3)	<.001
Asthma	1584 (10.8)	3328 (12.3)	33,223 (11.4)	<.001
Angina	679 (4.6)	1239 (4.6)	8720 (3.0)	<.001
Mental illness	505 (3.4)	1011 (3.7)	4855 (1.7)	<.001

Figure 1. Rate of telemedicine visits per 1000 eligible patients in Ontario by rurality, 2012-2020.

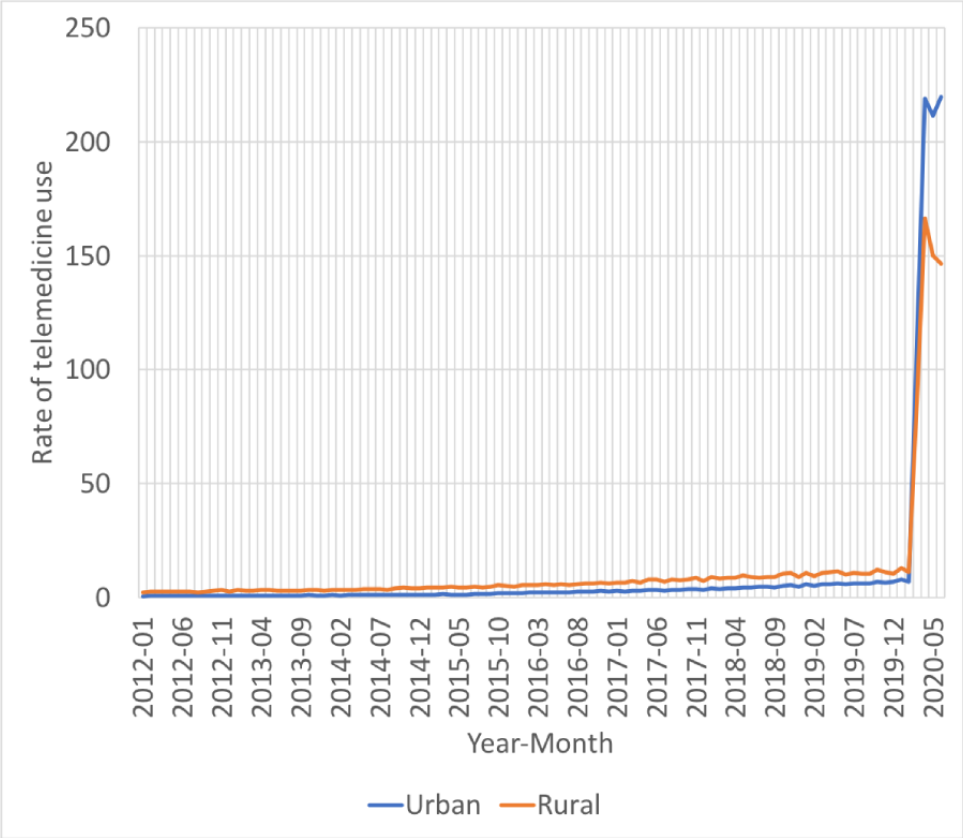


Figure 2. Rate of telemedicine visits per 100 rural patients by chronic disease, 2012-2020. COPD: chronic obstructive pulmonary disease.

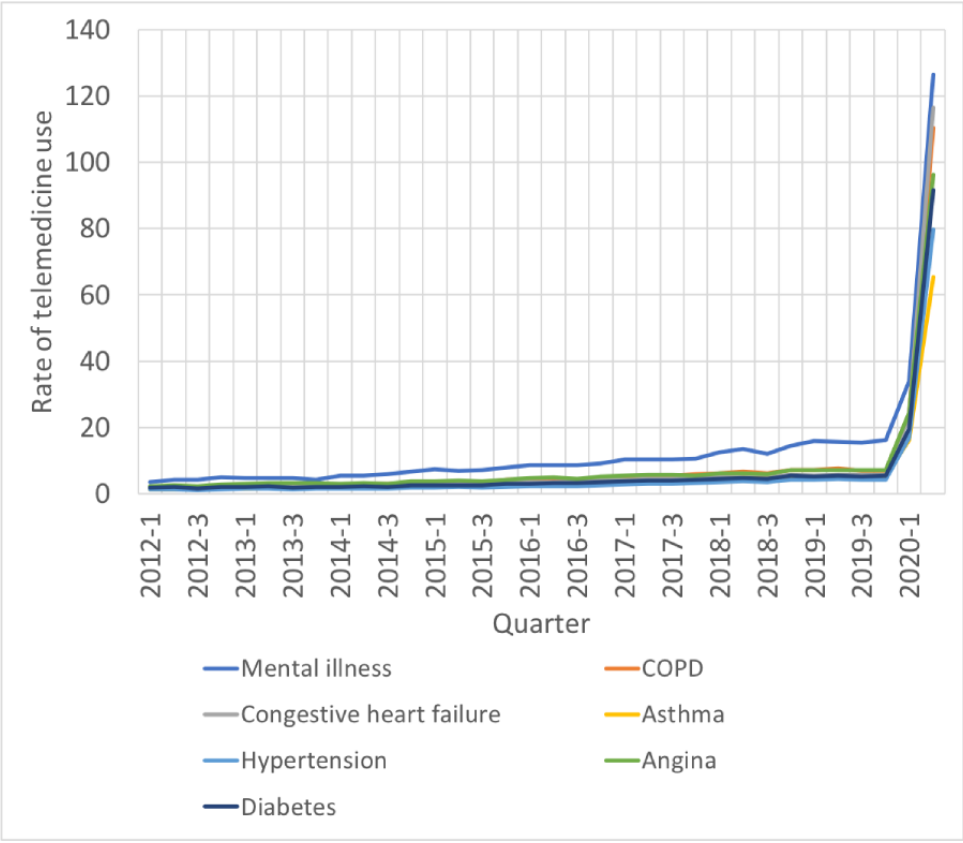
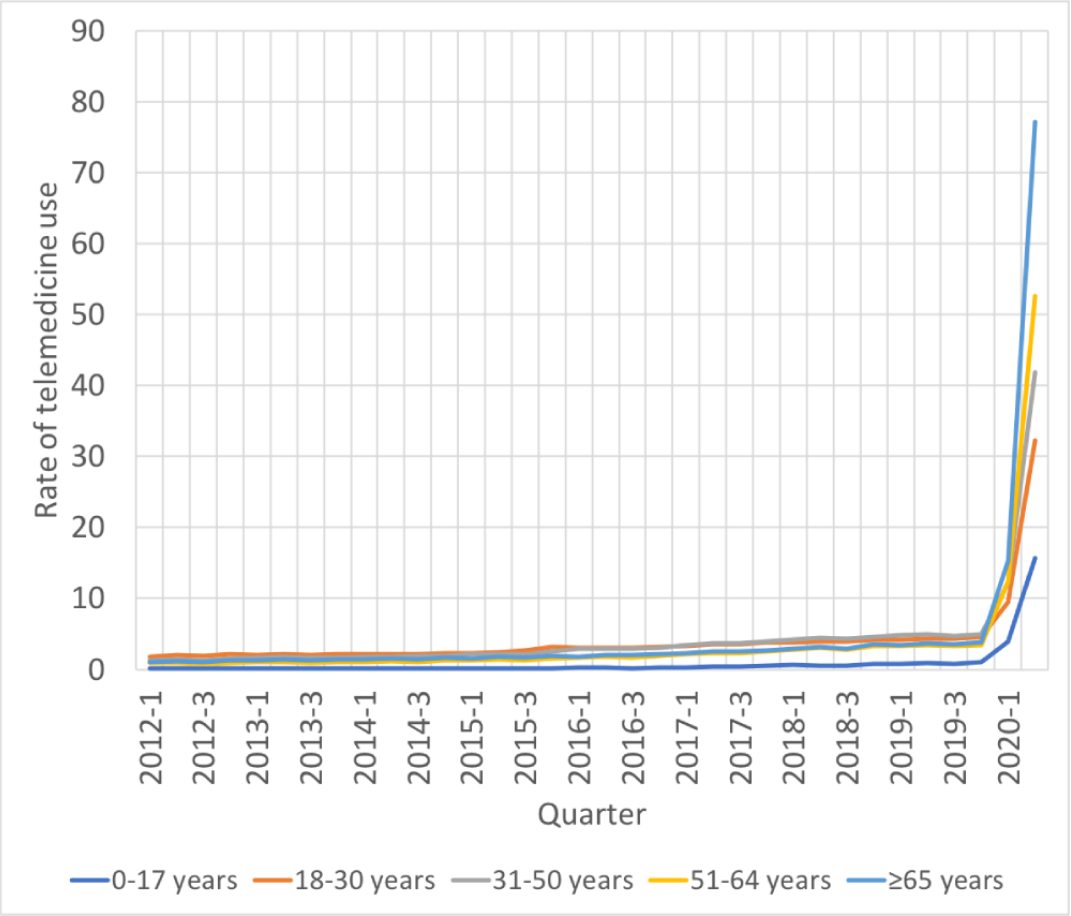


Figure 3. Rate of telemedicine visits per 100 rural patients by age group, 2012-2020.



Discussion

Principal Findings

In this population-based study, we sought to quantify the impact of the COVID-19 pandemic on rural telemedicine use across a large, geographically and demographically diverse region. While telemedicine use before the COVID-19 pandemic was higher in rural areas than in urban areas of Ontario, the use of telemedicine was modest. Broad telemedicine usage in Ontario was low in part owing to the lack of billing codes for many types of telemedicine services outside of the government-operated videoconferencing platform. With the initiation of these temporary fee-codes allowing reimbursement for other forms of telemedicine visits during the pandemic, usage increased rapidly, with the temporary visit codes comprising the vast majority of telemedicine visits. The onset of the pandemic led to a rapid increase in telemedicine usage in both rural and urban areas of Ontario, with high adoption of telemedicine among patients in urban areas and those in less rural areas, compared to those in the most remote regions of Ontario. We speculate that higher rates of urban telemedicine use during the pandemic coincide with generally higher rates of urban health care use as telemedicine became ubiquitous. We note that the slight reduction or plateau in the telemedicine usage rates among both urban and rural populations from April to June 2020 coincided with the loosening of lockdown restrictions after the first wave of the COVID-19 pandemic and may have resulted from an increase in in-person visits during this period.

Telemedicine usage increased rapidly among all age groups in rural areas, although the usage rate increased with age. Older adult patients used telemedicine at the highest rate during the pandemic and children and adolescents used telemedicine the least. Telemedicine usage varied among patients with different chronic diseases, those with mental illness and with congestive heart failure making maximum use of telemedicine, and those with other ambulatory sensitive chronic conditions making substantial use of telemedicine. This study is the first to quantify the rates of telemedicine use in rural populations, particularly among older patients and those with chronic diseases.

Health care access has been a major challenge for rural patients in many jurisdictions, particularly for chronic disease management and access to specialist care. Previous studies have described the increasing trend of telemedicine use specifically among rural beneficiaries even prior to the pandemic [13]. Numerous advantages are associated with the adoption of telemedicine in rural regions, such as saving of time and costs associated with traveling, reduction of hospital readmissions, prevention of emergency department visits, and increasing the availability of inpatient beds for patients requiring critical care [14]. Rural telemedicine has been shown to improve health care access and quality in various areas, ranging from trauma and critical care [15-17], family medicine [18], pediatrics [19], and cancer care [20]. In fact, several randomized controlled trials have reported that telemedicine-based care can lead to similar

and potentially better outcomes compared to in-person care among rural patients [21,22].

Furthermore, previous studies have reported that populations overall have increased their use of telemedicine and virtual care during the pandemic [23] to avail of both urgent and nonurgent care [24], and benefits have been observed within various subspecialties [25,26]. Telemedicine allows health care professionals to safely deliver care by reducing the risk of potential exposure to the care provider and the patient, preserving personal protective equipment, and lessening patient loads at hospitals and facilities [27]. This proves especially useful in rural settings, where a significant proportion of the population are older adults and are at a higher risk of COVID-19.

Telemedicine uptake has surged across several groups of patients with chronic diseases, potentially highlighting the importance of remote management of chronic conditions. This finding is concurrent with published data on the use of telemedicine among patients with various chronic diseases, such as diabetes [28] and congestive heart failure [29], and particularly among patients residing in rural or remote areas [30]. Although rates of telemedicine use increased among patients with chronic disease, the proportion of telemedicine users with chronic disease appeared to decrease slightly over the years, suggesting a potentially increased uptake of telemedicine overall. We hypothesize that with the widespread use of telemedicine, a greater proportion of healthier patients used telemedicine for their health care needs. Longitudinal studies on the changes in the trends of telemedicine usage, particularly in the postpandemic period, are needed.

Despite evidence on the effectiveness of virtual ambulatory care for older adults [31], some studies have reported the challenges of adopting telemedicine in this population, owing to barriers such as disabilities or lack of devices, proper internet connectivity, and experience with technology [32,33]. However, our study shows that telemedicine usage increased significantly among older adult patients during the pandemic. This was likely a result of the new COVID-19 billing codes, which allowed for the reimbursement of telephone visits, as telephone calls may be easier to access for older adults than videoconferencing. Our findings are concurrent with those of a similar study we conducted, which focused on the rates of telemedicine visits in the entire patient population in Ontario, and we previously reported that the increase in telemedicine usage rates from before the pandemic to during the pandemic was similar across age groups, with older adults showing the highest rates of telemedicine use during the pandemic [3].

Limitations

The limitations of this study include a shortage of data comparing different modalities of communication technology, specifically telephone versus video visits. The new COVID-19 billing codes reimburse for telemedicine delivered through various modalities, and the type of modality used is not available in administrative data sets. The lack of clinical granularity that accompanies the use of administrative data also implies our inability to deduce the reasons for telemedicine use and assess the quality of care in our study population.

Conclusions

Our study investigates the trend in telemedicine use among patients residing in rural Ontario before and during the COVID-19 pandemic. Rural patients globally face many barriers to care, and telemedicine has proved important in helping patients access the health care services they need. Uptake of telemedicine services increased after the onset of the pandemic for the rural patient population of Ontario and across various subgroups, including those who are older or those with chronic disease. Telemedicine use appears to have increased more significantly among patients residing in less rural areas compared to those residing in more rural areas during the pandemic. Further studies are required to assess the potential barriers to telemedicine experienced by rural populations compared to those experienced by urban populations and the impact of telemedicine compared to that of in-person care on other forms of health care utilization, outcomes, and quality of care among vulnerable and at-risk patient groups in the rural population.

Acknowledgments

This study was supported by the ICES, which is funded by an annual grant from the Ontario Ministry of Health and Long-Term Care. This study was also supported by a grant from the US National Institute on Aging provided to PC (R01AG058878). Parts of this material are based on data and information compiled and provided by the Ontario Ministry of Health and Long-Term Care and the Canadian Institute for Health Information. The analyses, conclusions, opinions, and statements expressed herein are solely those of the authors and do not reflect those of the funding or data sources; no endorsement is intended or should be inferred.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplemental Tables.
[DOCX File , 14 KB - jmir_v23i4e26960_app1.docx]

Multimedia Appendix 2

Rate of telemedicine visits per 1000 rural patients, by region.
[PNG File , 49 KB - jmir_v23i4e26960_app2.png]

Multimedia Appendix 3
Medical specialties of physicians who delivered any rural visit in 2012, 2016, and 2020.
[DOCX File , 14 KB - jmir_v23i4e26960_app3.docx]

References

1. Illustrated Glossary. Statistics Canada. 2017 Nov 15. URL: https://www150.statcan.gc.ca/n1/en/pub/92-195-x/92-195-x2016001-eng.pdf?st=BimwLQLd [accessed 2021-03-30]
2. Brown EM. The Ontario Telemedicine Network: a case report. Telemed J E Health 2013 May;19(5):373-376. [doi: 10.1089/tmj.2012.0299] [Medline: 23301768]
3. Bhatia RS, Chu C, Pang A, Tadrous M, Stamenova V, Cram P. Virtual care use before and during the COVID-19 pandemic: a repeated cross-sectional study. CMAJ Open 2021;9(1):E107-E114 [FREE Full text] [doi: 10.9778/cmajo.20200311] [Medline: 33597307]
4. Keeping Health Care Providers informed of payment, policy or program changes. Re: Changes to the Schedule of Benefits for Physician Services (Schedule) in response to COVID-19 influenza pandemic effective March 14, 2020. Ontario Ministry of Health and Long-Term Care. 2019 Nov 15. URL: http://www.health.gov.on.ca/en/pro/programs/ohip/bulletins/4000/bul4731.aspx [accessed 2021-03-30]
5. Eaton D, Hogenbirk J, McLean J, Nashkawa D, Simpson C, Skinner M, et al. Rural Ontario Foresight Papers. Rural Ontario Institute. 2019. URL: https://www.ruralontarioinstitute.ca/uploads/userfiles/files/Rural%20Ontario%20Foresight%20Papers_Oct%2023.pdf [accessed 2021-03-30]
6. Bhatia RS, Shojania KG, Levinson W. Cost of contact: redesigning healthcare in the age of COVID. BMJ Qual Saf 2021 Mar;30(3):236-239. [doi: 10.1136/bmjqs-2020-011624] [Medline: 32763977]
7. Keeping Health Care Providers informed of payment, policy or program changes. Re: Changes to the Schedule of Benefits for Physician Services (Schedule) in response to COVID-19 influenza pandemic effective March 14, 2020. Ontario Ministry of Health and Long-Term Care. 2020 Mar 13. URL: http://www.health.gov.on.ca/en/pro/programs/ohip/bulletins/4000/bul4745.aspx [accessed 2021-03-30]
8. Personal Health Information Protection Act, 2004. Government of Ontario. URL: https://www.ontario.ca/laws/statute/04p03#BK63 [accessed 2021-03-30]
9. Kralj B. Measuring Rurality - RIO2008_BASIC: Methodology and Results. Ontario Medical Association: Economics Department. 2009 Feb 02. URL: https://content.oma.org//wp-content/uploads/2008rio-fulltechnicalpaper.pdf [accessed 2021-03-30]
10. Glazier R, Zagorski B, Rayner J. Comparison of Primary Care Models in Ontario by Demographics, Case Mix and Emergency Department Use, 2008/09 to 2009/10. Data Discovery Better Health. 2012 Mar. URL: https://www.ices.on.ca/Publications/Atlases-and-Reports/2012/Comparison-of-Primary-Care-Models [accessed 2021-03-30]
11. Schull MJ, Azimaee M, Marra M, Cartagena RG, Vermeulen MJ, Ho MM, et al. ICES: Data, Discovery, Better Health. IJPDS 2020 Mar 11;4(2):1135. [doi: 10.23889/ijpds.v4i2.1135] [Medline: 32935037]
12. Turcotte M, Schellenberg G. A Portrait of Seniors in Canada. Statistics Canada. 2006. URL: https://www150.statcan.gc.ca/n1/en/pub/89-519-x/89-519-x2006001-eng.pdf?st=wBkQ3eIv [accessed 2021-03-30]
13. Mehrotra A, Jena AB, Busch AB, Souza J, Uscher-Pines L, Landon BE. Utilization of Telemedicine Among Rural Medicare Beneficiaries. JAMA 2016 May 10;315(18):2015-2016 [FREE Full text] [doi: 10.1001/jama.2016.2186] [Medline: 27163991]
14. Telehealth Use in Rural Healthcare. Rural Health Information Hub. URL: https://www.ruralhealthinfo.org/topics/telehealth#:~:text=Telehealth%20allows%20small%20rural%20hospitals,distances%20to%20access%20specialty%20care [accessed 2021-03-30]
15. Ricci MA, Caputo M, Amour J, Rogers FB, Sartorelli K, Callas PW, et al. Telemedicine reduces discrepancies in rural trauma care. Telemed J E Health 2003;9(1):3-11. [doi: 10.1089/153056203763317602] [Medline: 12699603]
16. Heath B, Salerno R, Hopkins A, Hertzig J, Caputo M. Pediatric critical care telemedicine in rural underserved emergency departments. Pediatr Crit Care Med 2009 Sep;10(5):588-591. [doi: 10.1097/PCC.0b013e3181a63eac] [Medline: 19451850]
17. Zawada ET, Herr P, Larson D, Fromm R, Kapaska D, Erickson D. Impact of an intensive care unit telemedicine program on a rural health care system. Postgrad Med 2009 May;121(3):160-170. [doi: 10.3810/pgm.2009.05.2016] [Medline: 19491554]
18. Norris T, Hart G, Larson E, Tarczy-Hornoch P, Masuda D, Fuller S. Low-bandwidth, low-cost telemedicine consultations in rural family practice. J Am Broad Fam Med 2002;15(2):123. [Medline: 12002196]
19. Marcin JP, Ellis J, Mawis R, Nagrampa E, Nesbitt TS, Dimand RJ. Using telemedicine to provide pediatric subspecialty care to children with special health care needs in an underserved rural community. Pediatrics 2004 Jan;113(1 Pt 1):1-6. [doi: 10.1542/peds.113.1.1] [Medline: 14702439]

20. Sabesan S, Larkins S, Evans R, Varma S, Andrews A, Beuttner P, et al. Telemedicine for rural cancer care in North Queensland: bringing cancer care home. Aust J Rural Health 2012 Oct;20(5):259-264. [doi: 10.1111/j.1440-1584.2012.01299.x] [Medline: 22998200]
21. Fortney JC, Pyne JM, Mouden SB, Mittal D, Hudson TJ, Schroeder GW, et al. Practice-Based Versus Telemedicine-Based Collaborative Care for Depression in Rural Federally Qualified Health Centers: A Pragmatic Randomized Comparative Effectiveness Trial. AJP 2013 Apr;170(4):414-425. [doi: 10.1176/appi.ajp.2012.12050696] [Medline: 32015700]
22. Morland LA, Greene CJ, Rosen CS, Foy D, Reilly P, Shore J, et al. Telemedicine for anger management therapy in a rural population of combat veterans with posttraumatic stress disorder: a randomized noninferiority trial. J Clin Psychiatry 2010 Jul;71(7):855-863. [doi: 10.4088/JCP.09m05604blu] [Medline: 20122374]
23. Webster P. Virtual health care in the era of COVID-19. The Lancet 2020 Apr;395(10231):1180-1181. [doi: 10.1016/s0140-6736(20)30818-7] [Medline: 32278374]
24. Mann D, Chen J, Chunara R, Testa P, Nov O. COVID-19 transforms health care through telemedicine: Evidence from the field. J Am Med Inform Assoc 2020 Jul 01;27(7):1132-1135. [doi: 10.1093/jamia/ocaa072] [Medline: 32324855]
25. Smrke A, Younger E, Wilson R, Husson O, Farag S, Merry E, et al. Telemedicine During the COVID-19 Pandemic: Impact on Care for Rare Cancers. JCO Glob Oncol 2020 Jul;6:1046-1051 [FREE Full text] [doi: 10.1200/GO.20.00220] [Medline: 32639877]
26. Novara G, Checcucci E, Crestani A, Abrate A, Esperto F, Pavan N, Research Urology Network (RUN). Telehealth in Urology: A Systematic Review of the Literature. How Much Can Telemedicine Be Useful During and After the COVID-19 Pandemic? Eur Urol 2020 Dec;78(6):786-811. [doi: 10.1016/j.eururo.2020.06.025] [Medline: 32616405]
27. Using Telehealth to Expand Access to Essential Health Services during the COVID-19 Pandemic. Centers for Disease Control and Prevention. 2020 Jun 10. URL: https://www.cdc.gov/coronavirus/2019-ncov/hcp/telehealth.html [accessed 2021-03-30]
28. Jones MS, Goley AL, Alexander BE, Keller SB, Caldwell MM, Buse JB. Inpatient Transition to Virtual Care During COVID-19 Pandemic. Diabetes Technol Ther 2020 Jun;22(6):444-448. [doi: 10.1089/dia.2020.0206] [Medline: 32396395]
29. Jerant AF, Azari R, Nesbitt TS. Reducing the cost of frequent hospital admissions for congestive heart failure: a randomized trial of a home telecare intervention. Med Care 2001 Nov;39(11):1234-1245. [doi: 10.1097/00005650-200111000-00010] [Medline: 11606877]
30. Jaglal SB, Haroun VA, Salbach NM, Hawker G, Voth J, Lou W, et al. Increasing access to chronic disease self-management programs in rural and remote communities using telehealth. Telemed J E Health 2013 Jun;19(6):467-473 [FREE Full text] [doi: 10.1089/tmj.2012.0197] [Medline: 23570277]
31. Batsis JA, DiMilia PR, Seo LM, Fortuna KL, Kennedy MA, Blunt HB, et al. Effectiveness of Ambulatory Telemedicine Care in Older Adults: A Systematic Review. J Am Geriatr Soc 2019 Aug;67(8):1737-1749. [doi: 10.1111/jgs.15959] [Medline: 31066916]
32. Lam K, Lu AD, Shi Y, Covinsky KE. Assessing Telemedicine Unreadiness Among Older Adults in the United States During the COVID-19 Pandemic. JAMA Intern Med 2020 Oct 01;180(10):1389-1391. [doi: 10.1001/jamainternmed.2020.2671] [Medline: 32744593]
33. Fisk M, Livingstone A, Pit SW. Telehealth in the Context of COVID-19: Changing Perspectives in Australia, the United Kingdom, and the United States. J Med Internet Res 2020 Jun 09;22(6):e19264 [FREE Full text] [doi: 10.2196/19264] [Medline: 32463377]

Abbreviations

ICES: Institute for Clinical Evaluative Sciences
OHIP: Ontario Health Insurance Plan
OTN: Ontario Telemedicine Network
RIO: rurality index of Ontario

Edited by R Kukafka; submitted 05.01.21; peer-reviewed by C Hawke, E van der Velde, M Mueller; comments to author 08.02.21; revised version received 03.03.21; accepted 24.03.21; published 05.04.21.

Please cite as:
Chu C, Cram P, Pang A, Stamenova V, Tadrous M, Bhatia RS
Rural Telemedicine Use Before and During the COVID-19 Pandemic: Repeated Cross-sectional Study
J Med Internet Res 2021;23(4):e26960
URL: https://www.jmir.org/2021/4/e26960
doi:10.2196/26960
PMID:33769942

Original Paper

Teleassistance for Patients With Type 1 Diabetes During the COVID-19 Pandemic: Results of a Pilot Study

Martina Parise[1], MD; Linda Tartaglione[2], MD; Antonio Cutruzzolà[1], MD; Maria Ida Maiorino[3], MD; Katherine Esposito[3], MD; Dario Pitocco[2], MD; Agostino Gnasso[4], MD; Concetta Irace[5], MD

[1]Department of Experimental and Clinical Medicine, University Magna Graecia Catanzaro, Catanzaro, Italy

[2]IRCCS, University Agostino Gemelli, Rome, Italy

[3]Department of Advanced Medical and Surgical Science, University Vanvitelli, Napoli, Naples, Italy

[4]Department of Advanced Medical and Surgical Science, University Magna Graecia Catanzaro, Naples, Italy

[5]Department of Health Science, University Magna Graecia Catanzaro, Catanzaro, Italy

Corresponding Author:
Concetta Irace, MD
Department of Health Science
University Magna Graecia Catanzaro
Viale Europa
Catanzaro, 88100
Italy
Phone: 39 3316718143
Email: irace@unicz.it

Abstract

Background: Telemedicine use in chronic disease management has markedly increased during health emergencies due to COVID-19. Diabetes and technologies supporting diabetes care, including glucose monitoring devices, software analyzing glucose data, and insulin delivering systems, would facilitate remote and structured disease management. Indeed, most of the currently available technologies to store and transfer web-based data to be shared with health care providers.

Objective: During the COVID-19 pandemic, we provided our patients the opportunity to manage their diabetes remotely by implementing technology. Therefore, this study aimed to evaluate the effectiveness of 2 virtual visits on glycemic control parameters among patients with type 1 diabetes (T1D) during the lockdown period.

Methods: This prospective observational study included T1D patients who completed 2 virtual visits during the lockdown period. The glucose outcomes that reflected the benefits of the virtual consultation were time in range (TIR), time above range, time below range, mean daily glucose, glucose management indicator (GMI), and glycemic variability. This metric was generated using specific computer programs that automatically upload data from the devices used to monitor blood or interstitial glucose levels. If needed, we changed the ongoing treatment at the first virtual visit.

Results: Among 209 eligible patients with T1D, 166 completed 2 virtual visits, 35 failed to download glucose data, and 8 declined the visit. Among the patients not included in the study, we observed a significantly lower proportion of continuous glucose monitoring (CGM) and continuous subcutaneous insulin infusion (CSII) users (n=7/43, 16% vs n=155/166, 93.4% and n=9/43, 21% vs n=128/166, 77.1%, respectively; *P*<.001) compared to patients who completed the study. TIR significantly increased from the first (62%, SD 18%) to the second (65%, SD 16%) virtual visit (*P*=.02); this increase was more marked among patients using the traditional meter (n=11; baseline TIR=55%, SD 17% and follow-up TIR=66%, SD 13%; *P*=.01) than among those using CGM, and in those with a baseline GMI of ≥7.5% (n=46; baseline TIR=45%, SD 15% and follow-up TIR=53%, SD 18%; *P*<.001) than in those with a GMI of <7.5% (n=120; baseline TIR=68%, SD 15% and follow-up TIR=69%, SD 15%; *P*=.98). The only variable independently associated with TIR was the change of ongoing therapy. The unstandardized beta coefficient (B) and 95% CI were 5 (95% CI 0.7-8.0) (*P*=.02). The type of glucose monitoring device and insulin delivery systems did not influence glucometric parameters.

Conclusions: These findings indicate that the structured virtual visits help maintain and improve glycemic control in situations where in-person visits are not feasible.

(J Med Internet Res 2021;23(4):e24552) doi:10.2196/24552

KEYWORDS

chronic disease; COVID-19; diabetes; effectiveness; management; technology; teleassistance; telehealth; telemedicine; time in range; type 1 diabetes

Introduction

Background

The COVID-19 pandemic has restricted access to medical clinics in order to prevent the risk of infection. In this context, routine care for patients with chronic diseases such as diabetes has emerged as a major challenge. Several strategies have been implemented to support patients with type 1 diabetes (T1D) and type 2 diabetes during the COVID-19 pandemic and to provide adequate assistance to avoid disease exacerbation [1-4]. Patients with diabetes have been recommended to follow general guidelines on infection risk reduction, blood glucose monitoring, taking medication, injecting insulin and noninsulin drugs adequately, and maintaining a healthy lifestyle (through diet and physical activity) [5].

Furthermore, experts have strongly suggested implementing the practice of downloading glucose data with dedicated software and sharing information with the health care providers to facilitate the provision of remote assistance [6]. More generally, the efficacy of virtual visits through telephone calls, text messages, mobile apps, or electronic visits—outside the context of the COVID-19 pandemic—has been described in some studies and systematic reviews. The current evidence is not univocal, and text messages seem more effective than web-based interventions [7-10]. Despite the number of published articles reporting a consensus on video consultations and guidelines for managing diabetes in the COVID-19 setting, no prospective evidence is available regarding the efficacy of teleassistance as an alternative to ambulatory visits during the pandemic [1,2,5,6]. Indeed, most studies, with a relevant number of patients, are retrospective and have analyzed data remotely with specific software that download glucose data without contacting the patients.

Aim and Hypothesis

As a necessary measure, we have implemented remote medical examination of patients with T1D during the lockdown period (March 10 to June 3, 2020). In this study, we assessed new glucometric parameters collected during 2 different virtual visits. We hypothesized that the COVID-19 pandemic might affect metabolic control in patients with diabetes owing to limited access to diabetes care centers and poor compliance to lifestyle recommendations. We accordingly designed a pilot study to prospectively analyze the effectiveness of 2 structured virtual visits (video or telephone consultations) on the basis of the time in range (TIR), time above range (TAR), time below range (TBR), mean daily glucose, glucose management indicator (GMI), and glycemic variability [11].

Methods

Patients

Our pilot study is a prospective, single-arm, observational study including patients with T1D who completed 2 virtual visits (baseline and follow-up visits) from the start (March 10, 2020) to the end (June 3, 2020) of the lockdown period. The protocol was preliminarily submitted and approved by the local Ethical Committee “Regione Calabria, Area Centro” (approval# 79-2020). In total, 3 diabetes care centers at teaching hospitals were involved in this study. A nurse or physician contacted the patients through the telephone before the first virtual visit (first contact) to explain the study's purpose to them. Patients who provided verbal consent were enrolled. Baseline characteristics of patients who declined to participate in the study were collected and compared to those of patients who agreed to participate. During the telephone call, patients were informed about or instructed, if necessary, on how to download the data from continuous glucose monitoring (CGM) systems or a traditional meter into the cloud by using specific software, such as Clarity, Care Link, Eversense diabetes management system (DMS), Accu-Chek Connect DMS, and share the data through the cloud or email. They were also invited to send recent laboratory findings or reports from other specialists via email or mobile apps. Patients using traditional meters were strongly suggested to check their blood glucose levels before each meal and in the presence of symptoms suggestive of hypoglycemia. During the second contact (baseline virtual visit), the physician verified the clinical conditions, ongoing treatment, and adherence to healthy lifestyle recommendations. Blood glucose data were monitored, and treatment was changed accordingly. The second virtual follow-up visit (third contact) was decided on the basis of clinical conditions and glycemic control and completed as the baseline visit. All information collected during the virtual visit were uploaded in the electronic medical files, and the summary of the visits was shared via email.

Data Collection

For the present analysis, the following parameters were collected: TIR, TAR, TBR, and mean (SD) daily glucose levels determined 2 weeks before the baseline and follow-up visits. The coefficient of variation (CV)—an estimate of glycemic variability—and the GMI—an estimate of glycated hemoglobin levels—were determined using the following formulae if not automatically generated by the software analyzing glucose data [12]:

$$CV\ (\%) = [(SD/Mean\ Glucose) \times 100] \quad \textbf{(1)}$$

$$GMI\ (\%) = 3.38 + 0.02345 \times (Mean\ glucose\ in\ mg/dL) \quad \textbf{(2)}$$

Statistical Analyses

Statistical analyses were performed using SPSS for Macintosh (version 23, IBM Corp). Patients who declined to participate or complete 2 virtual visits were excluded from the analyses. For the statistical analyses, patients were grouped as follows: those included and excluded from the study, those using the sensor and injecting insulin with a pump (CGM + continuous subcutaneous insulin infusion [CSII]), those using the sensor and with multiple daily insulin injection (CGM+MDI), and

those testing blood glucose with a traditional meter (ie, self-monitoring blood glucose [SMBG]) and injecting insulin with a pump or through MDI (CSII or MDI+SMBG). Sensor users included patients using intermittently scanned CGM (isCGM) and real-time CGM (rtCGM). Variables including baseline mean daily glucose, GMI, and TBR were not normally distributed. Parametric and nonparametric tests were performed accordingly.

The 2-tailed *t* test for unpaired data and the Mann-Whitney *U* test were used to compare between-group differences. The 2-tailed *t* test for paired data and the Wilcoxon Signed-Rank test were used to compare variables between baseline and follow-up visits. The chi-square test was used to compare percentages (CGM or sensor use and MDI or CSII treatment) between groups. Multivariable regression analysis was performed to evaluate variables independently associated with the absolute difference in TIR between baseline and follow-up visits. Independent variables included in the model were age, disease duration, use of the sensor or pump, and therapy change. The absolute difference in TIR, calculated as the difference between the follow-up TIR and the baseline TIR, was not normally distributed and hence log-transformed before regression analysis.

Results

Characteristics of Patients

In total, 209 patients were scheduled for in-person visits during the lockdown, from March 10 to June 3, 2020. Of them, 43 (20.6%) patients declined to participate in the study. Furthermore, 35 (16.7%) patients stated difficulties in downloading or sharing data via email, and 8 (3.8%) did not attend the virtual visit. The remaining 166 (79.4%) patients were enrolled and assessed. The average mean interval between the 2 virtual visits was 11 weeks. The characteristics of patients included and excluded from the study are summarized in Table 1. The proportion of CGM and CSII users was significantly higher among included patients than among excluded patients (n=155/166, 93.4% vs n=7/43, 16%, respectively; *P*<.001). Among 166 patients included in this study, 11 (6.6%) monitored their blood glucose levels with SMBG, 20 (12.0%) used the isCGM, and 135 (81.3%) used rtCGM. Among the excluded 43 patients, only 7 (16%) were using CGM, while the remaining 36 (84%) were using SMBG. All patients monitoring blood glucose through the traditional meter used the Accu-Chek Connect DMS. All the patients included in the study had strips and sensors to monitor their glucose levels during the lockdown. The mean number of tests per day was 4.4 [SD 1.9] and 5.1 [SD 2.4] (*P*=.21) among patients using SMBG and 9.5 [SD 4.9] and 9.8 [SD 6.1] (*P*=.70) among those using isCGM at baseline and follow-up visits, respectively.

Table 1. Characteristics of patients with type 1 diabetes included and excluded from the analyses (N=209).

Variables	Patients included (n=166)	Patients excluded (n=43)	*P* value
Age (years), mean (SD)	40 (14)	37 (15)	N/A[a]
Males, n (%)	80 (48.2)	21 (49)	N/A
Disease duration (years), mean (SD)	20 (11)	17 (9)	N/A
SMBG[b] users, n (%)	11 (6.6)	36 (84)	<.001
CGM[c] (isCGM[d] + rtCGM[e]) users, n (%)	155 (93.4)	7 (16)	<.001
MDI[f] users, n (%)	38 (22.9)	34 (79)	<.001
CSII[g] users, n (%)	128 (77.1)	9 (21)	<.001

[a]N/A: not applicable.

[b]SMBG: self-monitoring blood glucose.

[c]CGM: continuous glucose monitoring.

[d]isCGM: intermittently scanned continuous glucose monitoring.

[e]rtCGM: real-time continuous glucose monitoring.

[f]MDI: multiple daily insulin injection.

[g]CSII: continuous subcutaneous insulin infusion.

Glucometric Characteristics at Baseline and Follow-up Visits

We compared glucometric parameters measured at the baseline and follow-up visits among all patients included in the study and grouped in accordance with different combinations of insulin delivery methods and glucose monitoring systems (CSII+CGM, MDI+CGM, and CSII or MDI+SMBG) (Table 2). The TIR significantly increased from the baseline to the follow-up visit in all patients with T1D (62%, SD 18% vs 65%, SD 16%, respectively; *P*=.02) and in the CSII or MDI+SMBG group (55%, SD 17% vs 66%, SD 13%, respectively; *P*=.01). Furthermore, the CSII or MDI+SMBG group displayed a significant improvement in the TAR at baseline and follow-up visits (40%, SD 18% vs 28%, SD 15%, respectively; *P*=.03), mean daily glucose (176 [SD 49] mg/dL vs 150 [SD 25] mg/dL; *P*=.04), GMI (7.5%, SD 1.1% vs 6.9% SD 0.6%; *P*=.04), and

CV (36%, SD 8% vs 42%, SD 9%; *P*=.04) compared to the other groups.

Based on the baseline GMI findings among all patients included in the study, we observed that the TIR significantly improved from baseline to follow-up visits among those with a GMI of ≥7.5% (n=46; 45%, SD 15% vs 53%, SD 18%; *P*<.01) compared to those with a GMI of <7.5% (n=120; 68%, SD 15% vs 69%, SD 15%; *P*=.98).

Table 2. Glucometric parameters at baseline and follow-up virtual visits in all patients and in those grouped in accordance with the insulin delivery method and glucose monitoring system (N=166).

Patient groups	Parameters		*P* value[a]
	Baseline visit	Follow-up visit	
Time in range (%), mean (SD)			
All patients (n=166)	62 (18)	65 (16)	*.02*
CSII[b]+CGM[c] (n=122)	63 (17)	65 (17)	.24
MDI[d]+CGM (n=33)	62 (19)	64 (17)	.19
CSII or MDI+SMBG[e] (n=11)	55 (17)	66 (13)	*.01*
Time below range (%), mean (SD)			
All patients (n=166)	3.5 (4.1)	3.4 (3.8)	.58
CSII+CGM (n=122)	3.2 (4.0)	3.1 (3.7)	.86
MDI+CGM (n=33)	4.4 (4.3)	3.7 (3.3)	.34
CSII or MDI+SMBG (n=11)	4.7 (4.0)	5.8 (5.0)	.33
Time above range (%), mean (SD)			
All patients (n=166)	34 (18)	32 (18)	.08
CSII+CGM (n=122)	34 (18)	33 (18)	.40
MDI+CGM (n=33)	33 (21)	32 (18)	.52
CSII or MDI+SMBG (n=11)	40 (18)	28 (15)	*.03*
Mean daily glucose (mg/dL), mean (SD)			
All patients (n=166)	163 (29)	159 (25)	.25
CSII+CGM (n=122)	162 (25)	161 (24)	.90
MDI+CGM (n=33)	162 (37)	157 (26)	.17
CSII or MDI+SMBG (n=11)	176 (49)	150 (25)	*.04*
Coefficient of variation (%), mean (SD)			
All patients (n=166)	34 (6)	34 (7)	.32
CSII+CGM (n=122)	34 (6)	33 (7)	.93
MDI+CGM (n=33)	36 (7)	35 (7)	.55
CSII or MDI+SMBG (n=11)	36 (8)	42 (9)	*.04*
Glucose management indicator (%), mean (SD)			
All patients (n=166)	7.2 (0.7)	7.1 (0.6)	.23
CSII+CGM (n=122)	7.2 (0.6)	7.1 (0.6)	.90
MDI+CGM (n=33)	7.2 (0.8)	7.0 (0.6)	.12
CSII or MDI+SMBG (n=11)	7.5 (1.1)	6.9 (0.6)	*.04*

[a]Significant *P* values are shown in italics.

[b]CSII: continuous subcutaneous insulin infusion.

[c]CGM: continuous glucose monitoring.

[d]MDI: multiple daily insulin injection.

[e]SMBG: self-monitoring blood glucose.

In total, 104 (63%) patients were suggested a change of therapy during the baseline visit. Among them, 97 (93%) used CGM (isCGM or rtCGM), and 84 (81%) used CSII. These proportions were comparable with those of patients who were not suggested

a change of therapy (CGM: 93%; *P*=.60 and CSII: 71%; *P*=.10). The absolute difference in TIR between baseline and follow-up visits was significantly higher among patients who were suggested a changed of therapy (4%, SD 10%) than among those who were not suggested a change of therapy (0.1%, SD 10%) (*P*=.04). No significant difference was observed in the TBR, TAR, mean daily glucose, CV, and GMI between these 2 groups (data not shown).

None of the patients had diabetic ketoacidosis or severe hypoglycemia requiring hospitalization in the interval between the 2 virtual visits.

Multivariable regression analysis revealed that the change of therapy was the only variable independently associated with the absolute difference in TIR between baseline and follow-up visits. The unstandardized beta coefficient (B) and 95% CI were 5 (95% CI 0.7-8.0) (*P*=.02).

Discussion

Principal Findings

Our study shows that the structured virtual visits adequately maintain pre-existing glycemic control or improve the time spent in the target range among individuals with T1D during an emergency when in-person consultations are not permitted.

The benefits of virtual visits were discernible among patients using the traditional meters regardless of the type of insulin delivery modality (MDI or CSII) and among patients with a baseline GMI of ≥7.5%. These results suggest certain considerations. The telephone call preceding the baseline virtual visit may have encouraged patients using SMBG to download data from the meter into diabetes management systems, thus facilitating the subsequent interpretation of glucose data. In the actual scenario, it is common for patients using SMBG to not bring the meter along with them during in-person visits or to not have devices capable of connecting to the cloud, limiting appropriate adjustment of treatment. Patients using SMBG had higher values of GMI and a lower TIR, despite not presenting significant findings, than sensor users, which may have induced physicians to strengthen the more the general suggestions for diabetes management and convince patients to maintain a healthy lifestyle rather have a change of therapy. It could be argued that new metrics have been developed for glucose data collected by the sensor and that our results obtained from patients using the traditional meter might be due to chance. However, we have recently demonstrated that a strict correlation between TIR calculated by specific software, after downloading SMBG values, is significantly correlated with glycated hemoglobin levels. The strength of this association in our study was comparable to that between TIR calculated from CGM and glycated hemoglobin levels reported in other studies [13]. Patients using the meter reported a significant increase in CV compared to the other groups. This result can be explained by the higher daily fluctuation in the number of SMBG users compared to that of CGM users. Indeed, at the follow-up visit, mean glucose levels decreased by 11%, and the CV increased by 2.9%. This might be an unfavorable finding, if sustained in the long term. Some of the aforementioned considerations also apply to patients using CGM who had not experienced an exacerbation of glucose levels during the lockdown.

Patients with a GMI of ≥7.5% may have better managed their diabetes during the pandemic and may have had a healthy lifestyle. We believe that patients generally have had much more time available, owing to restrictions associated with stay-at-home orders. This may have stimulated a more in-depth analysis of glycemic data and potential interventions on incorrect habits.

Patients using the sensor or wearing the pump did not experience an exacerbation in glucose levels during the lockdown. In our opinion, this is a remarkable finding, given the current state of emergency. Sensor and pump users are in general well-educated and motivated to check and self-manage the disease. It is noteworthy that the proportion of pump and sensor users was very low among patients excluded from the study. The main limitation associated with patient recruitment was problems faced with downloading the glucose data.

Multivariable regression analysis revealed that a change of therapy during the virtual visit was the only variable independently associated with the absolute difference in TIR. Insulin therapy can be safely modified during a structured virtual visit. Indeed, none of the patients included in the teleassistance had a severe hypoglycemic episode or ketoacidosis.

Our results reinforce the evidence that virtual consultations may lead to appropriate care during a pandemic or when an in-person visit cannot be performed for any reason.

Comparison With Previous Studies

Recent case studies have reported that teleassistance can be provided safely and effectively for new-onset T1D and ketoacidosis, thus preventing hospital admission [14,15].

A retrospective study, including 13 adolescent patients with T1D, has reported that physical activity regularly performed during lockdown is associated with an improvement in the TIR [16]. Unfortunately, we did not collect information on physical activity from our patient cohort. However, we speculate that the patients included in our study stayed active at home in some way.

Another retrospective study, including 92 patients with T1D who use CGM systems, has reported an increase in the TIR from 59% to 63% during the lockdown. The study retrospectively reviewed glucose values downloaded into the cloud [17]. Similarly, a smaller study, including 33 patients with T1D who use the isCGM system that is connected to the clinic, has reported an increase in the TIR from 54% to 65% during the lockdown. However, that study did not indicate whether virtual contact was proposed during the lockdown [18].

We would like to remark on the potential educational role of virtual assistance. Indeed, a training session on the use of the pump and sensor and carbo-counting can be scheduled with specialists including nutritionists or nurses, favoring the access to technologies despite physical distancing [19].

Certain questions regarding teleassistance remain open, including those on cost, reimbursement, authorization, liability,

demographic characteristics of the people to be engaged, choice of the method for the virtual visit, involvement of the health care provider, duration of the visit, and adequate time interval between visits [20]. Furthermore, it is important to identify patients who need in-person visits despite the state of emergency, such as the current pandemic. Finally, it is important to consider patients with type 2 diabetes, who represent the majority of patients with diabetes and, in general, have limited access to technology and are less educated in self-managing diabetes. We should probably consider different strategies for type 2 diabetes.

Limitations

Our prospective study demonstrates the effectiveness of teleassistance in managing disease during the lockdown. Data sharing and remote visits help maintain or achieve adequate glycemic control through data analyses and therapy adjustment. However, our study has some limitations of note. For instance, the patient groups based on different therapeutic strategies are relatively small. Patients with T1D, who constitute the minority of patients with diabetes, frequently adopt personalized insulin delivery schedules and monitoring systems. It is therefore difficult to have large homogeneous patient groups. Furthermore, well-educated patients would have been more likely to provide their consent for virtual visits. This renders our study findings more reliable for patients who have received adequate therapeutic education and are cooperative. Ultimately, the selection criteria of our study were arbitrary; however, they were selected to include the largest number of patients, and we are confident that our findings are potentially applicable to most of our patients with T1D.

Conclusions

In conclusion, our results show that COVID-19 restrictions have provided an opportunity to bring teleassistance to the frontline in diabetes care. The advancement of technology and the development of new connected devices will further facilitate information exchange among patients, health care providers, and physicians.

Conflicts of Interest

None declared.

References

1. Rayman G, Lumb A, Kennon B, Cottrell C, Nagi D, Page E, London Inpatient Diabetes Network-COVID-19. Guidelines for the management of diabetes services and patients during the COVID-19 pandemic. Diabet Med 2020 Jul;37(7):1087-1089. [doi: 10.1111/dme.14316] [Medline: 32365233]
2. Ghosh A, Gupta R, Misra A. Telemedicine for diabetes care in India during COVID19 pandemic and national lockdown period: Guidelines for physicians. Diabetes Metab Syndr 2020;14(4):273-276 [FREE Full text] [doi: 10.1016/j.dsx.2020.04.001] [Medline: 32283497]
3. Jones MS, Goley AL, Alexander BE, Keller SB, Caldwell MM, Buse JB. Inpatient Transition to Virtual Care During COVID-19 Pandemic. Diabetes Technol Ther 2020 Jun;22(6):444-448. [doi: 10.1089/dia.2020.0206] [Medline: 32396395]
4. Garg SK, Rodbard D, Hirsch IB, Forlenza GP. Managing New-Onset Type 1 Diabetes During the COVID-19 Pandemic: Challenges and Opportunities. Diabetes Technol Ther 2020 Jun;22(6):431-439. [doi: 10.1089/dia.2020.0161] [Medline: 32302499]
5. Baral SD, Poteat T, Strömdahl S, Wirtz AL, Guadamuz TE, Beyrer C. Worldwide burden of HIV in transgender women: a systematic review and meta-analysis. Lancet Infect Dis 2013 Mar;13(3):214-222. [doi: 10.1016/S1473-3099(12)70315-8] [Medline: 23260128]
6. Società Italiana di Diabetologia. URL: http://www.siditalia.it [accessed 2021-03-31]
7. Aminuddin H, Jiao N, Jiang Y, Hong J, Wang W. Effectiveness of smartphone-based self-management interventions on self-efficacy, self-care activities, health-related quality of life and clinical outcomes in patients with type 2 diabetes: A systematic review and meta-analysis. Int J Nurs Stud 2019 Mar 08:103286. [doi: 10.1016/j.ijnurstu.2019.02.003] [Medline: 30827741]
8. Carallo C, Scavelli FB, Cipolla M, Merante V, Medaglia V, Irace C, Progetto Diabete Calabria. Management of Type 2 Diabetes Mellitus through Telemedicine. PLoS One 2015;10(5):e0126858 [FREE Full text] [doi: 10.1371/journal.pone.0126858] [Medline: 25974092]
9. Irace C, Schweitzer MA, Tripolino C, Scavelli FB, Gnasso A. Diabetes Data Management System to Improve Glycemic Control in People With Type 1 Diabetes: Prospective Cohort Study. JMIR Mhealth Uhealth 2017 Dec 21;5(11):e170 [FREE Full text] [doi: 10.2196/mhealth.8532] [Medline: 29162560]
10. Verhoeven F, van Gemert-Pijnen L, Dijkstra K, Nijland N, Seydel E, Steehouder M. The contribution of teleconsultation and videoconferencing to diabetes care: a systematic literature review. J Med Internet Res 2007 Dec 14;9(5):e37 [FREE Full text] [doi: 10.2196/jmir.9.5.e37] [Medline: 18093904]
11. Battelino T, Danne T, Bergenstal RM, Amiel SA, Beck R, Biester T, et al. Clinical Targets for Continuous Glucose Monitoring Data Interpretation: Recommendations From the International Consensus on Time in Range. Diabetes Care 2019 Aug;42(8):1593-1603 [FREE Full text] [doi: 10.2337/dci19-0028] [Medline: 31177185]

12. Bergenstal RM, Beck RW, Close KL, Grunberger G, Sacks DB, Kowalski A, et al. Glucose Management Indicator (GMI): A New Term for Estimating A1C From Continuous Glucose Monitoring. Diabetes Care 2018 Nov;41(11):2275-2280 [FREE Full text] [doi: 10.2337/dc18-1581] [Medline: 30224348]
13. Cutruzzolà A, Irace C, Parise M, Fiorentino R, Pio Tripodi PF, Ungaro S, et al. Time spent in target range assessed by self-monitoring blood glucose associates with glycated hemoglobin in insulin treated patients with diabetes. Nutr Metab Cardiovasc Dis 2020 Sep 24;30(10):1800-1805. [doi: 10.1016/j.numecd.2020.06.009] [Medline: 32669240]
14. Peters AL, Garg SK. The Silver Lining to COVID-19: Avoiding Diabetic Ketoacidosis Admissions with Telehealth. Diabetes Technol Ther 2020 Jul;22(6):449-453. [doi: 10.1089/dia.2020.0187] [Medline: 32383989]
15. Garg SK, Rodbard D, Hirsch IB, Forlenza GP. Managing New-Onset Type 1 Diabetes During the COVID-19 Pandemic: Challenges and Opportunities. Diabetes Technol Ther 2020 Jun;22(6):431-439. [doi: 10.1089/dia.2020.0161] [Medline: 32302499]
16. Tornese G, Ceconi V, Monasta L, Carletti C, Faleschini E, Barbi E. Glycemic Control in Type 1 Diabetes Mellitus During COVID-19 Quarantine and the Role of In-Home Physical Activity. Diabetes Technol Ther 2020 Jul;22(6):462-467. [doi: 10.1089/dia.2020.0169] [Medline: 32421355]
17. Mesa A, Viñals C, Pueyo I, Roca D, Vidal M, Giménez M, et al. The impact of strict COVID-19 lockdown in Spain on glycemic profiles in patients with type 1 Diabetes prone to hypoglycemia using standalone continuous glucose monitoring. Diabetes Res Clin Pract 2020 Oct;167:108354 [FREE Full text] [doi: 10.1016/j.diabres.2020.108354] [Medline: 32739380]
18. Bonora BM, Boscari F, Avogaro A, Bruttomesso D, Fadini GP. Glycaemic Control Among People with Type 1 Diabetes During Lockdown for the SARS-CoV-2 Outbreak in Italy. Diabetes Ther 2020 May 11:1-11 [FREE Full text] [doi: 10.1007/s13300-020-00829-7] [Medline: 32395187]
19. Vigersky RA, Velado K, Zhong A, Agrawal P, Cordero TL. The Effectiveness of Virtual Training on the MiniMed™ 670G System in People with Type 1 Diabetes During the COVID-19 Pandemic. Diabetes Technol Ther 2021 Feb;23(2):104-109 [FREE Full text] [doi: 10.1089/dia.2020.0234] [Medline: 32678672]
20. Fisk M, Livingstone A, Pit SW. Telehealth in the Context of COVID-19: Changing Perspectives in Australia, the United Kingdom, and the United States. J Med Internet Res 2020 Jun 09;22(6):e19264 [FREE Full text] [doi: 10.2196/19264] [Medline: 32463377]

Abbreviations

CGM: continuous glucose monitoring
CSII: continuous subcutaneous insulin infusion
CV: coefficient of variation
DMS: diabetes management system
GMI: glucose management indicator
isCGM: intermittently scanned continuous glucose monitoring
MDI: multiple daily insulin injection
rtCGM: real-time continuous glucose monitoring
SMBG: self-monitoring blood glucose
T1D: type 1 diabetes
TAR: time above range
TBR: time below range
TIR: time in range

Edited by R Kukafka; submitted 24.09.20; peer-reviewed by D Rodbard, K Fitzner; comments to author 13.11.20; revised version received 29.12.20; accepted 16.03.21; published 06.04.21.

Please cite as:
Parise M, Tartaglione L, Cutruzzolà A, Maiorino MI, Esposito K, Pitocco D, Gnasso A, Irace C
Teleassistance for Patients With Type 1 Diabetes During the COVID-19 Pandemic: Results of a Pilot Study
J Med Internet Res 2021;23(4):e24552
URL: https://www.jmir.org/2021/4/e24552
doi:10.2196/24552
PMID:33769945

Original Paper

Evolution of Online Health-Related Information Seeking in France From 2010 to 2017: Results From Nationally Representative Surveys

Pauline Ducrot[1], PhD; Ilaria Montagni[2], PhD; Viet Nguyen Thanh[1], MSc; Anne-Juliette Serry[1], MBA; Jean-Baptiste Richard[1], MSc

[1]Santé publique France, French national public health agency, Saint-Maurice, France

[2]Bordeaux Population Health Research Center, Team HEALTHY, University of Bordeaux, Inserm, Bordeaux, France

Corresponding Author:
Pauline Ducrot, PhD
Santé publique France, French national public health agency
12 rue du Val d'Osne
Saint-Maurice
France
Phone: 33 1 71 80 16 19
Email: pauline.ducrot@santepubliquefrance.fr

Abstract

Background: Given the rapid ongoing progression of the internet and increase in health information available from disparate online sources, it is important to understand how these changes impact online health information-seeking behavior of the population and the way of managing one's health.

Objective: This paper aims at describing the evolution of internet use as a source of health information between 2010 and 2017, as well as the characteristics of online health information seekers, topics of interest, sources of information, and trust in retrieved information and potential impact on behavior.

Methods: Data from the French nationally representative surveys Health Barometers were used (N=4141 in 2010, 4811 in 2014, and 6255 in 2017). Evolutions over time were assessed using chi-square tests. Associations with sociodemographic characteristics and health status were evaluated using logistic regression models.

Results: The use of the internet as a source of health information rose between 2010 and 2014 (from 37.3% to 67.9%, *P*<.001) but decreased significantly in 2017 (60.3%, *P*<.001). Overall, the profile of health information seekers compared with nonseekers did not change over time. They were more likely to be women, to be younger, to have a higher educational level, to have a higher household income, and to be executives. Between 2014 and 2017, the proportion of those who did not pay attention to the source of information significantly increased to reach 39.7% (*P*<.001). In 2017 as in 2014, general health-related websites remained the first source of information (38.6%) while institutional websites were the third source (8.1%). Most information seekers trusted the information found online in 2010 (more than 80%), with a slight decrease between 2014 and 2017 (*P*=.048). Among individual characteristics, trust in the information was the main determinant of the way of managing one's health (odds ratio 4.06, 95% CI 3.26-5.06).

Conclusions: After a rapid growth in the internet use for seeking health information in the 2010 to 2014 period, a decrease was recorded in 2017, in parallel with a decrease in trust in the quality and reliability of information found online. These findings underline the need for public health authorities to increase citizens' eHealth literacy and to provide alternative trustworthy sources combining the popularity and accessibility of general health information websites.

(J Med Internet Res 2021;23(4):e18799) doi:10.2196/18799

KEYWORDS

internet; health information; information-seeking behavior; eHealth literacy

Introduction

In Europe, internet access is now democratized with 80% of the households using the internet for personal use [1]. Thus, over the last decade, the internet has become a major source of information including health-related information, with about 6 out of 10 Europeans reporting seeking health information online in the past year [1].

Using the internet as a health information source has many advantages. By offering quick, easy, timely, and low-cost access to the information, the internet tends to expand access to health messages [2], thus impacting citizens' management of their health. Providing the information is reliable, citizens may increase their health knowledge, better understand the risks and benefits of some treatments, and participate in their health care decision making [3,4]. The internet also provides the possibility of personalized feedback [2,5] and contributes to addressing issues of geographical or mobility isolation and anonymity [2,5]. Indeed, the privacy offered by online information is particularly valuable for individuals searching for information on sensitive topics [2,5,6]. However, these advantages can easily turn into disadvantages, since the quality and authority of health-related information is debatable, and identifying trustworthy versus not trustworthy sites is challenging [2]. In fact, the multiplication of health-related websites increasingly raised the issue of accessibility to accurate and trustworthy information [2,7-9]. In light of this, users must have appropriate capacities to access, understand, appraise, and use health-related information in a digital environment (ie, eHealth literacy) [10]. Previous studies have shown that a high level of eHealth literacy is associated with good management of one's health [11-13].

Online health information-seeking behavior depends on the characteristics of individual internet users, which might determine the reasons to search for health-related information online. Indeed, the literature has highlighted several predictors of online health information seeking, such as sociodemographic characteristics and overall health conditions [14]. Poor health, being female, being younger, having a diploma, and having higher income are associated with seeking health information online [15-19]. Similar trends were observed for eHealth literacy according to age and educational level. However, no significant difference has been reported between men and women [13,20].

A few studies have recently explored the trends in the use of the internet for health information seeking. The majority of them have been conducted in the United States [21-23] or have focused on specific population subgroups such as older people [24,25], pregnant women [26,27], children and adolescents [28-30], cancer survivors [31], or sexual minorities [32]. Two studies were conducted in Europe, in Finland [33] and Norway [34], but they considered data up to 2007 or 2009. In the European Union, eHealth literacy has been identified as a priority to address health inequalities in the eHealth Action Plan 2012-2020. The Eurobarometer on digital health literacy performed in 2014 supported this objective by assessing how Europeans use online information to help manage their own health. However, to our knowledge, no recent data exist on the evolution of such practice.

In addition, despite the rapid ongoing progression of the internet and overall perception of an increase in health information available from disparate online sources, the evolution of such sources used to seek information has been not documented so far. Finally, it is important to understand how these changes impact the online health information-seeking behavior of the population in order to offer appropriate solutions to disseminate reliable health messages.

The main objective of this article, therefore, was to describe, based on nationally representative surveys, the evolution of internet use as a source of health information between 2010 and 2017 in France. More specifically, this study was aimed at (1) describing the prevalence of internet use for health-related purposes and the characteristics of online health information seekers over time; (2) assessing the evolutions of topics of interest and sources of health-related information found online; and (3) investigating the attention paid to sources of information, the trust in retrieved information, and the potential impact on management of one's health.

Methods

Survey Methodology and Participants

Data were extracted from 3 national surveys (called Health Barometers) conducted in 2010, 2014, and 2017 by the French national public health agency (Santé publique France, formerly the French Institute for Prevention and Health Education or INPES) in consultation with the French Ministry of Health [20]. Health Barometers are cross-sectional surveys of random representative samples of the French population conducted using computer-assisted telephone interviewing. These surveys were designed to measure the evolution of key indicators regarding health-related behaviors, attitudes, and opinions in the general population. The questionnaires and the data collection methods are available on the official survey website [35,36]. Briefly, Health Barometers evaluate different health topics such as addiction (tobacco, alcohol, illegal drugs), mental health, sexuality, nutrition, or vaccination, as well as use of the internet for health. The full questionnaire is designed not to exceed 30 minutes completion time. The general part of the questionnaire, lasting 20 minutes, is addressed to all participants, and specific parts, lasting 10 minutes, are asked to different subsamples.

Health Barometers are based on a 2-stage random sampling design: sampling of telephone numbers covering all metropolitan French regions and random selection of one member of the household, using the method proposed by Kish [37]. In 2010, because of the increasing rate of households that had abandoned their landline telephones for cell phones, a cell-only sample was added (12% of the sample to keep the same rate as in 2010 in France). In 2014, since a section of the population including people also having a landline preferred using a mobile phone, 2 overlapping samples were constituted: one surveyed by landline and the other by mobile phone [38].

For households of the landline sample, one person was randomly selected among eligible persons living in the household (aged 15 to 75 years in 2014 and 2017 and 15 to 85 years in 2010, speaking French). In the cellular sample, selection was done

among persons sharing the cell telephone (when such sharing was reported).

If a household or individual refused to participate or could not be reached, they were not replaced in the study. For this reason, considerable efforts were made to reach households and increase the response rate: a formal request to participate explaining who was conducting the study and the goals of the survey was sent by electronic mail or letter to participant (when such information could be found using a national reverse directory). For every sampled number, up to 40 attempts were made to complete an interview. The calls were staggered over times of day and days of the week to maximize the chances of making contact with a potential respondent. When the selected individual was reached but unavailable, an appointment was made. Individuals unwilling to participate at first were contacted again by specialized interviewers in order to recruit them. Response rate of the 3 surveys was 53% in 2010, 61% in 2014, and 49% in 2017. All collected data were anonymous and self-reported. The survey was approved by the National Data Protection Authority and complied with the European Union General Data Protection Regulation.

Data Collection

Sociodemographic and Economic Characteristics and Health Status

Participants were asked to provide sociodemographic data, including gender (men, women), age in categories (18-24 years, 25-34 years, 35-44 years, 45-54 years, 55-54 years, 65-75 years, >75 years), educational level (primary, secondary, postsecondary), employment status (working, student, unemployed, retired, other), occupational category (farmers, artisans, executives, intermediate profession, employees, manual workers) and monthly income. Occupational category was reported for the respondent (using the last job position for unemployed and retired people) or for the reference person in the household if the respondent had never worked before (ie, student). Monthly household income was calculated per consumption unit (CU), where one CU is attributed for the first adult in the household, 0.5 CU for other persons aged 14 years or older, and 0.3 CU for children under 14 years, following national statistics methodology and guidelines [39]. Income categories were defined using the tertiles of the entire database, including the 3 years of data sets. In addition, participants were asked if they have a chronic disease (yes/no).

Internet Use for Health Information Seeking

In 2010, 2014, and 2017, participants were asked whether they had used the internet to search for health-related information or advice in the last 12 months (yes/no/no access to the internet). This question was used to identify health information seekers versus non–health information seekers. Only online health seekers were further asked about the trust they had in health-related information obtained on the internet. The responses were rated on a 4-point Likert scale ranging from 1 (not trustworthy at all) to 4 (totally trustworthy) and grouped into 2 categories (not trustworthy vs trustworthy). In order to understand the effect of using the internet on the doctor/patient relationship, individuals were asked whether the information and advice found on the internet had changed the way they were taking care of their health (4-point Likert scale from not at all to definitely yes, further grouped in 2 categories, yes vs no). In addition, they were asked if their use of the internet led them to visit their doctor more often, less often, or to the same extent as they did before using the internet for health purposes.

In 2014 and 2017, online health information seekers were also asked about the topics of their searches including (1) general health and illnesses, medical news, and treatments; (2) sexually-related health risk; (3) contraception and methods to avoid pregnancy; (4) nutrition, weight gain, or eating disorders; (5) pregnancy or maternity; (6) child health and illness; (7) alcohol; (8) tobacco; (9) cannabis and other drugs; and (10) electronic cigarettes (the latter assessed in 2017 only).

In 2014 and 2017, the questionnaire also included questions about the source of health information in general (forum, health information website, or did not pay attention to the information source) and the specific websites used for searching health information. Spontaneous reponses of participants were then categorized into different types of websites including general health-related websites; Doctissimo (a popular French website dedicated to general health mentioned by name by a large number of participants); social media; Wikipedia; institutional websites, Websites from health professional, patient association, scientific database; and others.

In 2010, questions were asked to assess why some people did not seek online health information (sufficiently informed, not interested in such information, better to ask these questions to a doctor, distrust in retrieved online information, do not think about it). Online health information seekers were also asked how often they seek online health information. However, since these variables were not assessed in 2014 and 2017, they were not analyzed in this article.

The list of the different variables and the corresponding questions asked each year are presented in Multimedia Appendix 1.

Ethical Consideration

According to French law, this study was not required to obtain the approval of a national ethics committee, as it is not legally considered research involving human beings and it relied on the collection of anonymous data only.

Statistical Analysis

Chi-square tests were used to compare the population characteristics over time, including gender, age, educational level, income, employment status, occupational category, and health status (chronic disease). Chi-square tests were also performed to assess the evolution of online health seekers over time and, between 2014 and 2017, of (1) health-related search topics, (2) sources of online health-related information, and (3) types of website used for these searches. The same tests were also used to describe the trends of the trust in health information found online and the potential impact on health management.

Multivariate logistic regression models were performed to investigate the profile of online health seekers versus non–health seekers (defined as the dependent variable), as well as the

evolution of the use of the internet for seeking health information over time. Independent variables included in the model were time, all sociodemographic variables, and health status. Interactions between all independent variables and time were assessed to evaluate potential differences in the profile of health seekers over time.

The same models including the same independent variables were performed to evaluate how individuals' characteristics and time are related with (1) the fact of not paying attention to the source of the health information found online, (2) trust in the information found online, and (3) the change in taking care of one's health.

Data were weighted by the number of telephone lines and eligible persons in the household. They were also adjusted to represent the French population structure (labor force survey 2008, 2012, and 2014) according to age, gender, educational level, region of residence, and level of urbanization [39].

Given that the maximum age limit was fixed at 75 years in 2014 and 2017 and the minimum at 18 years in 2017, participants aged 15 to 17 years in the 2010 and 2014 surveys and those aged 76 to 85 years in the 2010 survey were excluded in order to allow comparison over time. Given the low rate of missing values among the independent variables (ie, 1.3%), no specific imputation method was employed. Participants were therefore excluded if they had at least one missing value among the covariates used in the models. All tests of statistical significance were 2-sided, and the type I error was set at 5%. Statistical analyses were performed using Stata software version 13 (StataCorp LLC).

Results

The final sample comprised 15,277 individuals across the 3 time points, respectively 4141 in 2010, 4811 in 2014, and 6255 in 2017. A total of 581 participants were excluded because they were aged younger than 18 years or older than 75 years, and 202 because they had missing data.

Table 1 shows sociodemographic and economic characteristics and health status of included participants over time. Significant differences were found for age, educational level, income, occupational category, and chronic disease. Overall, figures showed that individuals tended to have higher educational levels and household incomes and more chronic diseases over time.

The evolution in internet access and use as a source of health information from 2010 to 2017 are presented in Figure 1. Although internet access increased steadily during this period, from 72.8% in 2010 to 92.4% in 2017 (an increase of 27%), the use of the internet as a source of health information rose between 2010 and 2014 (from 37.3% to 67.9%, $P<.001$) and decreased significantly in 2017 (60.3%, $P<.001$).

Table 1. Sociodemographic and economic characteristics of included participants over time (2010, 2014, and 2017; N=15,277).

Characteristic	Survey year 2010 (N=4141), n (%)[a]	Survey year 2014 (N=4881), n (%)[a]	Survey year 2017 (N=6255), n (%)[a]	*P* value[b]
Gender	—[c]	—	—	.91
Men	1814 (48.36)	2268 (48.74)	2831 (48.88)	—
Women	2327 (51.64)	2613 (51.26)	3424 (51.12)	—
Age in years	—	—	—	—
18-24	433 (12.08)	460 (10.62)	503 (10.48)	—
25-34	745 (17.61)	781 (16.12)	943 (17.28)	—
35-44	843 (20.02)	1006 (20.54)	1097 (18.55)	—
45-54	755 (19.65)	1015 (20.46)	1298 (19.63)	—
55-64	836 (17.68)	923 (18.42)	1308 (18.40)	—
65-75	529 (12.96)	696 (13.84)	1106 (15.68)	—
Educational level	—	—	—	<.001
Primary	1963 (58.01)	1994 (53.62)	2385 (48.68)	—
Secondary	791 (18.29)	980 (19.39)	1320 (20.12)	—
Post-secondary	1387 (23.70)	1907 (27.00)	2550 (31.20)	—
Income (€/CU[d])	—	—	—	<.001
0-1100	994 (34.89)	1109 (33.47)	1825 (30.76)	—
1101-1799	1264 (31.25)	1660 (28.83)	2042 (30.62)	—
≥1800	1606 (25.92)	1854 (31.63)	2106 (33.30)	—
Not willing to answer	277 (7.94)	258 (6.07)	282 (5.32)	—
Employment status	—	—	—	.53
Working	2440 (57.24)	2989 (57.66)	3619 (56.62)	—
Student	243 (6.60)	252 (5.57)	332 (6.60)	—
Unemployed	307 (8.41)	371 (9.56)	442 (8.68)	—
Retired	915 (20.96)	1035 (20.21)	1503 (21.16)	—
Other	236 (6.79)	234 (7.00)	359 (6.94)	—
Occupational category	—	—	—	.04
Farmers	70 (1.71)	71 (1.43)	116 (1.82)	—
Artisans	237 (6.15)	292 (6.50)	379 (6.96)	—
Executives	759 (14.99)	972 (14.59)	1181 (14.68)	—
Intermediate profession	1203 (25.37)	1244 (21.61)	1716 (23.69)	—
Employees	1106 (28.11)	1367 (30.60)	1727 (29.13)	—
Manual workers	766 (23.68)	935 (25.25)	1136 (23.72)	—
Chronic disease	—	—	—	<.001
No	3138 (76.28)	3135 (65.36)	3948 (64.50)	—
Yes	1003 (23.72)	1746 (34.64)	2307 (35.50)	—

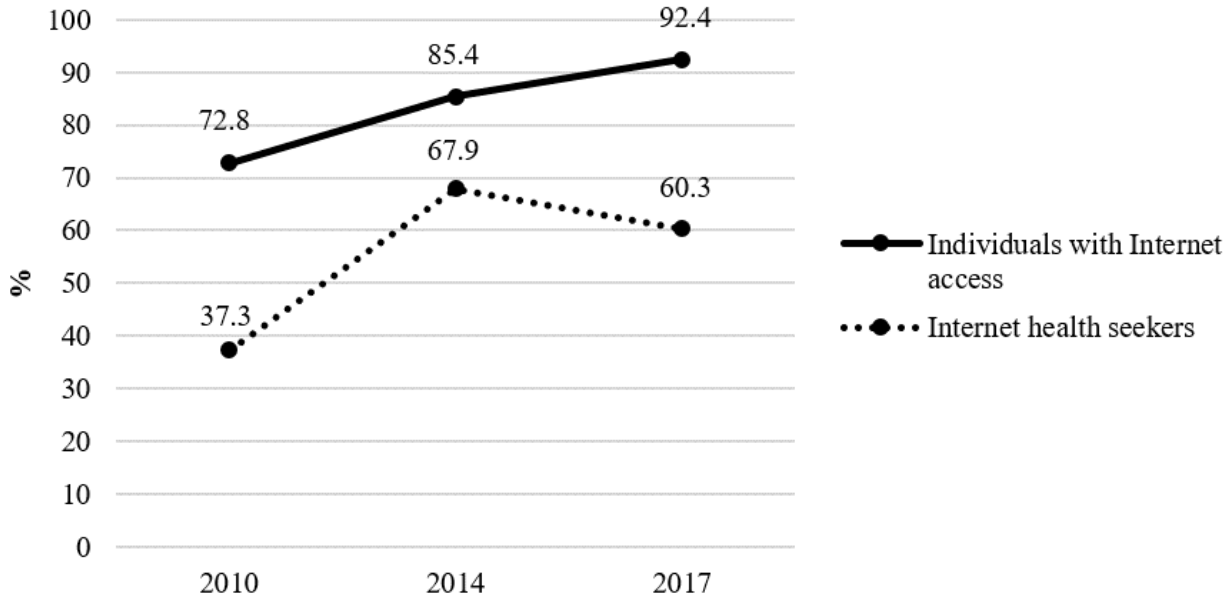
[a]Percentages are adjusted to represent the French population structure.

[b]On the basis of chi-square tests.

[c]Not applicable.

[d]CU: Household consumer units. One CU is attributed for the first adult in the household, 0.5 for other persons aged 14 years or older and 0.3 for children under 14 years.

Figure 1. Evolution in internet access and internet use as a source of health information from 2010 to 2017 (2010, N=4141; 2014, N=4881; 2017, N=6255).



Characteristics of online health information seekers across the 3 time points are presented in Table 2. Overall, health information seekers compared with nonseekers were more likely to be women, to be younger, to have a higher educational level, to have a higher household income, and to be executives. As regards the employment status, students and unemployed people were more likely to be health information seekers compared with the working group. Finally, individuals having a chronic disease were more likely to be health information seekers. Significant interaction with time was observed for age, educational level, and income. Results by the year of the survey were therefore explored (see Multimedia Appendix 2). Over time, the gap between generations seemed to be widening. In 2010, no differences in the use of the internet for seeking health information were found with people aged 45 to 54 years and younger, whereas this was the case in 2014 and 2017. As regards educational level, trends were comparable over time, but the odds of being online health seekers varied, in particular among those with a secondary education level. Finally, income was found to be less predictive of online seeking behavior after 2010, and particularly in 2014, since there was no more significant difference between the lower and intermediate income levels.

Table 2. Multivariate regression logistic models showing the association of internet use for seeking health information with time and individual characteristics (2010, N=4141; 2014, N=4881; 2017, N=6255).

Characteristic	OR[a] (95% CI)	*P* value[b]	*P* value of the time interaction[c]
Year	__[d]	—	—
2010	1	—	—
2014	4.23 (3.76-4.75)	<.001	—
2017	2.71 (2.43-3.02)	<.001	—
Gender	—	—	.14
Men	1	—	—
Women	1.81 (1.64-1.99)	<.001	—
Age in years	—	—	.0001
18-24	1	—	—
25-34	1.30 (1.05-1.61)	.02	—
35-44	0.90 (0.73-1.12)	.35	—
45-54	0.57 (0.46-0.70)	<.001	—
55-64	0.45 (0.36-0.57)	<.001	—
65-75	0.26 (0.19-0.34)	<.001	—
Educational level	—	—	.02
Primary	1	—	—
Secondary	1.62 (1.44-1.82)	<.001	—
Up to secondary	2.13 (1.89-2.41)	<.001	—
Income (€/CU[e])	—	—	.03
1-1100	1	—	—
1101-1799	1.26 (1.12-1.42)	<.001	—
≥1800	1.56 (1.37-1.78)	<.001	—
Not willing to answer	0.74 (0.60-0.91)	.004	—
Employment status	—	—	.49
Working	1	—	—
Student	1.67 (1.28-2.17)	<.001	—
Unemployed	1.22 (1.02-1.45)	.03	—
Retired	0.90 (0.74-1.09)	.27	—
Other	0.90 (0.73-1.12)	.35	—
Occupational category	—	—	.57
Executive	1	—	—
Intermediate profession	0.85 (0.74-0.98)	.03	—
Employee	0.70 (0.60-0.82)	<.001	—
Artisan	0.71 (0.57-0.88)	.002	—
Manual worker	0.52 (0.44-0.62)	<.001	—
Farmer	0.43 (0.30-0.61)	<.001	—
Chronic disease	—	—	.76
No	1	—	—
Yes	1.58 (1.43-1.75)	<.001	—

[a]OR: odds ratio.

[b]Multivariate logistic regression adjusted for year, gender, age, educational level, household income, employment status, occupational category, and

chronic disease.

[c]*P* value of the interaction term when adding interaction between each variable and year of survey in the logistic model.

[d]Not applicable.

[e]CU: consumption unit. One CU is attributed for the first adult in the household, 0.5 for other persons aged 14 years or older and 0.3 for children under 14 years.

Table 3 shows the topics of online health information research in 2014 and 2017, as well as the online sources of information and the type of websites used. Overall, the main topics searched online remained the same in 2014 and 2017 (ie, general health and illnesses; medical news and treatments; nutrition, weight gain, or eating disorders; and child health and illness). Nonetheless, while the percentage of respondents searching for information about general health and child health decreased over time, the percentage concerning nutrition-related topics remained constant. The most significant decreases were observed for searches about tobacco (–41%), alcohol (–70%), and cannabis and other drugs (–38%).

Overall, while the proportion of people using known health information websites and forums decreased, the proportion of those who did not pay attention to the source significantly increased (+12.1 percentage points). Thus, in 2017, this was the case for about 4 out of 10 internet health information seekers.

When focusing on the type of websites used for the last health-related internet searches by individuals who paid attention to the information source (48.7% in 2014 and 46.8% in 2017), figures indicates that between 2014 and 2017 general health-related websites remained the main source of information. Social media and commercial websites were the second source of information in 2014 as in 2017. And, in 2017, institutional websites were the third source. However, even if the visits to these institutional websites increased between 2014 and 2017, they remained at relatively low level, with only 8.1% of individuals declaring they used these sources.

Figure 2 shows the evolution of the trust in health information found on the internet and the change in taking care of one's health. Globally, the evolutions were symmetrical. While trust in the information increased from 2010 and 2014 ($P<.001$) and decreased slightly from 2014 to 2017 ($P=.048$), the change in health management rose between 2010 and 2014 ($P<.001$) but remained stable between 2014 and 2017. Nonetheless, the perception of health-related information found online was relatively positive, with at least 80% reporting that the most recent information found was trustworthy.

Table 3. Evolution between 2014 and 2017 among internet health seekers of (1) health-related search topics, (2) sources of online health-related information, and (3) types of websites used for internet health-related searches.

Search topics	Survey year 2014, n (%)[a]	Survey year 2017, n (%)[a]	*P* value[b]
(1) Health-related search topics	n=2036	n=3917	
General health and illnesses, medical news, and treatments[c]	1022 (71.80)	2159 (64.72)	.001
Nutrition, weight gain, or eating disorders	921 (45.03)	1741 (44.96)	.23
Child health and illness	625 (33.13)	975 (27.16)	<.001
Pregnancy or maternity	268 (14.63)	388 (11.78)	.02
Tobacco	240 (12.40)	245 (7.30)	<.001
Electronic cigarette	—[d]	191 (4.97)	—
Alcohol	197 (11.54)	133 (3.39)	<.001
Contraception and method to avoid pregnancy	164 (9.60)	237 (7.38)	.047
Sexually-related health risk	131 (7.41)	211 (6.59)	.37
Cannabis and other drugs	137 (7.12)	155 (4.36)	.001
(2) Sources of online health-related information	n=1396	n=3917	
Health information website	816 (55.75)	2035 (48.83)	<.001
Forum	432 (32.38)	801 (21.55)	<.001
active on the forum (n=1233)	20 (4.41)	40 (5.44)	.48
Did not pay attention to the source of the information	370 (27.63)	1475 (39.74)	<.001
(3) Type of websites used for internet health-related searches (among those who consulted information websites)	n=1478	n=2442	
General health-related website	443 (30.25)	937 (38.58)	<.001
Doctissimo[e]	371 (25.43)	733 (29.99)	.009
Social medias (YouTube, Facebook, blogs, TV...) and commercial websites	209 (13.70)	285 (11.86)	.16
Wikipedia	94 (6.16)	87 (3.16)	<.001
Institutional websites	103 (5.98)	212 (8.08)	.02
Websites from health professional, patient association, scientific database	66 (3.86)	161 (6.01)	.006
Other	18 (1.14)	6 (0.16)	.002

[a]Percentages are adjusted to represent the French population structure.

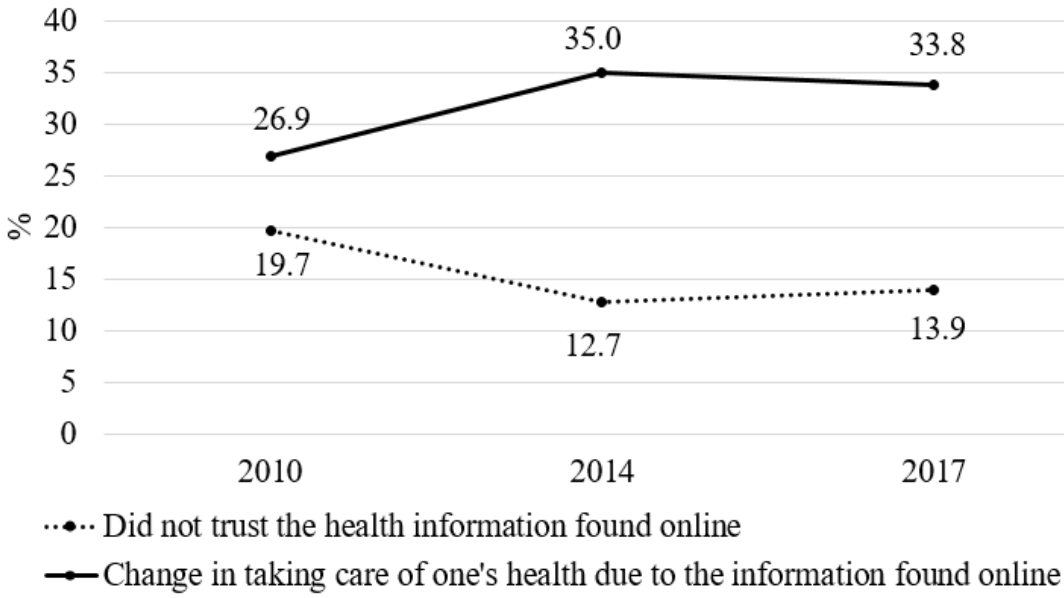
[b]On the basis of chi-square tests.

[c]For the “general health and illnesses, medical news, and treatments” topic, the question was asked to a subsample of participants to the study (n=4607).

[d]Not applicable.

[e]A popular French website dedicated to general health mentioned by name by a large number of participants.

Figure 2. Evolution of the trust in health information found on the internet and the change in taking care of one's health from 2010 to 2017 (2010, N=1707; 2014, N=3582; 2017, N=3965).



The results of multivariate models showing how individuals' characteristics and time are related to (1) the fact of not paying attention to the source of the health information found online, (2) trust in the information found online, and (3) the change in taking care of one's health are shown in Table 4. Not paying attention to the health-related information source significantly increased between 2014 and 2017. Such practice was associated with a higher probability of having lower education levels and being artisans or manual workers. In turn, students were more likely to pay attention to the source of their searches.

Trust in the health information found online and change in taking care of one's health both significantly increased in 2014 and slightly decreased in 2017. Regarding the associations with sociodemographic characteristics, only individuals with higher educational levels were more likely to trust the information found online. This trust strongly influenced the way of managing one's health (odds ratio 4.07). Individuals aged 25 to 34 years were more likely to have changed the way they managed their health due to the information found online, whereas those aged 65 to 75 years were less likely to do so. Finally, those having higher educational levels and higher incomes were also less influenced by the information found online.

Table 4. Multivariate logistic regression models showing the association of individuals' characteristics and time with (1) the fact of not paying attention to the source of the health information found online (2014, N=1422; 2017, N=3965), (2) trust in the information found online, and (3) the change in taking care of one's health (both 2010, N=1707; 2014, N=3582; 2017, N=3965).

Characteristic	Not paying attention to information source (N=5313), OR[a] (95% CI)	*P* value[b]	Trust in the last health information found online, OR (95% CI)	*P* value[b]	Change in taking care of one's health, OR (95% CI)	*P* value[b]
Year						
2010	__[c]	—	1	—	1	—
2014	1	—	1.66 (1.35-2.04)	<.001	1.41 (1.19-1.65)	<.001
2017	1.78 (1.50-2.11)	<.001	1.53 (1.27-1.85)	<.001	1.32 (1.13-1.55)	<.001
Trust in health information found online						
Not trustworthy	—	—	—	—	1	—
Trustworthy	—	—	—	—	4.06 (3.26-5.06)	<.001
Gender						
Men	1	—	1	—	1	—
Women	1.02 (0.87-1.19)	.80	0.93 (0.78-1.10)	.40	0.89 (0.79-1.01)	.07
Age in years						
18-24	1	—	1	—	1	—
25-34	0.96 (0.68-1.34)	.79	0.99 (0.72-1.37)	.96	1.37 (1.05-1.78)	.02
35-44	0.93 (0.66-1.32)	.69	1.32 (0.94-1.85)	.11	1.20 (0.92-1.58)	.18
45-54	1.12 (0.79-1.59)	.53	1.31 (0.92-1.85)	.13	1.21 (0.92-1.60)	.17
55-64	1.27 (0.87-1.85)	.21	1.10 (0.74-1.64)	.63	0.77 (0.57-1.05)	.11
65-75	1.39 (0.86-2.25)	.18	0.91 (0.55-1.51)	.72	0.70 (0.46-1.05)	.08
Educational level						
Primary	1	—	1	—	1	—
Secondary	0.66 (0.54-0.80)	<.001	1.09 (0.88-1.33)	.43	0.86 (0.73-1.01)	.06
Up to secondary	0.57 (0.47-0.70)	<.001	1.31 (1.05-1.63)	.03	0.87 (0.74-1.02)	.08
Income (€/CU[d])						
0-1100	1	—	1	—	1	—
1101-1799	0.90 (0.74-1.10)	.31	1.15 (0.93-1.43)	.19	0.82 (0.69-0.96)	.01
≥1800	0.91 (0.74-1.13)	.39	1.28 (1.02-1.62)	.04	0.79 (0.67-0.94)	.007
Not willing to answer	1.05 (0.71-1.55)	.81	0.91 (0.61-1.36)	.65	1.13 (0.84-1.53)	.42
Employment status						
Working	1	—	1	—	1	—
Student	0.65 (0.44-0.96)	.03	1.31 (0.90-1.90)	.16	1.25 (0.93-1.67)	.14
Unemployed	0.95 (0.72-1.25)	.71	1.21 (0.88-1.65)	.25	1.54 (1.23-1.92)	<.001
Retired	0.91 (0.66-1.25)	.56	1.32 (0.92-1.90)	.13	1.33 (1.02-1.73)	.04
Other	0.87 (0.62-1.22)	.43	1.02 (0.71-1.47)	.91	1.58 (1.20-2.09)	.001
Occupational category						
Executive	1	—	1	—	1	—
Intermediate profession	1.05 (0.85-1.29)	.66	1.04 (0.84-1.28)	.73	1.54 (0.83-2.84)	.17
Employee	1.19 (0.94-1.51)	.15	1.14 (0.88-1.47)	.33	1.31 (0.73-2.35)	.36
Artisan	1.91 (1.35-2.71)	<.001	1.39 (0.90-2.14)	.14	1.41 (0.79-2.50)	.25
Manual worker	1.48 (1.13-1.95)	.005	0.95 (0.71-1.27)	.75	1.52 (0.85-2.70)	.16
Farmer	1.28 (0.65-2.52)	.47	0.78 (0.38-1.58)	.49	1.60 (0.89-2.87)	.12

Characteristic	Not paying attention to information source (N=5313), OR[a] (95% CI)	*P* value[b]	Trust in the last health information found online, OR (95% CI)	*P* value[b]	Change in taking care of one's health, OR (95% CI)	*P* value[b]
Chronic disease						
No	1	—	1	—	1	—
Yes	0.46 (0.30-0.72)	.001	1.08 (0.91-1.28)	.40	0.97 (0.85-1.10)	.63

[a]OR: odds ratio.

[b]Multivariate logistic regression adjusted for year, gender, age, educational level, household income, employment status, occupational category, and chronic disease.

[c]Not applicable.

[d]CU: consumption unit. One CU is attributed for the first adult in the household, 0.5 for other persons aged 14 years or older, and 0.3 for children under 14 years.

Discussion

Principal Findings and Interpretation

This was one of the first studies describing the evolution of online health information seeking in a European country, based on nationally representative time-series survey data. We observed an increase of the use of the internet for health-related information between 2010 and 2014 but a decrease between 2014 and 2017. In parallel, trust in the health information found online followed the same trend, thus suggesting a potential relationship between these 2 variables. Indeed, the growing phenomenon of misinformation and fake news might restrain citizens from using the internet for health-related information [40]. They might prefer consulting a health professional or just avoid looking for health information online [41]. However, promoting access to trustworthy information online represents a key lever to help people managing their health. Therefore, a growing body of research exists on the potential of interventions designed to develop eHealth literacy [42], which has been described as a necessary competence to mitigate health inequalities [43].

The rise in distrust was paradoxically complemented by a higher proportion of respondents reporting not paying attention to the information sources. This might be explained by the fact that it is often difficult to identify the source of information and assess its credibility. Apart from institutional health-related websites (eg, the website of the Ministry of Health, the website of a local hospital), determining the online source of information has become challenging and even frustrating [44]. On the other hand, general websites are easy to access and consult, while institutional websites remain less consulted, even if a slight increase was reported between 2014 and 2017 in our study. The complexity and density of the information they provide might explain their scarce use, despite their trustworthiness. Those who trusted more online contents were respondents having higher educational levels and incomes, which might be explained by the fact that they are supposed to have more developed eHealth literacy and technological skills, allowing them to better evaluate the accuracy of the information retrieved online [13,45]. These citizens were also less likely to change the management of their health based on health-related information found online, differentiating their information-seeking behavior from their health behavior [46].

Sociodemographic characteristics and health conditions of online health information seekers were similar to those found in previous studies [14,18,47]; being a woman, being younger, being an executive, having a higher educational level, having a higher household income, and having a chronic disease were all associated with use of the internet for health information seeking. Interestingly, our results showed that unemployed people were more likely to be health information seekers compared with other groups such as working or retired people. This result is in contrast with previous research [48,49] but in line with other studies [19,50] showing that this point is controversial and might depend on the specific characteristics of unemployed citizens when they are not taken into account in a model, like the fact of being a woman or having an illness, or rather on the country's unemployment rate or medical care coverage for these people. What can be said with more confidence is that higher educational levels are associated with higher use of the internet for health information seeking, independently from being employed or not, since eHealth literacy skills are higher among people having a diploma [51]. In our study, students, independently of the level of education, were more likely to be health information seekers compared with other people (eg, working, retired), which can be explained by the fact that they are used to seeking online information in general as part of their study curriculum.

Trends about searched topics were similar across time periods, with general health and illness being the most searched terms together with nutrition, weight gain, or eating disorders and child health and illness. The prevalence of these topics is explained by the characteristics of likely online health information seekers (ie, women aged 35 to 54 years). Gender differences have been frequently reported as relevant for health information seeking, including topics of interest [52]. Decreasing interest in topics like tobacco, alcohol, or other drug consumption might be explained by the fact that users prefer browsing the web for general health-related information, while for more specific problems like addiction, they prefer other sources of information. This is in line with the trustworthiness of online information and the risk of encountering fake news for sensitive topics like drug consumption and is also documented in previous research [53].

Strengths and Limitations

Strengths of this study included the use of large datasets from nationally representative surveys including the general population with various sociodemographic characteristics. The time-series design was also important to robustly assess the evolution of online health information-seeking patterns in the French population.

This study is not without limitations. First, while being based on large samples, the response rates were between 48.5% and 61%, which means that selection bias cannot be excluded and that some specific population groups like homeless people or immigrants were likely to be underrepresented. Second, as a population-based study specific to France, these results are not generalizable to other countries, although online health information-seeking behavior patterns are supposed to be similar in most of the European countries. Third, data on trust in the information found online were only available for the subsample of health information seekers: this prevented the evaluation of the association between trust and decrease in the use of the internet for health information seeking. Fourth, Health Barometer surveys do not report on important aspects related to online health information seeking such as technical skills and eHealth literacy. Finally, the reliability of some answers may be affected by a memory bias, and other data concerning online health information seeking were not assessed like the frequency of use and the use of social media or mobile apps providing health tips and information. A complete picture of online health information-seeking behaviors might benefit from more data on digital health use in general.

Conclusions

Our results showed a rapid growth in internet use in the 2010 to 2014 period with a decrease in the year 2017, in parallel with a decreasing trust in the quality and reliability of information found online. The trends in the use of and trust in online health information need to be constantly monitored, but our findings already underlined the need for alternative trustworthy sources of information on the internet. In particular, it is recommended that official health institutions promote initiatives to help citizens navigate health-related information available on the internet. These initiatives might range from interventions aimed at promoting citizens' eHealth literacy, such as educational programs, to official websites and online portals providing reliable but simple and usable information. Effective interventions should combine the popularity and accessibility of general health-related websites with the authority of the institutional websites.

Acknowledgments

We would like to acknowledge all participants in the 2010, 2014, and 2017 Health Barometers. We would also like to thank François Beck for coordinating the 2010 Health Barometer survey and Abdelkrim Zeghnoun for the statistical support. The 2010, 2014, and 2017 Health Barometers were funded by Santé publique France, the national public health agency (and formerly the National Institute for Health Prevention and Education, or INPES).

Authors' Contributions

The 2017 Health Barometer group participated in the design and implementation of the 2017 Health Barometer survey. JBR coordinated the 2010, 2014, and 2017 Health Barometers. JBR was responsible for the design and protocol of the study. PD conducted the literature review, performed the statistical analyses, and drafted the manuscript. IM, VNT, AJS, and JBR were involved in the interpretation of results and critically reviewed the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
List of the variables used in the study and the corresponding questions asked by year of survey.
[DOCX File , 13 KB - jmir_v23i4e18799_app1.docx]

Multimedia Appendix 2
Multivariate regression logistic models showing the association of internet use for seeking health information with individual characteristics by the year of the survey (2010, N=4141; 2014, N=4881; 2017, N=6255).
[DOCX File , 14 KB - jmir_v23i4e18799_app2.docx]

References

1. European Citizens' Digital Health Literacy. European Commission. 2014. URL: https://ec.europa.eu/commfrontoffice/publicopinion/flash/fl_404_en.pdf [accessed 2020-02-03]
2. Cline RJ, Haynes KM. Consumer health information seeking on the Internet: the state of the art. Health Educ Res 2001 Dec;16(6):671-692. [doi: 10.1093/her/16.6.671] [Medline: 11780707]

3. Santana S, Lausen B, Bujnowska-Fedak M, Chronaki CE, Prokosch H, Wynn R. Informed citizen and empowered citizen in health: results from an European survey. BMC Fam Pract 2011;12:20 [FREE Full text] [doi: 10.1186/1471-2296-12-20] [Medline: 21496309]
4. Coulter A, Ellins J, Swain D, Clarke A, Heron P, Rasul F. Assessing the quality of information to support people in making decisions about their health and healthcare. Oxford: Picker Institute Europe URL: http://www.picker.org/assets/content/pdf/Project_Reports/Health-information-quality-web-version-FINAL.pdf [accessed 2020-02-03]
5. Griffiths F, Lindenmeyer A, Powell J, Lowe P, Thorogood M. Why are health care interventions delivered over the internet? A systematic review of the published literature. J Med Internet Res 2006;8(2):e10 [FREE Full text] [doi: 10.2196/jmir.8.2.e10] [Medline: 16867965]
6. Berger M, Wagner TH, Baker LC. Internet use and stigmatized illness. Soc Sci Med 2005 Oct;61(8):1821-1827. [doi: 10.1016/j.socscimed.2005.03.025] [Medline: 16029778]
7. Joury A, Joraid A, Alqahtani F, Alghamdi A, Batwa A, Pines JM. The variation in quality and content of patient-focused health information on the Internet for otitis media. Child Care Health Dev 2018 Mar;44(2):221-226. [doi: 10.1111/cch.12524] [Medline: 28913967]
8. Lawrentschuk N, Abouassaly R, Hewitt E, Mulcahy A, Bolton DM, Jobling T. Health information quality on the internet in gynecological oncology: a multilingual evaluation. Eur J Gynaecol Oncol 2016 Aug;37(4):478-483. [Medline: 29894070]
9. Novillo-Ortiz D, Hernández-Pérez T, Saigí-Rubió F. Availability of information in Public Health on the Internet: an analysis of national health authorities in the Spanish-speaking Latin American and Caribbean countries. Int J Med Inform 2017 Apr;100:46-55 [FREE Full text] [doi: 10.1016/j.ijmedinf.2017.01.013] [Medline: 28241937]
10. Norman CD, Skinner HA. eHealth literacy: essential skills for consumer health in a networked world. J Med Internet Res 2006 Jun;8(2):e9 [FREE Full text] [doi: 10.2196/jmir.8.2.e9] [Medline: 16867972]
11. Hsu W, Chiang C, Yang S. The effect of individual factors on health behaviors among college students: the mediating effects of eHealth literacy. J Med Internet Res 2014;16(12):e287 [FREE Full text] [doi: 10.2196/jmir.3542] [Medline: 25499086]
12. Mitsutake S, Shibata A, Ishii K, Oka K. Associations of eHealth literacy with health behavior among adult internet users. J Med Internet Res 2016 Jul;18(7):e192 [FREE Full text] [doi: 10.2196/jmir.5413] [Medline: 27432783]
13. Neter E, Brainin E. eHealth literacy: extending the digital divide to the realm of health information. J Med Internet Res 2012 Jan;14(1):e19 [FREE Full text] [doi: 10.2196/jmir.1619] [Medline: 22357448]
14. Li J, Theng Y, Foo S. Predictors of online health information seeking behavior: changes between 2002 and 2012. Health Informatics J 2015 Aug 10;22(4):804-814. [doi: 10.1177/1460458215595851] [Medline: 26261218]
15. Cotten SR, Gupta SS. Characteristics of online and offline health information seekers and factors that discriminate between them. Soc Sci Med 2004 Nov;59(9):1795-1806. [doi: 10.1016/j.socscimed.2004.02.020] [Medline: 15312915]
16. Powell J, Inglis N, Ronnie J, Large S. The characteristics and motivations of online health information seekers: cross-sectional survey and qualitative interview study. J Med Internet Res 2011 Feb;13(1):e20 [FREE Full text] [doi: 10.2196/jmir.1600] [Medline: 21345783]
17. Rice RE. Influences, usage, and outcomes of Internet health information searching: multivariate results from the Pew surveys. Int J Med Inform 2006 Jan;75(1):8-28. [doi: 10.1016/j.ijmedinf.2005.07.032] [Medline: 16125453]
18. Ghweeba M, Lindenmeyer A, Shishi S, Abbas M, Waheed A, Amer S. What predicts online health information-seeking behavior among Egyptian adults? A cross-sectional study. J Med Internet Res 2017 Jun 22;19(6):e216 [FREE Full text] [doi: 10.2196/jmir.6855] [Medline: 28642216]
19. Nölke L, Mensing M, Krämer A, Hornberg C. Sociodemographic and health-(care-)related characteristics of online health information seekers: a cross-sectional German study. BMC Public Health 2015;15:31 [FREE Full text] [doi: 10.1186/s12889-015-1423-0] [Medline: 25631456]
20. Xesfingi S, Vozikis A. eHealth literacy: in the quest of the contributing factors. Interact J Med Res 2016 May 25;5(2):e16 [FREE Full text] [doi: 10.2196/ijmr.4749] [Medline: 27226146]
21. Din HN, McDaniels-Davidson C, Nodora J, Madanat H. Profiles of a health information-seeking population and the current digital divide: cross-sectional analysis of the 2015-2016 California Health Interview Survey. J Med Internet Res 2019 May 14;21(5):e11931 [FREE Full text] [doi: 10.2196/11931] [Medline: 31094350]
22. Miller LMS, Bell RA. Online health information seeking: the influence of age, information trustworthiness, and search challenges. J Aging Health 2012 Apr;24(3):525-541. [doi: 10.1177/0898264311428167] [Medline: 22187092]
23. Massey PM. Where do U.S. adults who do not use the internet get health information? Examining digital health information disparities from 2008 to 2013. J Health Commun 2016;21(1):118-124. [doi: 10.1080/10810730.2015.1058444] [Medline: 26166484]
24. Choi N. Relationship between health service use and health information technology use among older adults: analysis of the US National Health Interview Survey. J Med Internet Res 2011 Apr 20;13(2):e33 [FREE Full text] [doi: 10.2196/jmir.1753] [Medline: 21752784]
25. Yoon H, Jang Y, Vaughan PW, Garcia M. Older adults' internet use for health information: digital divide by race/ethnicity and socioeconomic status. J Appl Gerontol 2020 Jan;39(1):105-110. [doi: 10.1177/0733464818770772] [Medline: 29661052]

26. Javanmardi M, Noroozi M, Mostafavi F, Ashrafi-Rizi H. Internet usage among pregnant women for seeking health information: a review article. Iran J Nurs Midwifery Res 2018;23(2):79-86 [FREE Full text] [doi: 10.4103/ijnmr.IJNMR_82_17] [Medline: 29628953]
27. Sayakhot P, Carolan-Olah M. Internet use by pregnant women seeking pregnancy-related information: a systematic review. BMC Pregnancy Childbirth 2016;16:65 [FREE Full text] [doi: 10.1186/s12884-016-0856-5] [Medline: 27021727]
28. Freeman JL, Caldwell PHY, Bennett PA, Scott KM. How adolescents search for and appraise online health information: a systematic review. J Pediatr 2018 Apr;195:244-255. [doi: 10.1016/j.jpeds.2017.11.031] [Medline: 29398062]
29. Park E, Kwon M. Health-related internet use by children and adolescents: systematic review. J Med Internet Res 2018 Apr 03;20(4):e120 [FREE Full text] [doi: 10.2196/jmir.7731] [Medline: 29615385]
30. Beck F, Richard J, Nguyen-Thanh V, Montagni I, Parizot I, Renahy E. Use of the internet as a health information resource among French young adults: results from a nationally representative survey. J Med Internet Res 2014;16(5):e128 [FREE Full text] [doi: 10.2196/jmir.2934] [Medline: 24824164]
31. Claridy MD, Hudson MM, Caplan L, Mitby PA, Leisenring W, Smith SA, et al. Patterns of Internet-based health information seeking in adult survivors of childhood cancer. Pediatr Blood Cancer 2018 Dec;65(5):e26954 [FREE Full text] [doi: 10.1002/pbc.26954] [Medline: 29350454]
32. Jabson JM, Patterson JG, Kamen C. Understanding health information seeking on the internet among sexual minority people: cross-sectional analysis from the health information national trends survey. JMIR Public Health Surveill 2017 Jun 19;3(2):e39 [FREE Full text] [doi: 10.2196/publichealth.7526] [Medline: 28630036]
33. Ek S, Eriksson-Backa K, Niemelä R. Use of and trust in health information on the Internet: a nationwide eight-year follow-up survey. Inform Health Soc Care 2013 Sep;38(3):236-245. [doi: 10.3109/17538157.2013.764305] [Medline: 23514041]
34. Wangberg S, Andreassen H, Kummervold P, Wynn R, Sørensen T. Use of the internet for health purposes: trends in Norway 2000-2010. Scand J Caring Sci 2009 Dec;23(4):691-696. [doi: 10.1111/j.1471-6712.2008.00662.x] [Medline: 19473317]
35. Questionnaire of the Health Barometer 2017. Santé publique France. URL: https://www.santepubliquefrance.fr/content/download/197532/2364898 [accessed 2020-03-04]
36. Methods of the Health Barometer 2017. Santé publique France. URL: https://www.santepubliquefrance.fr/content/download/153340/2186141 [accessed 2020-03-04]
37. Kish L. A procedure for objective respondent selection within the household. J Am Stat Assoc 1949 Sep;44(247):380-387. [doi: 10.1080/01621459.1949.10483314]
38. Richard J, Andler R, Gautier A, Guignard R, Leon C, Beck F. Effects of using an overlapping dual-frame design on estimates of health behaviors: a French general population telephone survey. J Surv Stat Methodol 2016 Nov 17;5(2):smw028. [doi: 10.1093/jssam/smw028]
39. Labour force survey 2008, 2012, 2016. French National Institute for Statistics and Economic Studies. URL: https://www.insee.fr/en/metadonnees/source/serie/s1223/ [accessed 2020-03-04]
40. Sbaffi L, Rowley J. Trust and credibility in web-based health information: a review and agenda for future research. J Med Internet Res 2017 Jun 19;19(6):e218 [FREE Full text] [doi: 10.2196/jmir.7579] [Medline: 28630033]
41. Jacobs W, Amuta AO, Jeon KC. Health information seeking in the digital age: an analysis of health information seeking behavior among US adults. Cogent Soc Sci 2017 Mar 13;3(1):1302785. [doi: 10.1080/23311886.2017.1302785]
42. Watkins I, Xie B. eHealth literacy interventions for older adults: a systematic review of the literature. J Med Internet Res 2014;16(11):e225 [FREE Full text] [doi: 10.2196/jmir.3318] [Medline: 25386719]
43. Azzopardi-Muscat N, Sørensen K. Towards an equitable digital public health era: promoting equity through a health literacy perspective. Eur J Public Health 2019 Oct 01;29(Supplement_3):13-17 [FREE Full text] [doi: 10.1093/eurpub/ckz166] [Medline: 31738443]
44. LaValley SA, Kiviniemi MT, Gage-Bouchard EA. Where people look for online health information. Health Info Libr J 2016 May 21;34(2):146-155. [doi: 10.1111/hir.12143] [Medline: 27207817]
45. Paige SR, Krieger JL, Stellefson ML. The influence of eHealth literacy on perceived trust in online health communication channels and sources. J Health Commun 2017 Jan;22(1):53-65. [doi: 10.1080/10810730.2016.1250846] [Medline: 28001489]
46. Greyson DL, Johnson JL. The role of information in health behavior: a scoping study and discussion of major public health models. J Assn Inf Sci Tec 2015 Jun 24;67(12):2831-2841. [doi: 10.1002/asi.23392]
47. Gonzalez M, Sanders-Jackson A, Emory J. Online health information-seeking behavior and confidence in filling out online forms among Latinos: a cross-sectional analysis of the California Health Interview Survey, 2011-2012. J Med Internet Res 2016 Jul 04;18(7):e184 [FREE Full text] [doi: 10.2196/jmir.5065] [Medline: 27377466]
48. Gallagher S, Doherty DT. Searching for health information online: characteristics of online health seekers. J Evid Based Med 2009 May;2(2):99-106. [doi: 10.1111/j.1756-5391.2009.01021.x] [Medline: 21348996]
49. Cohen RA, Adams PF. Use of the internet for health information: United States, 2009. NCHS Data Brief 2011 Jul(66):1-8 [FREE Full text] [Medline: 22142942]
50. Andreassen HK, Bujnowska-Fedak MM, Chronaki CE, Dumitru RC, Pudule I, Santana S, et al. European citizens' use of E-health services: a study of seven countries. BMC Public Health 2007;7:53 [FREE Full text] [doi: 10.1186/1471-2458-7-53] [Medline: 17425798]

51. Holt KA, Karnoe A, Overgaard D, Nielsen SE, Kayser L, Røder ME, et al. Differences in the level of electronic health literacy between users and nonusers of digital health services: an exploratory survey of a group of medical outpatients. Interact J Med Res 2019 Apr 05;8(2):e8423 [FREE Full text] [doi: 10.2196/ijmr.8423] [Medline: 30950809]
52. Baumann E, Czerwinski F, Reifegerste D. Gender-specific determinants and patterns of online health information seeking: results from a representative German health survey. J Med Internet Res 2017 Apr 04;19(4):e92 [FREE Full text] [doi: 10.2196/jmir.6668] [Medline: 28377367]
53. Fox S. Americans search online for mental health, insurance, and drug information. Washington: Pew Internet and American Life Project; 2003 Jul 16. URL: https://www.pewresearch.org/internet/2003/07/16/americans-search-online-for-mental-health-insurance-and-drug-information/ [accessed 2021-03-15]

Abbreviations

CU: consumption unit
INPES: French Institute for Prevention and Health Education

Edited by G Eysenbach; submitted 31.03.20; peer-reviewed by E Neter, H Scholten, A Gulliver; comments to author 12.06.20; revised version received 04.11.20; accepted 24.01.21; published 14.04.21.

Please cite as:
Ducrot P, Montagni I, Nguyen Thanh V, Serry AJ, Richard JB
Evolution of Online Health-Related Information Seeking in France From 2010 to 2017: Results From Nationally Representative Surveys
J Med Internet Res 2021;23(4):e18799
URL: https://www.jmir.org/2021/4/e18799
doi:10.2196/18799
PMID:33851927

Original Paper

Computer Mouse Movements as an Indicator of Work Stress: Longitudinal Observational Field Study

Nicolas Banholzer[1], MSc; Stefan Feuerriegel[1], Prof Dr; Elgar Fleisch[2,3], Prof Dr; Georg Friedrich Bauer[4], Prof Dr; Tobias Kowatsch[2,3], PhD

[1]Department of Management, Technology and Economics, ETH Zurich, Zurich, Switzerland

[2]Centre for Digital Health Interventions, Department of Management, Technology and Economics, ETH Zurich, Zurich, Switzerland

[3]Centre for Digital Health Interventions, Institute of Technology Management, University of St. Gallen, Zurich, Switzerland

[4]Center for Salutogenesis, Epidemiology, Biostatistics and Prevention Institute, University of Zurich, Zurich, Switzerland

Corresponding Author:
Tobias Kowatsch, PhD
Centre for Digital Health Interventions
Department of Management, Technology and Economics
ETH Zurich
Weinbergstr 56/58
Zurich, 8092
Switzerland
Phone: 41 799129133
Email: tkowatsch@ethz.ch

Abstract

Background: Work stress affects individual health and well-being. These negative effects could be mitigated through regular monitoring of employees' stress. Such monitoring becomes even more important as the digital transformation of the economy implies profound changes in working conditions.

Objective: The goal of this study was to investigate the association between computer mouse movements and work stress in the field.

Methods: We hypothesized that stress is associated with a speed-accuracy trade-off in computer mouse movements. To test this hypothesis, we conducted a longitudinal field study at a large business organization, where computer mouse movements from regular work activities were monitored over 7 weeks; the study included 70 subjects and 1829 observations. A Bayesian regression model was used to estimate whether self-reported acute work stress was associated with a speed-accuracy trade-off in computer mouse movements.

Results: There was a negative association between stress and the two-way interaction term of mouse speed and accuracy (mean −0.32, 95% highest posterior density interval −0.58 to −0.08), which means that stress was associated with a speed-accuracy trade-off. The estimated association was not sensitive to different processing of the data and remained negative after controlling for the demographics, health, and personality traits of subjects.

Conclusions: Self-reported acute stress is associated with computer mouse movements, specifically in the form of a speed-accuracy trade-off. This finding suggests that the regular analysis of computer mouse movements could indicate work stress.

(J Med Internet Res 2021;23(4):e27121) doi:10.2196/27121

KEYWORDS

work stress; psychological stress; stress indicator; computer mouse movements; human-computer interactions

Introduction

Stress in the workplace is responsible for over 120,000 deaths and US $187 billion in annual health care spending in the United States [1]. To mitigate this burden, work stress must be monitored and managed. The need for workplace stress management increases even further as the digital transformation of the economy implies profound changes in working conditions [2]. At the same time, digital transformation offers opportunities for better stress management. Human-computer interactions with ubiquitous digital devices could be used for real-time monitoring of work-related stress. In particular, it has been

shown that the computer mouse responds to changes in muscular activity as a result of stress [3-6]. Thus, previous studies have investigated the association between stress and the use of the computer mouse [7-11], for instance, by analyzing computer mouse movements [8,10,11]. However, the evidence from these studies is, so far, based on lab experiments using artificially designed computer tasks. Hence, it remains unclear whether an association between stress and the use of the computer mouse can also be observed in the field.

For this study, we hypothesized that there is an association between stress and computer mouse movements. Our hypothesized association is based on the theory of neuromotor noise [12-16]. Stress, induced by time pressure or multitasking, leads to higher neuromotor noise [15,16], which is the noise in control signals steering motor movements. Lower signal-to-noise ratios and limited capacity to process information lead to adaptive movement behavior [12]. For instance, if subjects are required to execute fast movements, then neuromotor noise will lead to greater variability in the direction of movement [15]. The reason for this is that high or low execution speeds induce neuromotor noise, which makes it more difficult to hit the intended target of the movement accurately and requires more or fewer corrections, respectively, along the trajectory [13,14]. That is, the accuracy of the movement has to adjust relative to the movement speed.

In short, the previous literature suggests that stress induces neuromotor noise, resulting in a speed-accuracy trade-off in motor movements. This trade-off is particularly documented in rapid aimed movements [13,14]; based on this, we can expect that it also applies to computer mouse movements. We tested our hypothesis with data from a longitudinal observational field study that included 70 subjects and 1829 observations. Thereby, we collected computer mouse movements and self-reported stress levels from employees during their regular office work for 7 weeks. Using a Bayesian regression model, we present findings that support our hypothesis that work stress is characterized by a speed-accuracy trade-off in computer mouse movements.

Methods

Study Design

A 7-week longitudinal field study was conducted at a large European technology company. The company's human resources director asked 496 employees from different service units (ie, accounting, human resources, information technology, marketing, quality management, logistics, and business development) to participate through an email invitation. The invitation described the study's objective of understanding the association between computer mouse movements and work stress.

Subjects were not offered financial incentives. However, they were invited to a debriefing event at the end of the study, where the aggregated results were presented. Further, their self-reports were made available to them through graphical diagrams so they could monitor their stress levels over the course of the study.

Among all invited employees, 71 subjects decided to participate. They installed our study software by clicking on a link in the invitation. When subjects first opened the study software, a tutorial explained how the software was used to report stress. During the 7-week study period, the study software asked subjects twice a day to report their stress level. The timings were randomly triggered by our software, namely, once between 9 AM and 11 AM and once between 2 PM and 4 PM. Prior to these self-reports, our software recorded all computer mouse movements for 30 minutes. If subjects were not using their computer at that time (eg, due to a meeting), then no data were recorded.

Data about subjects' computer mouse movements and self-reports were securely transferred to a server at the organization, from which they were gathered by our research team to perform subsequent analyses. At the beginning of the study, subjects were further asked to report their sociodemographics (ie, age, gender, and education), behavioral attributes regarding health and nutrition (ie, sports, nutrition, smoking, and drinking habits), and expression of the big five personality traits as measured by an established inventory [17]. All variables are described in Table 1.

Table 1. Variables and descriptions.

Variable	Description
Target variable	
Valence	Self-reported valence on a scale from 1 (low) to 7 (high)
Arousal	Self-reported arousal on a scale from 1 (low) to 7 (high)
Stress	Dummy with 1 if valence <4 and arousal >4 (stress), 0 otherwise (no stress)
Mouse movements	
Speed	Distance computer mouse is moved divided by the duration of the movement
Accuracy	Proportion of mouse events where the movement direction remained equal along the x-axis and y-axis
Mouse events	
Clicks	Proportion of mouse tracks with clicks in a recording
Wheels	Proportion of mouse tracks with wheels in a recording
Recording time	
Weekday	Categorical {1: Monday, 2: Tuesday, 3: Wednesday, 4: Thursday, 5: Friday, 6: Saturday and Sunday}
Daytime	Dummy with 1 if recording was in the morning, 0 otherwise (in the afternoon)
Sociodemographics	
Age	Subject age
Gender	Dummy with 1 if male, 0 otherwise (female)
Education	Dummy with 1 if university degree, 0 otherwise (ie, high school or lower)
Health and nutrition	
Sport	Hours of sport per week
Nutrition	Number of fruits or vegetables consumed per day
Alcohol	Categorical {1: never, 2: 2-4 times per month, 3: 2-3 times per week, 4: more than 4 times per week}
Smoking	Categorical {1: daily, 2: occasionally, 3: not anymore, 4: never smoked}
Personality traits	The big five personality traits, each measured on a scale from 1 (low expression of the trait) to 10 (high expression of the trait), based on an established inventory [17]

Processing Computer Mouse Movements

A Java application was developed to record computer mouse movements (ie, timestamp and x- and y-coordinates) and mouse events (ie, movement, click, and wheel). The application was built on the Windows operating system's standard software drivers with a sample rate of approximately 125 Hz. Computer mouse movements were recorded for 30 minutes and processed in the following way. Each recording was split into separate trajectories, where a trajectory started with a mouse movement and ended with a different mouse event (ie, a click or wheel). Thereby, trajectories were only considered if their duration was between 1 and 10 seconds. This approach was beneficial, as it omitted trajectories that were extremely short or included temporary phases where the mouse was not moving. For each trajectory, two variables were computed: (1) mouse speed, which is the average movement speed, and (2) mouse accuracy, which is the proportion of mouse events where the direction of the movement remained equal along the x- and y-axes (ie, the proportion of times the movement direction was *not* corrected). Both variables were then averaged over all trajectories. These provided the features that were inserted into our regression model.

Mouse speed was computed as the total distance the mouse moved between the start time t=1 of a trajectory and its end time T divided by the trajectory's total duration T. Hence, this yielded the following:

Mouse accuracy is the relative frequency of how often the movement in x- and y-directions was *not* changed. It is formalized by the following:

where the variable $eqdir_t$ indicates whether the movement in both x- and y-directions remained equal at time t. It returns a value of 1 if this is the case and 0 otherwise. Formally, it is specified by the following:

Accordingly, the larger the accuracy value is, the less the movement direction was altered. If the value for accuracy is 1, then the movement direction was never altered, and if the value for accuracy is 0, then the movement direction was always

altered. In other words, the more accurate movement was the one with fewer corrections. This directly relates our measure of accuracy to the theory of neuromotor noise, which predicts more corrections as the movement speed is increased.

The proportion of direction changes is commonly used as a measure of accuracy in related work [18,19]. Another measure for accuracy is the deviation from an optimal trajectory [19,20]. However, the theoretical model underlying the speed-accuracy trade-off predicts that higher movement speed leads to more corrective submovements [13,14], not necessarily to a larger deviation from the optimal line between the start and end point of the trajectory. For that reason, we specifically chose the proportion of direction changes as our measure of accuracy in this study.

Stress Measurement

Acute stress was measured according to the circumplex model of affect [21]. This model relates affective states to two underlying neurophysiological systems: valence, a pleasure-displeasure continuum, and arousal or alertness [22]. Both were collected using self-assessment manikins [23] on a 7-point Likert scale, with a value of 1 referring to a very negative valence (very low arousal) and a value of 7 indicating a very positive valence (very high arousal). Acute stress was then defined as a combination of low valence and high arousal, which has been shown to be related to work stressors in empirical research [24]. Specifically, stress is encoded as a dichotomous variable that equals 1 if subjects reported low valence and high arousal (ie, valence below the neutral midpoint of 4 and arousal above the neutral midpoint of 4) and 0 otherwise. Hence, our encoding translates into an analysis that focuses on distinguishing negative stress from positive or no stress.

Statistical Analysis

A logistic regression model was estimated with stress as the dichotomous outcome variable and with features from computer mouse movements as the independent variables. The model is specified as follows:

where $stress_{ik}$ is the dichotomous outcome variable for subject i=1,...,M and recording k=1,...,N. Subject-specific variation in average stress levels is captured by the varying intercept α_i. Note that subject-specific characteristics such as age or gender could explain between-subject variation of average stress levels, but beyond that, time-varying variables such as computer mouse movements are needed to explain within-subject variation of stress levels over time. The association of mouse speed and accuracy with stress is estimated by β_1 to β_3. In particular, the two-way interaction between mouse speed and accuracy (β_3) tests whether a speed-accuracy trade-off in computer mouse movements is associated with stress. Note that mouse speed and accuracy were centered and scaled by their empirical mean and standard deviation. By centering both variables, the sign of β_3 indicates the direction of the speed-accuracy trade-off. That is, a negative sign of β_3 would indicate that a simultaneous increase in mouse speed and decrease in mouse accuracy or a simultaneous decrease in mouse speed and increase in mouse accuracy is associated with a higher probability of stress.

Further independent variables were included in the above regression model as part of the sensitivity analysis. For instance, to control for mouse usage, we computed the number of events where the mouse was clicked or wheeled. Note that access to other human-computer interactions (eg, keyboard strokes) was not granted in this study due to privacy concerns.

Model Estimation

A Bayesian approach was used for model estimation. Compared to classical statistics, the Bayesian approach requires the specification of priors for all model parameters. When choosing flat priors, the classical and Bayesian approaches are the same. However, when choosing a Bayesian prior (eg, a normal prior), the results are different, and sign errors are less frequent with a Bayesian prior [25]. In other words, a Bayesian approach is less prone to making wrong claims about the sign of a parameter. In our study setting, this would most likely give more conservative estimates; hence, a Bayesian approach was used for data analysis. We chose weakly informative priors for all model parameters, thereby following recommendations on the choice of priors [26]. Our priors are as follows:

The model was estimated with Markov chain Monte Carlo using four chains. Each chain performed 2000 iterations divided into 1000 iterations for a warm-up and 1000 iterations for sampling. Samples were drawn with the No-U-Turn Sampler [27]. Thereby, it was ensured that all Markov chains converged successfully so that inference could be performed. In the Results section, we report the posterior distribution, the posterior mean, and the 95% highest posterior density interval (HPDI) of the estimated parameters.

Statistical analysis was performed with the programming language R, version 4.0.2 (The R Foundation), and the probabilistic programming language Stan, version 2.21.0 [28], using the interface provided by the R package brms, version 2.13.5 [29].

Data Inclusion and Exclusion

All participants deciding to participate were included in the study (ie, no additional inclusion or exclusion criteria were applied). Our raw data contained 2029 recordings from 71 subjects. The number of recordings per subject varied due to absences or because the subjects decided to stop participating. Further, recordings were excluded when no computer mouse movements were recorded (5 recordings), the recorded computer mouse movements contained tracking errors (ie, incorrect time

stamps) (92 recordings), or when the recordings contained less than 10 computer mouse trajectories (200 recordings). This led to the removal of 297 recordings from 62 subjects—between 1 and 12 per subject—and the exclusion of 1 subject from the study.

Data and Code Availability

Preprocessed data and a script to replicate all model results are provided [30]. Raw data may be used to identify individual study participants (eg, because mouse movements can be very specific to a person or may reveal sensitive information, such as passwords, when using a software keyboard) and, thus, cannot be made available; this decision was made by the research team and the institutional review board that evaluated the study.

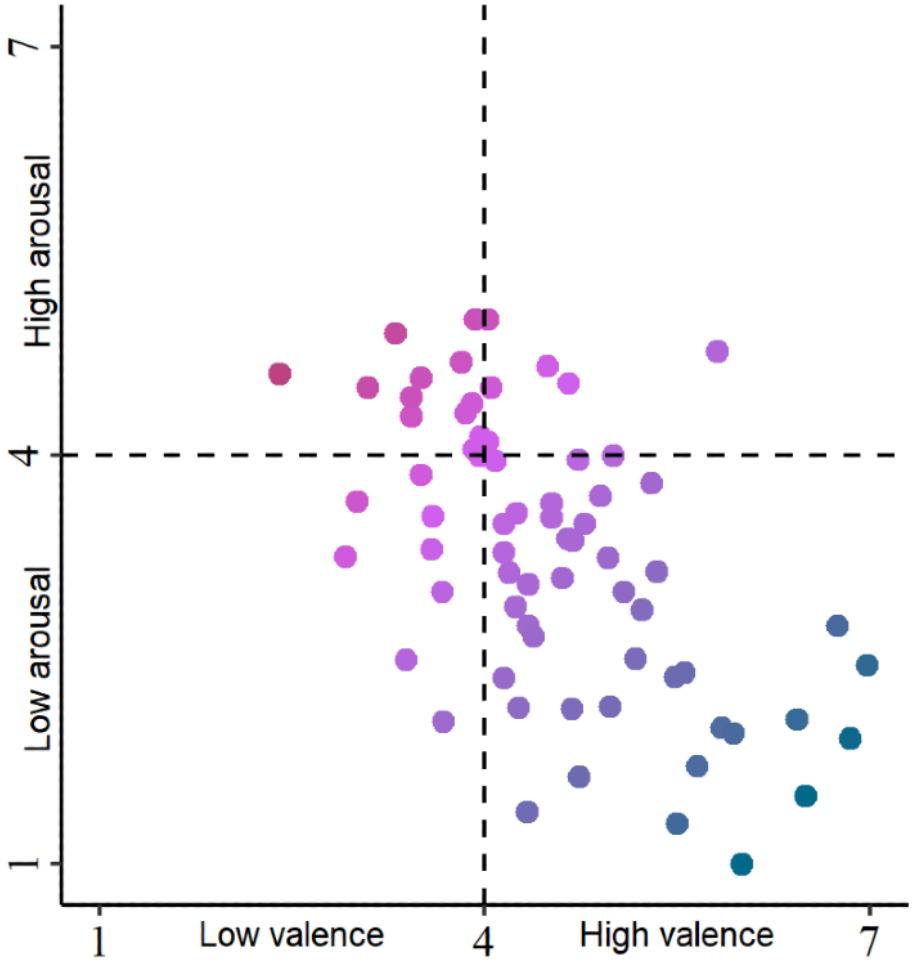
Results

Subject Statistics

Our results are based on 70 subjects and 1829 recordings (mean 26.13, SD 14.33). Subjects were between 20 and 61 years old, with a median age of 39.5 years (IQR 31.0-49.0). Further, 46% (32/70) of the participants were female, and 59% (41/70) held a university degree; all others had high school diplomas or lower. Recordings were roughly balanced by daytime hours (951/1829, 52.0% in the morning and 878/1829, 48.0% in the afternoon) and weekdays (329/1829, 18.0% to 384/1829, 21.0% per weekday and 18/1829, 1% on the weekend).

Both valence and arousal varied across subjects (see Figure 1). Average valence per subject was slightly above the neutral midpoint (mean 4.53, SD 0.98), and average arousal was slightly below the neutral midpoint (mean 3.28, SD 1.02). When averaged over the study period, a combination of low valence and high arousal (the top-left quadrant in Figure 1) was observed in 12 out of the 70 (17%) subjects. Applying our encoding of stress following the circumplex model of affect [21], 185 out of the 1829 self-reports (10.1%) were classified as stressful.

Figure 1. Perceived valence and arousal by subject. Shown are the average self-reported valence and arousal levels by subject in the field study. Red points indicate high levels and blue points indicate low levels of average stress.



Association Between Stress and Computer Mouse Movements

It was hypothesized that stress is characterized by a speed-accuracy trade-off. This trade-off is illustrated in Figure 2. When subjects perceived no stress, computer mouse movements were typically not characterized by a speed-accuracy trade-off. In contrast to that, when subjects perceive stress, computer mouse movements were typically characterized by a trade-off where the mouse was moved quickly but less

accurately (ie, many direction changes) or slowly but more accurately (ie, few direction changes). Descriptives for average mouse speed and accuracy are provided in Multimedia Appendix 1, Figure S1.

The estimated parameters of mouse speed and accuracy were as follows. The individual parameters of mouse speed (β_1) and accuracy (β_2) were not significant based on the observation that the 95% HPDIs include zero (see Figure 3). However, the parameter for the two-way interaction between speed and accuracy (β_3) was significant (mean −0.32, 95% HPDI −0.58 to −0.08). On average, a simultaneous 1 SD increase in mouse speed and 1 SD decrease in mouse accuracy, or vice versa, was associated with a change in the odds for perceiving stress by 1.53. In other words, work stress was characterized by a speed-accuracy trade-off.

Figure 4 depicts the partial dependence of both mouse speed and mouse accuracy on the probability of perceiving stress. Based on the plot, two findings can be derived. First, stress was more likely when there was a speed-accuracy trade-off. Second, this trade-off seemed slightly more prevalent for low mouse speed and high mouse accuracy, as indicated by a higher share of observations in the lower-right corner. This means that although both directions of the trade-off are present in our data, subjects perceiving stress were slightly more frequently increasing accuracy at the cost of speed.

Figure 2. Illustrative examples of the speed-accuracy trade-off in computer mouse movements. Shown are typical computer mouse movements (blue dot: beginning of movement; red dot: click) from the field study. Circles correspond to recordings at 125 Hz. When subjects perceived no stress, computer mouse movements were typically not characterized by a speed-accuracy trade-off. When subjects perceived stress, computer mouse movements were typically characterized by a speed-accuracy trade-off. Mouse speed and accuracy were standardized to indicate the direction of the trade-off; that is, high speed (+) and low accuracy (−) or low speed (−) and high accuracy (+).

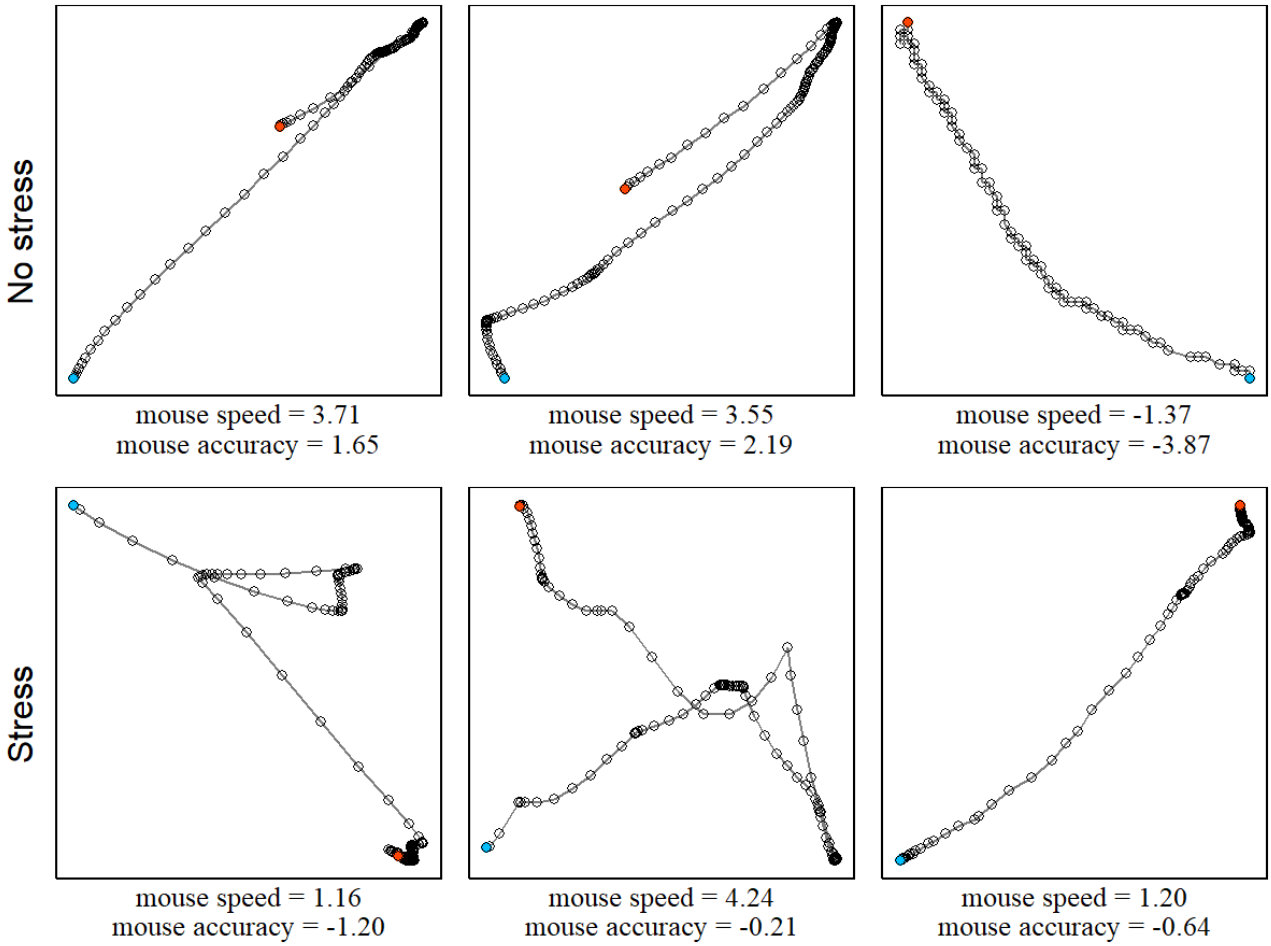


Figure 3. Association between work stress and computer mouse movements. Shown is the estimated effect (posterior and prior density and mean as solid and dashed grey lines, respectively, and 95% highest posterior density interval as shaded area) of mouse speed (β_1), mouse accuracy (β_2), and the two-way interaction between mouse speed and accuracy (β_2).

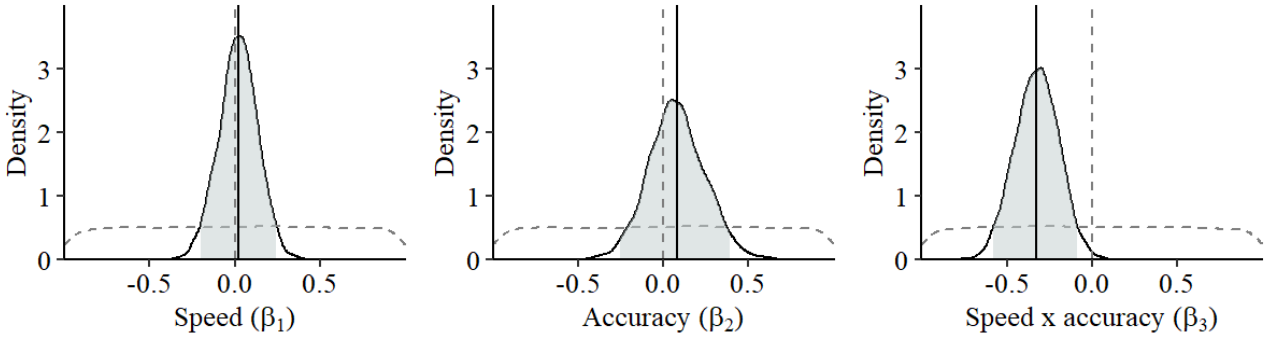
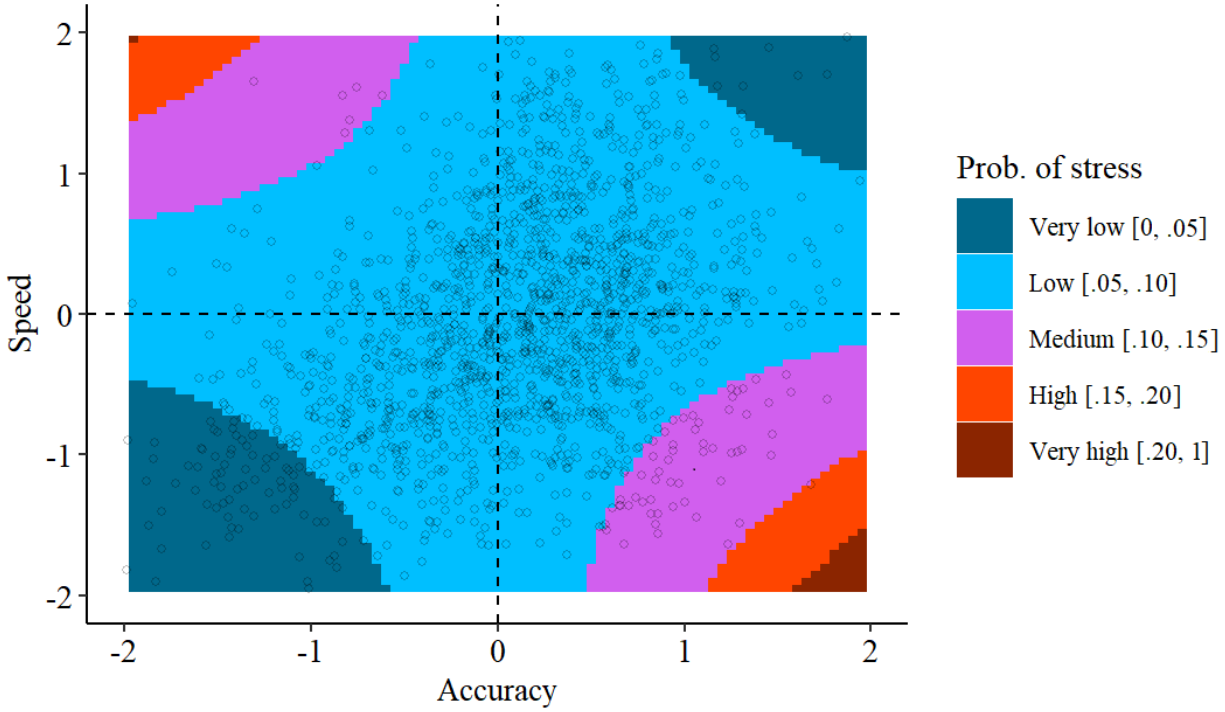


Figure 4. Probability (Prob.) of perceived stress based on mouse speed and accuracy. Shown is the partial dependence of stress on mouse speed and accuracy in the range of −2 SD to +2 SD. Red areas indicate high levels and blue areas indicate low levels of stress.



Sensitivity Analysis

The sensitivity of the estimated parameters was assessed in the following ways. First, different processing of the data led to conclusive findings. In the above analysis, recordings were removed when fewer than 10 computer mouse trajectories were counted over 30 minutes. When varying this number, the estimated parameter of the mouse speed-accuracy trade-off remained stable (see Multimedia Appendix 1, Figure S2). Similarly, the maximum duration for a trajectory was set to 10 seconds. When varying the maximum duration from 5 to 20 seconds, the estimated parameter of the mouse speed-accuracy trade-off also remained stable (see Multimedia Appendix 1, Figure S3). Furthermore, recordings from 2 subjects revealed unusually low mouse accuracy (see Multimedia Appendix 1, Figure S1). Excluding all recordings from these subjects slightly reduced the size of the estimated parameter for the trade-off (mean −0.22, 95% HPDI −0.42 to −0.03).

Second, the sensitivity of the estimated parameter for the speed-accuracy trade-off was assessed with respect to the inclusion of varying slopes for computer mouse movement variables and additional controls, such as mouse events and sociodemographics. Including varying slopes or adding more controls led to comparable estimates for the parameter of the mouse speed-accuracy trade-off (see Multimedia Appendix 1, Figure S4).

Third, the association of computer mouse movements with valence, arousal, and a discrete measure of stress (defined as arousal – valence + 6) was estimated. Results from Poisson regressions with the same model specification showed no significant associations but a tendency that arousal and the discrete measure of stress were negatively associated with the speed-accuracy trade-off (see Multimedia Appendix 1, Figure S5). However, the results from the regression with the discrete measure of stress as the outcome should be interpreted with caution, as an arousal level of 7 and valence level of 4 would

result in the equivalent level of stress as an arousal level of 5 and a valence level of 2, whereas only the second self-report would be labeled as stress according to the circumplex model of affect.

Fourth, the possibility of selection bias was investigated, with a statistical comparison between those subjects with few (n≤10) and many (n>10) recordings. The proportion of recordings with stress from subjects with few recordings (6/43, 14%) was higher than the proportion of recordings with stress from subjects with many recordings (188/1986, 9.5%). However, the difference was not statistically significant (χ^2_1=0.5, *P*=.47). This result suggests that participation intensity was not significantly related to stress. Other than that, it could not be investigated whether individuals outside this study were more or less stressed than the subjects participating in this study.

Discussion

Principal Findings

The goal of this study was to examine whether computer mouse movements indicate work stress. Data from a 7-week longitudinal field study supported the hypothesis. Despite the heterogeneity of computer tasks and the resulting complexity of computer mouse movements, we found a significant association with work stress. That is, work stress was characterized by a speed-accuracy trade-off in computer mouse movements.

Comparison With Prior Work

This is the first study to infer stress from the computer mouse in the field (ie, at the workplace). In prior work, lab studies were conducted to investigate the association between stress and the use of the computer mouse [7-11]. In these lab studies, subjects performed artificial tasks (eg, point-and-click tasks) in a controlled environment. In contrast to that, our data were collected unobtrusively while subjects were performing office work in a real-world environment. On the one hand, this made data processing and analysis challenging. On the other hand, it provided us with the unique opportunity to present the first empirical evidence as to whether stress is associated with computer mouse movements in the field.

A drawback of our field study in comparison to the lab studies is that we are not able to estimate a causal link. The reason is that there are potentially unmeasured confounders. In particular, computer mouse movements as well as stress may depend on the difficulty of the task, with more difficult tasks resulting in higher levels of stress. In the lab, it is possible to control which task is performed, whereas this is not possible in the field without obtrusive monitoring of tasks. However, precisely because unobtrusive and continuous monitoring of tasks is not feasible in the field, computer mouse movements may be a good proxy for how stressful a task is perceived and may thus provide an indirect way to measure stress.

Benefits

Monitoring of computer mouse movements provide a number of benefits for stress management in the workplace. Most office work involves computer tasks; as such, computer mouse movement data are readily available. Unlike other forms of stress monitoring, computer mouse movements present a viable tool for monitoring stress at scale because they can be collected in an unobtrusive fashion and continuously over time [10]. The latter becomes important when offering on-demand stress management interventions by organizations and for monitoring their effectiveness [31]. It is also possible to monitor stress by monitoring physiological changes (eg, heart rate variability or skin conductivity) through wearable devices. However, when introduced by employers, the broad usage of physiological data in a corporate context raises issues regarding their acceptance and legitimacy [32]. When compared to such physiological stress measurements, many employees might consider the measurement of computer mouse movements as a clearly work-related behavior and as a less intrusive and more legitimate monitoring method at work. As computer mouse movements are bound to currently performed work, their measurement will trigger a more balanced action on the part of employees to mitigate work stress: both reducing their own receptivity to stress and improving the underlying working conditions, as it is also recommended by the European Union [33]. Thus, the measurement of computer mouse movements offers a valuable, complementary approach to physiological measurements.

Limitations

Our study also has limitations. First, our work constitutes an observational study with an explanatory analysis of the data. As a consequence, a causal interpretation of the estimated association is precluded. Second, computer mouse movements were only linked to the presence of stress, which was defined according to the circumplex model of affect [21]. The severity of stress could not be assessed due to the low prevalence of high levels of stress. Third, computer mouse movements were only linked to acute stress. The association of computer mouse movements with chronic stress is subject to future work. Fourth, the outcome of this study was psychological stress, which was measured based on self-reports. It is unclear if, and to what extent, psychological and physiological measures of stress are alternative or complementary by nature [34]. Thus, collecting physiological data from wearable devices to monitor stress [35,36] could be used to validate the association with computer mouse movements. Fifth, the sources of stress were not identified, which is important for managing stress. However, other work suggests that human-computer interactions also correlate with workplace stressors [37]. Sixth, the determinants for the directions of the speed-accuracy trade-off were not explored. This would probably require a different research setting, most likely a controlled lab experiment, in order to investigate what causes subjects to increase speed at the cost of accuracy, or vice versa.

Conclusions

To summarize, the findings of this study suggest that the computer mouse can be used to infer work stress. These findings could be combined with findings from other forms of human-computer interactions (eg, computer trackpads [38] or keyboard strokes [39]) in order to develop digital tools for detecting stress.

Acknowledgments

We would like to thank Andreas Filler for developing the software that allowed us to perform the computer mouse movement and self-report recordings. NB and SF acknowledge funding from the Swiss National Science Foundation outside of this study.

Conflicts of Interest

EF and TK are affiliated with the Center for Digital Health Interventions, a joint initiative of the Department of Management, Technology and Economics at ETH Zurich and the Institute of Technology Management at the University of St. Gallen, which is funded in part by the Swiss health insurer CSS. EF and TK are also cofounders of Pathmate Technologies, a university spin-off company that creates and delivers digital clinical pathways. However, Pathmate Technologies was not involved in the study described in this paper.

Multimedia Appendix 1
Supplementary figures for descriptives and sensitivity analyses.
[DOCX File , 142 KB - jmir_v23i4e27121_app1.docx]

References

1. Goh J, Pfeffer J, Zenios SA. The relationship between workplace stressors and mortality and health costs in the United States. Manage Sci 2016 Feb;62(2):608-628. [doi: 10.1287/mnsc.2014.2115]
2. Stacey N, Ellwood P, Bradbrook S, Reynolds J, Williams H, Lye D. Foresight on New and Emerging Occupational Safety and Health Risks Associated with Digitalization by 2025. Luxembourg: European Agency for Safety and Health at Work; 2018. URL: https://osha.europa.eu/pt/publications/foresight-new-and-emerging-occupational-safety-and-health-risks-associated [accessed 2021-03-16]
3. Wahlström J, Hagberg M, Johnson P, Svensson J, Rempel D. Influence of time pressure and verbal provocation on physiological and psychological reactions during work with a computer mouse. Eur J Appl Physiol 2002 Jul;87(3):257-263. [doi: 10.1007/s00421-002-0611-7] [Medline: 12111287]
4. Wahlström J, Lindegård A, Ahlborg G, Ekman A, Hagberg M. Perceived muscular tension, emotional stress, psychological demands and physical load during VDU work. Int Arch Occup Environ Health 2003 Oct;76(8):584-590. [doi: 10.1007/s00420-003-0454-5] [Medline: 12898271]
5. Laursen B, Jensen B, Garde AH, Jørgensen AH. Effect of mental and physical demands on muscular activity during the use of a computer mouse and a keyboard. Scand J Work Environ Health 2002 Aug;28(4):215-221 [FREE Full text] [doi: 10.5271/sjweh.668] [Medline: 12199422]
6. Visser B, De Looze MP, De Graaff MP, Van Dieën J. Effects of precision demands and mental pressure on muscle activation and hand forces in computer mouse tasks. Ergonomics 2004 Feb 05;47(2):202-217. [doi: 10.1080/00140130310001617967] [Medline: 14660213]
7. Hernandez J, Paredes P, Roseway A, Czerwinski M. Under pressure: Sensing stress of computer users. In: Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems (CHI '14). 2014 Presented at: 32nd Annual ACM Conference on Human Factors in Computing Systems (CHI '14); April 26-May 1, 2014; Toronto, ON p. 51-60. [doi: 10.1145/2556288.2557165]
8. Sun D, Paredes P, Canny J. MouStress: Detecting stress from mouse motion. In: Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems (CHI '14). 2014 Presented at: 32nd Annual ACM Conference on Human Factors in Computing Systems (CHI '14); April 26-May 1, 2014; Toronto, ON p. 61-70. [doi: 10.1145/2556288.2557243]
9. Wang J, Fu EY, Ngai G, Leong HV, Huang MX. Detecting stress from mouse-gaze attraction. In: Proceedings of the 34th ACM/SIGAPP Symposium on Applied Computing (SAC '19). 2019 Presented at: 34th ACM/SIGAPP Symposium on Applied Computing (SAC '19); April 8-12, 2019; Limassol, Cyprus p. 692-700. [doi: 10.1145/3297280.3297345]
10. Freihaut P, Göritz AS. Using the computer mouse for stress measurement – An empirical investigation and critical review. Int J Hum Comput Stud 2021 Jan;145:102520. [doi: 10.1016/j.ijhcs.2020.102520]
11. Pepa L, Sabatelli A, Ciabattoni L, Monteriu A, Lamberti F, Morra L. Stress detection in computer users from keyboard and mouse dynamics. IEEE Trans Consum Electron 2021 Feb;67(1):12-19. [doi: 10.1109/tce.2020.3045228]
12. Fitts PM. The information capacity of the human motor system in controlling the amplitude of movement. J Exp Psychol 1954;47(6):381-391. [doi: 10.1037/h0055392]
13. Meyer DE, Abrams RA, Kornblum S, Wright CE, Smith JE. Optimality in human motor performance: Ideal control of rapid aimed movements. Psychol Rev 1988 Jul;95(3):340-370. [doi: 10.1037/0033-295x.95.3.340] [Medline: 3406245]
14. Meyer DE, Smith JEK, Kornblum S, Abrams RA, Wright CE. Speed-accuracy tradeoffs in aimed movements: Toward a theory of rapid voluntary action. In: Proceedings of the Thirteenth International Symposium on Attention and Performance.

Mahwah, NJ: Lawrence Erlbaum Associates, Inc; 1990 Presented at: Thirteenth International Symposium on Attention and Performance; June 27-July 2, 1988; Arc-et-Senans, France p. 173-226. [doi: 10.4324/9780203772010-6]
15. Van Galen GP, Van Huygevoort M. Error, stress and the role of neuromotor noise in space oriented behaviour. Biol Psychol 2000 Jan;51(2-3):151-171. [doi: 10.1016/s0301-0511(99)00037-x]
16. Van Gemmert AWA, Van Galen GP. Stress, neuromotor noise, and human performance: A theoretical perspective. J Exp Psychol 1997;23(5):1299-1313. [doi: 10.1037/0096-1523.23.5.1299]
17. Rammstedt B, John OP. Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. J Res Pers 2007 Feb;41(1):203-212. [doi: 10.1016/j.jrp.2006.02.001]
18. Grimes M, Jenkins J, Valacich J. Exploring the effect of arousal and valence on mouse interaction. In: Proceedings of the 34th International Conference on Information Systems (ICIS) in Human-Computer Interaction. 2013 Presented at: 34th International Conference on Information Systems (ICIS) in Human-Computer Interaction; December 15-18, 2013; Milano, Italy.
19. Yamauchi T, Xiao K. Reading emotion from mouse cursor motions: Affective computing approach. Cogn Sci 2018 Apr;42(3):771-819. [doi: 10.1111/cogs.12557] [Medline: 29131372]
20. Hibbeln MT, Jenkins JL, Schneider C, Valacich J, Weinmann M. How is your user feeling? Inferring emotion through human-computer interaction devices. MIS Q 2017 Jan 1;41(1):1-21. [doi: 10.25300/misq/2017/41.1.01]
21. Russell JA. A circumplex model of affect. J Pers Soc Psychol 1980;39(6):1161-1178. [doi: 10.1037/h0077714]
22. Posner J, Russell JA, Peterson BS. The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. Dev Psychopathol 2005 Nov 1;17(03):715-734. [doi: 10.1017/s0954579405050340]
23. Bradley MM, Lang PJ. Measuring emotion: The self-assessment manikin and the semantic differential. J Behav Ther Exp Psychiatry 1994 Mar;25(1):49-59. [doi: 10.1016/0005-7916(94)90063-9]
24. Van Katwyk PT, Fox S, Spector PE, Kelloway EK. Using the Job-Related Affective Well-Being Scale (JAWS) to investigate affective responses to work stressors. J Occup Health Psychol 2000;5(2):219-230. [doi: 10.1037/1076-8998.5.2.219]
25. Gelman A, Tuerlinckx F. Type S error rates for classical and Bayesian single and multiple comparison procedures. Comput Stat 2000 Sep 1;15(3):373-390. [doi: 10.1007/s001800000040]
26. Stan Development Team. Prior choice recommendations. GitHub. 2020. URL: https://github.com/stan-dev/stan/wiki/Prior-Choice-Recommendations [accessed 2021-03-15]
27. Hoffman MD, Gelman A. The No-U-Turn Sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. J Mach Learn Res 2014 Jan;15(1):1593-1623 [FREE Full text] [doi: 10.5555/2627435.2638586]
28. Carpenter B, Gelman A, Hoffman MD, Lee D, Goodrich B, Betancourt M, et al. Stan: A probabilistic programming language. J Stat Softw 2017;76(1):1-32 [FREE Full text] [doi: 10.18637/jss.v076.i01]
29. Bürkner PC. brms: An R package for Bayesian multilevel models using Stan. J Stat Softw 2017;80(1):1-28 [FREE Full text] [doi: 10.18637/jss.v080.i01]
30. Banholzer N, Feuerriegel S, Fleisch E, Bauer GF, Kowatsch T. Data and code: Computer mouse movements as an indicator of work stress: Longitudinal observational field study. OSF. 2021. URL: https://doi.org/10.17605/OSF.IO/HE3F2 [accessed 2021-03-15]
31. Ganster DC, Rosen CC. Work stress and employee health. J Manage 2013 Feb 19;39(5):1085-1122. [doi: 10.1177/0149206313475815]
32. Lupton D. Health promotion in the digital era: A critical commentary. Health Promot Int 2015 Mar;30(1):174-183. [doi: 10.1093/heapro/dau091] [Medline: 25320120]
33. Leka S, Jain A. Mental Health in the Workplace in Europe.: EU Compass for Action on Mental Health and Well-being; 2017. URL: https://www.lisboninstitutegmh.org/assets/files/Mental%20health%20in%20the%20workplace%20in%20Europe.pdf [accessed 2021-03-15]
34. Tams S, Hill K, Guinea A, Thatcher J, Grover V. NeuroIS—Alternative or complement to existing methods? Illustrating the holistic effects of neuroscience and self-reported data in the context of technostress research. J Assoc Inf Syst 2014 Oct;15(10):723-753. [doi: 10.17705/1jais.00374]
35. Smets E, Rios Velazquez E, Schiavone G, Chakroun I, D'Hondt E, De Raedt W, et al. Large-scale wearable data reveal digital phenotypes for daily-life stress detection. NPJ Digit Med 2018;1:67 [FREE Full text] [doi: 10.1038/s41746-018-0074-9] [Medline: 31304344]
36. Natarajan A, Pantelopoulos A, Emir-Farinas H, Natarajan P. Heart rate variability with photoplethysmography in 8 million individuals: A cross-sectional study. Lancet Digit Health 2020 Dec;2(12):e650-e657. [doi: 10.1016/s2589-7500(20)30246-6]
37. Eijckelhof BHW, Huysmans MA, Blatter BM, Leider PC, Johnson PW, van Dieën JH, et al. Office workers' computer use patterns are associated with workplace stressors. Appl Ergon 2014 Nov;45(6):1660-1667. [doi: 10.1016/j.apergo.2014.05.013] [Medline: 25005311]
38. Goel R, An M, Alayrangues H, Koneshloo A, Lincoln ET, Paredes PE. Stress tracker-detecting acute stress from a trackpad: Controlled study. J Med Internet Res 2020 Oct 23;22(10):e22743 [FREE Full text] [doi: 10.2196/22743] [Medline: 33095176]
39. Vizer LM, Zhou L, Sears A. Automated stress detection using keystroke and linguistic features: An exploratory study. Int J Hum Comput Stud 2009 Oct;67(10):870-886. [doi: 10.1016/j.ijhcs.2009.07.005]

Abbreviations

HPDI: highest posterior density interval

Edited by G Eysenbach; submitted 13.01.21; peer-reviewed by A Göritz, J Zhang; comments to author 05.02.21; revised version received 16.02.21; accepted 25.02.21; published 02.04.21.

Please cite as:
Banholzer N, Feuerriegel S, Fleisch E, Bauer GF, Kowatsch T
Computer Mouse Movements as an Indicator of Work Stress: Longitudinal Observational Field Study
J Med Internet Res 2021;23(4):e27121
URL: https://www.jmir.org/2021/4/e27121
doi:10.2196/27121
PMID:33632675

Original Paper

The Feasibility of Using Instagram Data to Predict Exercise Identity and Physical Activity Levels: Cross-sectional Observational Study

Sam Liu[1*], PhD; Megan Perdew[1*], MSc; Alexander Lithopoulos[1*], PhD; Ryan E Rhodes[1,2*], PhD

[1]School of Exercise Science, Physical and Health Education, University of Victoria, Victoria, BC, Canada

[2]Department of Psychology, University of Victoria, Victoria, BC, Canada

[*]all authors contributed equally

Corresponding Author:
Sam Liu, PhD
School of Exercise Science, Physical and Health Education
University of Victoria
PO Box 3010 STN CSC
Victoria, BC, V8W 3N4
Canada
Phone: 1 2507218392
Email: samliu@uvic.ca

Abstract

Background: Exercise identity is an important predictor for regular physical activity (PA). There is a lack of research on the potential mechanisms or antecedents of identity development. Theories of exercise identity have proposed that investment, commitment and self-referential (eg, I am an exerciser) statements, and social activation (comparison, support) may be crucial to identity development. Social media may be a potential mechanism to shape identity.

Objective: The objectives of this study were to (1) explore whether participants were willing to share their Instagram data with researchers to predict their lifestyle behaviors; (2) examine whether PA-related Instagram uses (ie, the percentage of PA-related Instagram posts, fitness-related followings, and the number of likes received on PA-related posts) were positively associated with exercise identity; and (3) evaluate whether exercise identity mediates the relationship between PA-related Instagram use and weekly PA minutes.

Methods: Participants (18-30 years old) were asked to complete a questionnaire to evaluate their current levels of exercise identity and PA. Participants' Instagram data for the past 12 months before the completion of the questionnaire were extracted and analyzed with their permission. Instagram posts related to PA in the 12 months before their assessment, the number of likes received for each PA-related post, and verified fitness- or PA-related followings by the participants were extracted and analyzed. Pearson correlation analyses were used to evaluate the relationship among exercise identity, PA, and Instagram uses. We conducted mediation analyses using the PROCESS macro modeling tool to examine whether exercise identity mediated the relationship between Instagram use variables and PA. Descriptive statistical analyses were used to compare the number of willing participants versus those who were not willing to share their Instagram data.

Results: Of the 76 participants recruited to participate, 54% (n=41) shared their Instagram data. The percentage of PA-related Instagram posts (r=0.38; P=.01) and fitness-related Instagram followings (r=0.39; P=.01) were significantly associated with exercise identity. The average number of "likes" received (r=0.05, P=.75) was not significantly associated with exercise identity. Exercise identity significantly influenced the relationship between Instagram usage metrics (ie, the percentage of PA-related Instagram posts [P=.01] and verified fitness-related Instagram accounts [P=.01]) and PA level. Exercise identity did not significantly influence the relationship between the average number of "likes" received for the PA-related Instagram posts and PA level.

Conclusions: Our results suggest that an increase in PA-related Instagram posts and fitness-related followings were associated with a greater sense of exercise identity. Higher exercise identity led to higher PA levels. Exercise identity significantly influenced the relationship between PA-related Instagram posts (P=.01) and fitness-related followings on PA levels (P=.01). These results suggest that Instagram may influence a person's exercise identity and PA levels. Future intervention studies are warranted.

(J Med Internet Res 2021;23(4):e20954) doi:10.2196/20954

KEYWORDS

social media; exercise identity; physical activity; physical fitness

Introduction

Regular physical activity (PA) is critical to prevent chronic diseases and maintain overall well-being [1]. However, physical inactivity (ie, not meeting recommended guidelines) continues to be a major public health concern worldwide [2]. In Canada, approximately 79% of adults aged 18-39 did not meet the PA guidelines (150 minutes of moderate-to-vigorous PA per week) in 2015 [3]. Thus, there is an urgent need to find innovative solutions to better target and tailor PA interventions based on a person's needs to maximize public health impact [4]. Much of the PA promotion research literature has been based in the social cognitive tradition, which is focused on increasing expectations about the benefits of PA, improving perceptions of capability, and developing goals and self-regulation tactics to enable behavior change [5-8]. The social cognitive–based behavior change theories suggest that intentions are proximal antecedents of behavior; thus, supporting positive PA intentions can lead to subsequent behavioral enactment [9]. This approach has shown some effectiveness in terms of behavior change, albeit with modest effects [5]. There is clearly a need to continue to investigate additional mechanisms that can assist in PA continuance.

One of these mechanisms may be identity. Identity represents a person's self-categorization into a particular role [10]. This process assists in developing personal rules and standards of conduct with a particular behavior, such as PA [11]. As such, the central mechanism that sustains behavior becomes regulation around the identity standard [12]. Exercise identity refers to the degree to which an individual identifies himself/herself as an "exerciser" or someone who values and regularly engages in PA [13]. Discrepancies between the person's exercise identity and the present situation (eg, being physically inactive) can create negative emotions, which are proposed to motivate the individual to close the gap between the present behavior and the identity standard [12,14]. In essence, identity is a reflexive self-regulating system of behavioral continuance/stability [10]. Exercise identity has emerged as an important construct for PA promotion because of the reciprocal relationship between role identities and behavior [15]. Self-report surveys, such as the Exercise Identity Scale, have been developed and rigorously tested for assessing an individual's exercise identity, which has allowed for further investigation of the association between exercise identity and PA behavior [16].

Observational evidence has shown that PA behavior and identity have a reliable association in the medium to large effect size range [17]. Furthermore, the relationship is often independent of more traditional social–cognitive determinants (eg, intention) [17] and the mechanisms of identity dissonance and negative affect have been well-validated in hypothetical PA abstinence situations [18-20]. What has seen less attention, however, is the potential mechanisms or antecedents of identity development. Theories of exercise identity have proposed that investment, commitment and self-referential statements (eg, I am an exerciser), and social activation (comparison, support) may be crucial to identity development [13,21-23]. Another possible mechanism of identity is self-expression. Bem's self-perception theory [24], for example, suggests that identity may be formed largely by observing one's behavior and inferring self-categorization. This premise has some support in PA interventions from 2 small feasibility trials where participants who were encouraged to present PA-related imagery and self-expression (eg, a picture on the mantle) showed corresponding increases in identity [25,26]. However, the large focus on social media and self-expression across much of the population may be an even more effective mechanism of shaping identity.

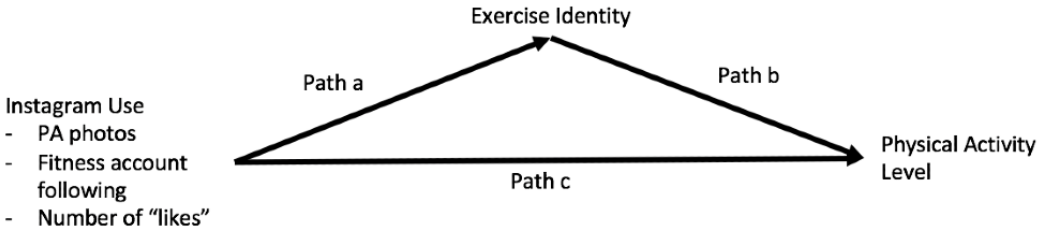
Social media websites, such as Instagram, allow users to easily and freely communicate and express themselves by posting photos, status messages, or links that become instantly available to the public [27]. Infodemiology is a new area of study where researchers examine ways to use data generated on social media to better understand and inform public health problems in real time [28]. Infoveillance has referred to applications where infodemiology methods are utilized, particularly for surveillance purposes [29]. PA promotion interventions can potentially extract useful information from user-generated social data for infoveillance and to tailor health promotion efforts to improve PA levels. In fact, recent studies have shown the possibility of using social media and digital interventions to promote PA [30-32].

Researchers have already shown that behavior characteristics extracted from social media sites, such as Twitter, can be used to model alongside other biomedical data sets to inform PA on a population level [30-34]. We believe that it is critical to explore the use of social data from other platforms, such as Instagram, to expand the field of infodemiology. Instagram is a highly unique and user-friendly platform that allows users to share photos and videos on their mobile devices anywhere at any time. This platform has evolved into a visual-oriented culture, promoting photos first and text second. Instagram enables users to express their personalities and lifestyles with others and form relationships with users that share similar interests and values. In contrast to Twitter and Facebook, Instagram's "photo-first" culture may contribute to different user behavior and motivation [35]. Approximately 71% of online adults in America have an Instagram account [36] compared with 38% of adults (18-29 years of age) that use Twitter [37]. In Canada, 37% of adults have a social networking account, and among adults aged 18-34, about 65% have an Instagram account [38]. An Instagram study suggested that analysis of the content of Instagram pictures, color analysis (amount of saturation brightness, photo filters), frequency of posts, patterns of followings, algorithmic face detection, and the number of likes are associated with symptoms of depression [39].

However, it remains unclear whether PA-related markers extracted from Instagram are associated with a person's exercise identity and subsequent PA level. Furthermore, Twitter data are designed to be publicly available, whereas Instagram data are

not publicly available by default [40]. Users may perceive their Instagram data as being more private and sensitive [41]. The percentage of participants who are willing to share their Instagram data for health behavior monitoring is currently unclear. Thus, the objectives of this study were to (1) explore whether participants were willing to share their Instagram data with researchers to predict their lifestyle behaviors; (2) examine whether PA-related Instagram uses (ie, the percentage of PA-related Instagram posts, fitness-related followings, and the number of likes received on PA-related posts) were positively associated with exercise identity; and (3) evaluate whether exercise identity mediates the relationship between PA-related Instagram use and weekly PA minutes. Commensurate with self-perception theory [24] and past research [25,26], we hypothesized (1) more than 50% of the participants would be willing to share their Instagram data for health infoveillance, (2) PA-related Instagram use was positively associated with exercise identity, and (3) exercise identity would mediate the relationship between PA-related Instagram use (ie, number of likes) and PA minutes. Figure 1 presents a visual representation of the hypothesized mediation model.

Figure 1. Hypothesized mediation model including Instagram use, exercise identity, and physical activity level.



Methods

Participants and Procedure

This cross-sectional study recruited university students (n=76) from the University of Victoria, British Columbia, Canada, using research posters and online advertisements. The rolling recruitment strategy took place between June 30, 2018, and July 1, 2019. Inclusion criteria required that participants be Instagram users, and between the ages of 18 and 29. Exclusion criteria included the inability to comprehend English. Interested participants were asked to contact the research team. The research team obtained consent from all eligible participants prior to data collection. During the consent process, participants were notified of the study objectives and that the sharing of their Instagram data was voluntary. All data obtained in connection with the study remained confidential. Confidentiality was maintained by means of encrypting on a secure server. All survey data and ID numbers were used on all databases. Researchers had access to participant Instagram data for up to 1 year from the start of the study.

Eligible participants who gave consent to participate in the study were asked to complete a questionnaire to record their demographic characteristics, daily Instagram use duration, and evaluate their current exercise identity and PA levels. Participants were also asked if they are willing to share their Instagram data. Participants that refused to share their Instagram data were given the option to write down reasons for not sharing. Those participants who agreed to share their Instagram data were asked to follow an Instagram research account in order for researchers to extract their Instagram data (photos, likes, followings). Researchers analyzed the previous 12 months of Instagram data from the completion date of the exercise identity and PA questionnaires. This study received research ethics approval from the research ethics board at the University of Victoria (protocol number: 17-294).

Measures

Demographic information (age, sex, and ethnicity) and daily Instagram usage time were collected using self-report instruments. PA levels were assessed using the Recent Physical Activity Questionnaire (RPAQ) [42]. Minutes of PA at low (metabolic equivalent of task [MET] < 6/week), moderate (MET 3-6/week), and high (MET > 6/week) intensities were calculated based on the RPAQ scoring instruction [42].

Exercise identity was assessed using the Exercise Identity Scale [16], which is a 9-item questionnaire on a 5-point Likert scale from 1 (strongly disagree) to 5 (strongly agree) (α=.92). Participants' exercise identities were determined by calculating a total score from the 9-item questionnaire; thus, scores ranged from 9 to 45 whereby a higher score represented a stronger exercise identity.

Participants' Instagram data were extracted using a customized program [43], which enabled the researchers to export users' posts, likes received for each post, and followings into separate Microsoft Excel files. Instagram posts extracted during the past 12 months before the completion of the study questionnaire were manually coded by 2 trained research assistants using standardized criteria to determine whether posts were PA related or not (eg, binary coded). PA-related posts were defined as those posts suggesting participants performed physical activities defined by the guidelines for exercise testing published by the American College of Sports Medicine [44]. Specifically, Instagram picture and video posts were identified as PA related if they included the person in sporting attire or the person in a location to perform PA (eg, kayaking, skiing, swimsuit, hiking) individually or in a group setting. Any discrepancies between the coders were resolved through discussion. The percentage of the number of PA-related posts relative to the total number of Instagram posts was calculated for each participant. The average number of likes for PA-related posts for each participant was also calculated.

An analysis of participants' Instagram "followings" was completed to identify the number of verified fitness or exercise-related Instagram accounts each participant followed. The verified fitness- or exercise-related Instagram accounts means Instagram has confirmed that an account is the authentic presence of the public figure, celebrity, or global brand it represents. A list of the most popular 300 verified fitness- or exercise-related Instagram accounts was compiled from publicly available Instagram rankings [45]. The percentage of exercise-related Instagram accounts following relative to the total number of Instagram followings was computed for each participant.

Statistical Analysis

All analyses were performed using the standard SPSS version 26.0 for Mac (SPSS, Inc./IBM), with a significance level set at $P<.05$. Descriptive statistical analyses were used to compare the number of willing participants with those who were not willing to share their Instagram data. Participants that were not willing to share their Instagram data were asked to provide reasons for not sharing in an open-ended question. Responses from these questions were grouped to create common themes. Independent *t* tests were used to compare differences between those groups (Instagram data shared versus Instagram data not shared) on demographics, PA, and exercise identity. A chi-square test was used to compare differences between categorical variables.

We examined normality (eg, skewness > 2 and kurtosis > 3) of all variables to determine whether any transformations were required by conversion to z-scores [46]. Pearson correlation analyses were used to evaluate the relationship among exercise identity, PA, and Instagram use metrics (percentage of PA-related Instagram post, percentage of fitness-related Instagram followings, and average number of "likes" on PA-related posts). Mediation analyses were based on 5000 bootstrapped samples using Hayes PROCESS Macro version 3.5 [47]. Multiple mediation analyses was used to examine whether exercise identity mediated the relationship between Instagram use variables and PA level [47]. This process involved examining path a, the association between Instagram use (independent variable) and exercise identity; path b, the impact of exercise identity (mediator variable) on PA level; and path c, the effect of Instagram use on PA levels (outcome variable). All analyses were controlled for age and sex. The 95% CIs must not cross 0 to satisfy the criteria for mediation.

Results

Participant Characteristics

A total of 76 participants were recruited and completed the study. Participant characteristics are presented in Table 1. The mean age was 19.7 (SD 0.31) years and the sample consisted of 72% female (55/76). Overall, 46% (35/76) of participants did not agree to share their Instagram data. Of these participants, 14 participants provided reasons for not sharing. Privacy was a major concern; specifically, there were concerns with (1) data oversight to prevent misusage of data and technology (n=9), (2) limitations to remain anonymous using Instagram photos (n=4), and (3) data ownership (n=1). Participants that were not willing to share their Instagram data reported significantly fewer weekly PA minutes at moderate to high intensity (MET ≥3/week) relative to participants that agreed to share their Instagram data ($P=.04$); however, there were no significant differences between groups for demographic characteristics, exercise identity, weekly PA minutes at low intensity (MET <3/week), and daily Instagram usage (Table 1).

A total of 41 participants agreed to share their Instagram data. The mean number of posts during the 12-month study period was 29.6 (SD 4.7). Examination of shared Instagram data revealed that, on average 7.3 (SD 1.1) or 25% of Instagram posts shared during the previous 12 months were related to PA. The mean number of Instagram following per person was 439.3 (SD 258) accounts. The mean number of verified PA- or fitness-related Instagram accounts following per user was 3 (SD 3.14). The mean number of "likes" received on all posts was 102.2 (SD 72.4) and the mean number of "likes" received on PA-related Instagram posts was 99 (SD 73).

Table 1. Baseline characteristics.

Characteristics	Instagram data shared (n=41)	Instagram data not shared (n=35)	*P* value
Age, mean (SD)	19.66 (2.08)	20.03 (1.81)	.42
Sex (female), n (%)	27 (66)	28 (80)	.20
Ethnicity, n (%)			.70
White	29 (71)	24 (69)	
East Asian	6 (15)	6 (17)	
Black	0 (0)	1 (3)	
Other (mixed, Spanish, South Asian)	6 (15)	4 (11)	
Exercise identity, mean (SD)	34.00 (8.20)	34.79 (7.16)	.67
Weekly PA[a] minutes at low intensity (MET[b] <3/week), mean (SD)	38.51 (135.94)	22.92 (59.94)	.50
Weekly PA min at moderate to high intensity (MET ≥3/week), mean (SD)	476.37 (542.40)	263.82 (238.27)	.04
Instagram usage (hours/day), mean (SD)	1.87 (1.04)	2.32 (1.91)	.29

[a]PA: physical activity.

[b]MET: metabolic equivalent of task.

The Relationship Among Exercise Identity, Physical Activity, and Instagram Use Metrics

Normality analysis revealed that weekly PA minutes at moderate to high intensity (MET ≥3/week) and fitness-related Instagram followings percentage were kurtotic (ie, values ≥3). These variables were transformed by conversion to z-scores. The Pearson correlation analyses are shown in Table 2. The results indicated that the percentage of PA-related Instagram posts (r=0.38; P=.01) and fitness-related Instagram (r=0.39; P=.01) followings were significantly associated with exercise identity. The strengths of the associates were moderate. We did not find that the average number of "likes" received (r=0.05, P=.75) to be significantly associated with exercise identity.

Table 2. Correlation analyses among exercise identity, physical activity, and Instagram use metrics.

Metrics	Variables				
	1	2	3	4	5
Pearson *r*					
Exercise identity	—				
Weekly PA[a] minutes at moderate to high intensity (MET[b] ≥3/week)	0.49	—			
Percentage of physical activity–related Instagram posts	0.38	0.03	—		
Percentage of verified fitness-related Instagram following	0.39	0.23	0.19	—	
Average number of "likes" received for the physical activity–related Instagram posts	0.05	0.01	–0.04	–0.13	—
***P* value**					
Exercise identity	—				
Weekly PA[a] minutes at moderate to high intensity (MET[b] ≥3/week)	.001	—			
Percentage of physical activity–related Instagram posts	.01	.86	—		
Percentage of verified fitness-related Instagram following	.01	.14	.23	—	
Average number of "likes" received for the physical activity–related Instagram posts	.75	.95	.78	.43	—

[a]PA: physical activity.

[b]MET: metabolic equivalent of task.

Mediation Analysis: Exercise Identity, Percentage of Physical Activity–Related Instagram Post, and Physical Activity

Exercise identity significantly influenced the relationship between the percentage of PA-related Instagram posts and PA levels (indirect effect: 3.28, 95% CI 0.54 to 8.11; direct effect: –3.65, 95% CI –9.70 to 2.39). As indicated in Table 3, a unit increase in percentage of PA-related Instagram posts was associated with a 0.10-unit increase in exercise identity; a unit increase in exercise identity was associated with 34 minutes of

PA (MET ≥3/week) increase. Approximately 29% of the variance was accounted for by the predictors (R^2=0.29).

Table 3. Mediation analyses of the effects of exercise identity on Instagram use and physical activity levels.[a]

Effects	B	95% CI	β	*P* value
Instagram PA[b]photos—PA levels (MET[c] ≥3/week)[d]				
Path a: Instagram PA photos—exercise identity	0.10	0.01 to 0.18	.33	.02
Path b: Exercise identity—PA levels	34.0	11.66 to 56.36	.51	<.01
Path c: Instagram PA photos—PA levels	–3.65	–9.70 to 2.39	–.18	.22
Instagram PA following—PA levels (MET ≥3/week)[e]				
Path a: Instagram PA following—exercise identity	3.18	1.02 to 5.33	.48	<.001
Path b: Exercise identity—PA levels	29.34	7.16 to 51.52	.44	.01
Path c: Instagram PA following—PA levels	32.75	–129.17 to 194.68	.06	.68
Instagram PA "likes"—PA levels (MET ≥3/week)[f]				
Path a: Instagram PA "likes"—exercise identity	0.02	–0.01 to 0.06	.21	.17
Path b: Exercise identity—PA levels	28.6	6.75 to 50.47	.43	.01
Path c: Instagram PA "likes"—PA levels	0.28	–2.06 to 2.61	.04	.81

[a]Model covariates include age, gender, B (unstandardized beta), and β (standardized beta).

[b]PA: physical activity.

[c]MET: metabolic equivalent of task.

[d]Indirect effect: 3.28 (95% CI 0.54 to 8.11); direct effect: –3.65 (95% CI –9.70 to 2.39).

[e]Indirect effect: 91.9 (95% CI 36.93 to 174.82); direct effect: 60.53 (95% CI –115.15 to 236.21).

[f]Indirect effect: 0.69 (95% CI –0.10 to 2.28); direct effect: 0.28 (95% CI –2.06 to 2.61).

Mediation Analysis: Exercise Identity, Fitness-Related Instagram Followings, and Physical Activity

Exercise identity significantly influenced the relationship between the percentage of verified fitness-related Instagram following and PA level (MET ≥3/week; indirect effect: 91.9, 95% CI 36.93 to 174.82; direct effect: 60.53, 95% CI –115.15 to 236.21). A unit increase in percentage of fitness-related Instagram following leads to a 3.18-unit increase in exercise identity; a unit increase in exercise identity was associated with an increase of 29.34 minutes of PA (MET ≥3/week; Table 3). Approximately 26% of the variance was accounted for by the predictors (R^2=0.26).

Mediation Analysis: Exercise Identity, Physical Activity–Related Instagram "Likes," and Physical Activity

Exercise identity did not significantly influence the relationship between the average number of "likes" received for the PA-related Instagram posts and PA level (MET ≥3/week; indirect effect: 0.69, 95% CI –0.10 to 2.28). We did not observe a significant direct effect between the number of "likes" received on PA-related Instagram posts and PA minutes (direct effect: 0.28, 95% CI –2.06 to 2.61; Table 3).

Discussion

Principal Findings

The objectives of this study were to explore whether participants were willing to share their Instagram data with researchers to predict their lifestyle behaviors and examine the relationship among PA-related Instagram metrics, exercise identity, and PA. We found that the majority of participants were willing to share their Instagram data for monitoring lifestyle behaviors. We also found that only PA-related Instagram posts and fitness-related followings were significantly associated with an increase in exercise identity. Furthermore, we found that exercise identity significantly influenced the relationship between certain Instagram use metrics (eg, PA-related Instagram posts and fitness-related followings) and PA level. To our knowledge, this is one of the first studies to examine the relationship among Instagram use, exercise identity, and PA levels. The methods and the findings from this study can be used to inform future infodemiology PA-related studies.

Our results suggest that over 50% (41/76, 54%) of the users were willing to share their Instagram data for monitoring their PA and health-related behaviors. This observation was supported by our hypothesis. It was worth noting that participants that agreed to share their data reported significantly higher weekly PA minutes at moderate to high intensity (MET ≥3/week) compared with participants who were not willing to share their data. A potential explanation is that those who were less active felt guilty for not showing consistent behavior with their social media posts. Individuals may not always portray their real selves on social media sites due to factors such as lower self-esteem, self-reflection, and self-concept clarity [48]. Alternatively, those participants may feel being monitored or judged for what they were posting, given the purpose of this study [27,49]. Concerns over privacy were the main reasons for not sharing their Instagram data. Unlike Twitter data, Instagram data are not

publicly available by default, and the users mainly share pictures instead of text. Thus, this poses a limitation to remain anonymous when analyzing Instagram photos. Our findings reinforced the ethical challenges faced in using social data for public health monitoring. There is a need to establish "best-practice" standards for analyzing social media data while simultaneously acknowledging peoples' individual rights and respecting their privacy [50,51]. Previous studies have shown that users are more likely to accept the use of social media data for research if there is an oversight body that protects user data to ensure that their data are not being used beyond the intended purpose and that user data can remain anonymous to protect the identity of the user [52,53]. Overall, these privacy considerations need to be addressed for the future development of social data monitoring technology.

Based on the self-perception theory [24] and past research [25,26], we hypothesized that PA-related Instagram use would be positively associated with exercise identity and that exercise identity would mediate the relationship between PA-related Instagram use and PA minutes. Interestingly, we observed that only the percentage of PA-related Instagram posts and fitness-related followings were positively associated with exercise identity. Exercise identity also significantly influenced the relationship between these Instagram usage metrics and PA level (MET ≥3/week). Exercise identity did not influence the relationship between the number of "likes" and PA level. Previous research has demonstrated that social media "likes" can contribute to PA engagement [54,55]. The Instagram "likes" metric is unlike the number of followings and posts metrics because users do not have direct control over the number of "likes" received. The lack of consistency of our findings compared with previous literature may be attributed to the way individuals are using their Instagram in our sample. "Surveillance users" (eg, users focused on monitoring other users) may post lesser often and have a lesser number of followers, leading to a lower number of "likes" compared with "cool-ness users" who are motivated to become popular [56,57]. Thus, the types of users need to be taken into consideration for future studies.

Findings from this study have several implications for PA promotion research. First, Instagram may have the potential to target identity antecedents through the reception of both inspirational and informative posts about exercise from respected verified fitness accounts (ie, enhancing self-efficacy and congruency with the self) [58-60]. Based on the self-perception theory [24] and self-definition model [21], future PA promotion interventions using Instagram may target personal investment, perceived commitment (ie, engaging with relevant fitness groups/users), social activation (ie, contributing and receiving support from fitness groups/users), and self-expression (ie, share photos of their PA or fitness-related experiences or write about these activities using social media), as these may play a role in strengthening an individual's exercise identity and subsequent PA behaviors [22,24]. However, it is worth noting that Instagram as a medium for expressing one's PA behavior with the end goal being to strengthen individuals' exercise identities may only be appealing to certain segments of the population. Second, results from this study extend previous Infodemiology research that Instagram data may be used to provide insights into health-related outcomes [39]. Future studies can build upon the methods used in this study and may also explore other Instagram metrics (eg, frequency of post) in PA infodemiology studies. Finally, PA researchers can leverage these social media analysis techniques to build prediction tools to monitor PA on a population level in real time. These tools may in the future aid public health agencies in identifying particular PA-related trends in various geographical areas on which to focus health and wellness initiatives.

There are potential negative aspects associated with using Instagram for the purpose of promoting PA and forming or strengthening one's exercise identity that must be considered. For example, recent movements known as "fitspiration" and "fitspro" have become popular on Instagram to inspire and motivate others to eat a healthy diet and exercise regularly. Fitspiration provides many positive attributes, such as accountability for exercising and education about fitness [61]. However, some women sharing fitspiration posts on Instagram may indicate disordered eating and exercise behaviors [62]. Despite the potential of using Instagram to motivate adherence to a healthy lifestyle, it is important for future interventions to consider the potential negative effect of social media on mental and physical health. Future studies need to examine ways to create a positive environment where participants feel comfortable posting pictures of their exercise experience and journey to fitness.

In terms of the limitations of the study, our sample only included university students which may limit the ability to generalize our findings beyond our sample. Future studies should examine the relationship between social media posts and PA-related behavior in middle-aged and older adults. Another limitation is that we used a list of verified fitness-related Instagram account in our analysis. It is possible that users may follow fitness-related Instagram accounts that are not verified and did not make it onto our list. Thus, this means that the observed relationship between exercise identity and the percentage of the fitness-related Instagram following needs to be interpreted with caution. Self-report questionnaires may result in confirmation bias. Moreover, it may be possible that some users' may have deleted photos, and thus affecting the number of PA-related photos in our analysis. Finally, the lack of longitudinal data for exercise identity and PA levels was another limitation when conducting mediation analysis. Thus, the mediation analysis result needs to be interpreted with caution. Future study with longitudinal data is needed to build more complex multivariate models to better examine mediation effects.

Conclusion

This study examined the relationships among Instagram use, exercise identity, and PA levels. Findings from this study demonstrated that over 50% (41/76, 54%) of participants were willing to share their Instagram data with researchers for the purpose of monitoring and predicting PA behaviors. Exercise identity significantly influenced the relationship between Instagram use (eg, PA-related Instagram posts and fitness-related followings) and PA level. Results from this study suggest that there is an association between Instagram posts and PA-related

outcomes (ie, exercise identity and PA levels). Future studies should examine whether Instagram may be used for a future intervention to help form or strengthen an individual's exercise identity, particularly if the individual can be supported by their "social world."

Acknowledgments

We acknowledge Henry La and Si Ning Yeo for their assistance in coding the physical activity–related posts. This study is supported by the Internal Research Project Grant at the University of Victoria.

Conflicts of Interest

None declared.

References

1. Fletcher GF, Landolfo C, Niebauer J, Ozemek C, Arena R, Lavie CJ. Promoting Physical Activity and Exercise: JACC Health Promotion Series. J Am Coll Cardiol 2018 Oct 02;72(14):1622-1639 [FREE Full text] [doi: 10.1016/j.jacc.2018.08.2141] [Medline: 30261965]
2. Ozemek C, Lavie CJ, Rognmo Ø. Global physical activity levels - Need for intervention. Prog Cardiovasc Dis 2019 Mar;62(2):102-107. [doi: 10.1016/j.pcad.2019.02.004] [Medline: 30802461]
3. Statistics Canada. Distribution of the household population meeting/not meeting the Canadian physical activity guidelines, by sex and age group, occasional (percentage). 2015. URL: https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1310033701 [accessed 2021-04-12]
4. Liu S, Brooks D, Thomas SG, Eysenbach G, Nolan RP. Effectiveness of User- and Expert-Driven Web-based Hypertension Programs: an RCT. Am J Prev Med 2018 Apr;54(4):576-583. [doi: 10.1016/j.amepre.2018.01.009] [Medline: 29456025]
5. Rhodes R, McEwan D, Rebar A. Theories of physical activity behaviour change: A history and synthesis of approaches. Psychology of Sport and Exercise 2019 May;42:100-109. [doi: 10.1016/j.psychsport.2018.11.010]
6. Hagger M, Weed M. DEBATE: Do interventions based on behavioral theory work in the real world? Int J Behav Nutr Phys Act 2019 Apr 25;16(1):36 [FREE Full text] [doi: 10.1186/s12966-019-0795-4] [Medline: 31023328]
7. McEwan D, Beauchamp M, Kouvousis C, Ray C, Wyrough A, Rhodes R. Examining the active ingredients of physical activity interventions underpinned by theory versus no stated theory: a meta-analysis. Health Psychol Rev 2019 Mar;13(1):1-17. [doi: 10.1080/17437199.2018.1547120] [Medline: 30412685]
8. Brand R, Cheval B. Theories to Explain Exercise Motivation and Physical Inactivity: Ways of Expanding Our Current Theoretical Perspective. Front Psychol 2019;10:1147 [FREE Full text] [doi: 10.3389/fpsyg.2019.01147] [Medline: 31164856]
9. Rhodes RE, de BG. What predicts intention-behavior discordance? A review of the action control framework. Exerc Sport Sci Rev 2013 Oct;41(4):201-207. [doi: 10.1097/JES.0b013e3182a4e6ed] [Medline: 23873134]
10. Burke P, Stets J. Identity Theory. New York, NY: Oxford University Press; 2009.
11. Stryker S, Burke PJ. The Past, Present, and Future of an Identity Theory. Social Psychology Quarterly 2000 Dec;63(4):284. [doi: 10.2307/2695840]
12. Burke PJ. Identity Change. Soc Psychol Q 2016 Jun 23;69(1):81-96. [doi: 10.1177/019027250606900106]
13. Strachan SM, Fortier MS, Perras MG, Lugg C. Understanding variations in exercise-identity strength through identity theory and self-determination theory. International Journal of Sport and Exercise Psychology 2013 Sep;11(3):273-285. [doi: 10.1080/1612197x.2013.749005]
14. Stryker S, Burke PJ. The Past, Present, and Future of an Identity Theory. Social Psychology Quarterly 2000 Dec;63(4):284. [doi: 10.2307/2695840]
15. Biddle B, Bank B, Slavings R. Norms, Preferences, Identities and Retention Decisions. Social Psychology Quarterly 1987 Dec;50(4):322 [FREE Full text] [doi: 10.2307/2786817]
16. Anderson DF, Cychosz CM. Development of An Exercise Identity Scale. Percept Mot Skills 2017 May 30;78(3):747-751. [doi: 10.1177/003151259407800313]
17. Rhodes RE, Kaushal N, Quinlan A. Is physical activity a part of who I am? A review and meta-analysis of identity, schema and physical activity. Health Psychol Rev 2016 Jun 02;10(2):204-225. [doi: 10.1080/17437199.2016.1143334] [Medline: 26805431]
18. Strachan SM, Brawley LR. Reactions to a perceived challenge to identity: a focus on exercise and healthy eating. J Health Psychol 2008 Jul 01;13(5):575-588. [doi: 10.1177/1359105308090930] [Medline: 18519432]
19. Strachan SM, Brawley LR, Spink KS, Jung ME. Strength of exercise identity and identity-exercise consistency: affective and social cognitive relationships. J Health Psychol 2009 Nov 26;14(8):1196-1206. [doi: 10.1177/1359105309346340] [Medline: 19858339]
20. Strachan SM, Flora PK, Brawley LR, Spink KS. Varying the cause of a challenge to exercise identity behaviour: reactions of individuals of differing identity strength. J Health Psychol 2011 May 23;16(4):572-583. [doi: 10.1177/1359105310383602] [Medline: 21346019]

21. Kendzierski D, Morganstein M. Test, revision, and cross-validation of the Physical Activity Self-Definition Model. J Sport Exerc Psychol 2009 Aug;31(4):484-504. [doi: 10.1123/jsep.31.4.484] [Medline: 19842544]
22. Kendzierski D, Furr R, Schiavoni J. Physical activity self-definitions: Correlates and perceived criteria. J Sport Exerc Psychol 1998;20:193. [doi: 10.1123/jsep.20.2.176]
23. Rhodes R. Chapter Five - The Evolving Understanding of Physical Activity Behavior: A Multi-Process Action Control Approach. In: Inlliot AJ, editor. Advances in Motivation Science. Cambridge, MA: MAlsevier Academic Press; 2017:171-205.
24. Bem DJ. In: Berkowitz L, editor. Self-perception theory. New York, NY: Academic Press; 1972.
25. Husband CJ, Wharf-Higgins J, Rhodes RE. A feasibility randomized trial of an identity-based physical activity intervention among university students. Health Psychology and Behavioral Medicine 2019 Apr 10;7(1):128-146. [doi: 10.1080/21642850.2019.1600407]
26. Lim C. Working Out with F.I.D.O. (Frequency, Intensity, Duration, & Outcomes) - A Feasibility Randomized Controlled Trial (MSc Thesis). 2017. URL: https://www.researchgate.net/publication/319177116 [accessed 2021-04-14]
27. Alhabash S, Ma M. A Tale of Four Platforms: Motivations and Uses of Facebook, Twitter, Instagram, and Snapchat Among College Students? Social Media + Society 2017 Feb 01;3(1):205630511769154. [doi: 10.1177/2056305117691544]
28. Eysenbach G. Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the Internet. J Med Internet Res 2009 Mar 27;11(1):e11 [FREE Full text] [doi: 10.2196/jmir.1157] [Medline: 19329408]
29. Eysenbach G. Infodemiology and infoveillance tracking online health information and cyberbehavior for public health. Am J Prev Med 2011 May;40(5 Suppl 2):S154-S158. [doi: 10.1016/j.amepre.2011.02.006] [Medline: 21521589]
30. Liu S, Zhu M, Yu DJ, Rasin A, Young SD. Using Real-Time Social Media Technologies to Monitor Levels of Perceived Stress and Emotional State in College Students: A Web-Based Questionnaire Study. JMIR Ment Health 2017 Jan 10;4(1):e2 [FREE Full text] [doi: 10.2196/mental.5626] [Medline: 28073737]
31. Liu S, Young SD. A survey of social media data analysis for physical activity surveillance. J Forensic Leg Med 2018 Jul;57:33-36 [FREE Full text] [doi: 10.1016/j.jflm.2016.10.019] [Medline: 29801949]
32. Liu S, Chen B, Kuo A. Monitoring Physical Activity Levels Using Twitter Data: Infodemiology Study. J Med Internet Res 2019 Jun 03;21(6):e12394 [FREE Full text] [doi: 10.2196/12394] [Medline: 31162126]
33. Liu S, Zhu M, Young SD. Monitoring Freshman College Experience Through Content Analysis of Tweets: Observational Study. JMIR Public Health Surveill 2018 Jan 11;4(1):e5 [FREE Full text] [doi: 10.2196/publichealth.7444] [Medline: 29326096]
34. Paul M, Dredze M. You are what you tweet: Analyzing twitter for public health. 2011 Presented at: Fifth International AAAI Conference on Weblogs and Social Media; July 17–21, 2011; Baltimore, MD.
35. Lee E, Lee J, Moon JH, Sung Y. Pictures Speak Louder than Words: Motivations for Using Instagram. Cyberpsychol Behav Soc Netw 2015 Sep;18(9):552-556. [doi: 10.1089/cyber.2015.0157] [Medline: 26348817]
36. Greenwood S, Perrin A, Duggan M. Social Media Update 2016. Washington, D.C: Pew Research Center; 2016. URL: https://www.pewresearch.org/internet/wp-content/uploads/sites/9/2016/11/PI_2016.11.11_Social-Media-Update_FINAL.pdf [accessed 2020-08-01]
37. Perrin A, Anderson M. Share of U.S. adults using social media, including Facebook, is mostly unchanged since 2018. Washington, D.C: Pew Research Center; 2019. URL: https://www.pewresearch.org/fact-tank/2019/04/10/share-of-u-s-adults-using-social-media-including-facebook-is-mostly-unchanged-since-2018/ [accessed 2020-05-01]
38. Gruzd A, Jacobson J, Mai P, Dubois E. The State of Social Media in Canada 2017. SSRN Journal 2018:1-18. [doi: 10.2139/ssrn.3158771]
39. Reece AG, Danforth CM. Instagram photos reveal predictive markers of depression. EPJ Data Sci 2017 Aug 8;6(1):s13688. [doi: 10.1140/epjds/s13688-017-0110-z]
40. Tenkanen H, Di Minin E, Heikinheimo V, Hausmann A, Herbst M, Kajala L, et al. Instagram, Flickr, or Twitter: Assessing the usability of social media data for visitor monitoring in protected areas. Sci Rep 2017 Dec 14;7(1):17615 [FREE Full text] [doi: 10.1038/s41598-017-18007-4] [Medline: 29242619]
41. Talib S, Abdul RS, Olowolayemo A, Salependi M, Ahmad N, Kunhamoo S, et al. Perception analysis of social networks? privacy policy: Instagram as a case study. 2014 Presented at: 5th International Conference on Information and Communication Technology for the Muslim World (ICT4M); 17-18 Nov. 2014; Kuching, Malaysia. [doi: 10.1109/ict4m.2014.7020612]
42. InterAct Consortium, Peters T, Brage S, Westgate K, Franks PW, Gradmark A, et al. Validity of a short questionnaire to assess physical activity in 10 European countries. Eur J Epidemiol 2012 Jan;27(1):15-25 [FREE Full text] [doi: 10.1007/s10654-011-9625-y] [Medline: 22089423]
43. Mishra S. Instagram Data Download Internet. 2020. URL: https://codecanyon.net/item/instagram-scrapper/21146630 [accessed 2020-07-05]
44. American College of Sports Medicine. In: Lupash E, editor. ACSM's Guidelines for Exercise Testing Prescription. Ninth Edition. Philadelphia, PA: Wolters Kluwer Health, Lippincott Williams & Wilkins; 2014.
45. 30 Inspiring Fitness Girls To Follow On Instagram. 2020. URL: https://www.harpersbazaar.com/beauty/diet-fitness/g4018/inspiring-fitness-girls-on-instagram/ [accessed 2020-06-27]

46. Kim H. Statistical notes for clinical researchers: assessing normal distribution (2) using skewness and kurtosis. Restor Dent Endod 2013 Feb;38(1):52-54. [doi: 10.5395/rde.2013.38.1.52] [Medline: 23495371]
47. Hayes A. Introduction to mediation, moderation, and conditional process analysis: A regression-based approach. New York, NY: Guilford Publications; 2017:1.
48. Yang C, Brown BB. Online Self-Presentation on Facebook and Self Development During the College Transition. J Youth Adolesc 2016 Feb 3;45(2):402-416. [doi: 10.1007/s10964-015-0385-y] [Medline: 26534776]
49. Hausenblas HA, Brewer BW, Van Raalte JL. Self-Presentation and Exercise. Journal of Applied Sport Psychology 2004 Jan;16(1):3-18. [doi: 10.1080/10413200490260026]
50. Vayena E, Salathé M, Madoff LC, Brownstein JS. Ethical challenges of big data in public health. PLoS Comput Biol 2015 Feb;11(2):e1003904 [FREE Full text] [doi: 10.1371/journal.pcbi.1003904] [Medline: 25664461]
51. Conway M. Ethical issues in using Twitter for public health surveillance and research: developing a taxonomy of ethical concepts from the research literature. J Med Internet Res 2014 Dec 22;16(12):e290 [FREE Full text] [doi: 10.2196/jmir.3617] [Medline: 25533619]
52. Beninger K, Lepps H. Research using social media: users' views. NatCen Social Research. 2014. URL: https://www.researchgate.net/publication/261551701 [accessed 2021-04-14]
53. Mikal J, Hurst S, Conway M. Ethical issues in using Twitter for population-level depression monitoring: a qualitative study. BMC Med Ethics 2016 Apr 14;17:22 [FREE Full text] [doi: 10.1186/s12910-016-0105-5] [Medline: 27080238]
54. Tu R, Hsieh P, Feng W. Walking for fun or for "likes"? The impacts of different gamification orientations of fitness apps on consumers' physical activities. Sport Management Review 2019 Nov;22(5):682-693. [doi: 10.1016/j.smr.2018.10.005]
55. Hamari J, Koivisto J. "Working out for likes": An empirical study on social influence in exercise gamification. Computers in Human Behavior 2015 Sep;50:333-347. [doi: 10.1016/j.chb.2015.04.018]
56. Sheldon P, Bryant K. Instagram: Motives for its use and relationship to narcissism and contextual age. Computers in Human Behavior 2016 May;58:89-97. [doi: 10.1016/j.chb.2015.12.059]
57. Bakhshi S, Shamma D, Gilbert E. Faces engage us: Photos with faces attract more likes and comments on instagram. New York, NY: ACM; 2014 Presented at: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems; April 26 to May 1, 2014; Toronto, ON p. 965-974 URL: https://www.researchgate.net/publication/266655817 [doi: 10.1145/2556288.2557403]
58. Al-Eisa E, Al-Rushud A, Alghadir A, Anwer S, Al-Harbi B, Al-Sughaier N, et al. Effect of Motivation by "Instagram" on Adherence to Physical Activity among Female College Students. Biomed Res Int 2016;2016:1546013-1546016 [FREE Full text] [doi: 10.1155/2016/1546013] [Medline: 27034927]
59. Santarossa S, Coyne P, Lisinski C, Woodruff SJ. #fitspo on Instagram: A mixed-methods approach using Netlytic and photo analysis, uncovering the online discussion and author/image characteristics. J Health Psychol 2019 Mar 15;24(3):376-385. [doi: 10.1177/1359105316676334] [Medline: 27852889]
60. Vaterlaus JM, Patten EV, Roche C, Young JA. #Gettinghealthy: The perceived influence of social media on young adult health behaviors. Computers in Human Behavior 2015 Apr;45:151-157. [doi: 10.1016/j.chb.2014.12.013]
61. DiBisceglie S, Arigo D. Perceptions of #fitspiration activity on Instagram: Patterns of use, response, and preferences among fitstagrammers and followers. J Health Psychol 2019 Aug 28:1359105319871656 [FREE Full text] [doi: 10.1177/1359105319871656] [Medline: 31455118]
62. Holland G, Tiggemann M. "Strong beats skinny every time": Disordered eating and compulsive exercise in women who post fitspiration on Instagram. Int J Eat Disord 2017 Jan;50(1):76-79. [doi: 10.1002/eat.22559] [Medline: 27302867]

Abbreviations

MET: metabolic equivalent of task
PA: physical activity
RPAQ: Recent Physical Activity Questionnaire

Edited by R Kukafka; submitted 02.06.20; peer-reviewed by F Santos, M Bardus, A Roundtree, AM Auvinen; comments to author 06.07.20; revised version received 31.08.20; accepted 18.03.21; published 19.04.21.

Please cite as:
Liu S, Perdew M, Lithopoulos A, Rhodes RE
The Feasibility of Using Instagram Data to Predict Exercise Identity and Physical Activity Levels: Cross-sectional Observational Study
J Med Internet Res 2021;23(4):e20954
URL: https://www.jmir.org/2021/4/e20954
doi:10.2196/20954
PMID:33871380

Original Paper

Assessing Children's Fine Motor Skills With Sensor-Augmented Toys: Machine Learning Approach

Annette Brons[1,2], MSc; Antoine de Schipper[3], MSc; Svetlana Mironcika[4], MSc; Huub Toussaint[3], PhD; Ben Schouten[4,5], PhD; Sander Bakkes[2], PhD; Ben Kröse[1,6], PhD

[1]Digital Life Center, Amsterdam University of Applied Sciences, Amsterdam, Netherlands

[2]Department of Information and Computing Sciences, Utrecht University, Utrecht, Netherlands

[3]Academy for Physical Education, Amsterdam University of Applied Sciences, Amsterdam, Netherlands

[4]Play and Civic Media, Amsterdam University of Applied Sciences, Amsterdam, Netherlands

[5]Department of Industrial Design, Eindhoven University of Technology, Eindhoven, Netherlands

[6]Department of Computer Science, University of Amsterdam, Amsterdam, Netherlands

Corresponding Author:
Annette Brons, MSc
Digital Life Center
Amsterdam University of Applied Sciences
Wibautstraat 2-4
Amsterdam, 1091 GM
Netherlands
Phone: 31 621156976
Email: a.e.brons@hva.nl

Abstract

Background: Approximately 5%-10% of elementary school children show delayed development of fine motor skills. To address these problems, detection is required. Current assessment tools are time-consuming, require a trained supervisor, and are not motivating for children. Sensor-augmented toys and machine learning have been presented as possible solutions to address this problem.

Objective: This study examines whether sensor-augmented toys can be used to assess children's fine motor skills. The objectives were to (1) predict the outcome of the fine motor skill part of the Movement Assessment Battery for Children Second Edition (fine MABC-2) and (2) study the influence of the classification model, game, type of data, and level of difficulty of the game on the prediction.

Methods: Children in elementary school (n=95, age 7.8 [SD 0.7] years) performed the fine MABC-2 and played 2 games with a sensor-augmented toy called "Futuro Cube." The game "roadrunner" focused on speed while the game "maze" focused on precision. Each game had several levels of difficulty. While playing, both sensor and game data were collected. Four supervised machine learning classifiers were trained with these data to predict the fine MABC-2 outcome: k-nearest neighbor (KNN), logistic regression (LR), decision tree (DT), and support vector machine (SVM). First, we compared the performances of the games and classifiers. Subsequently, we compared the levels of difficulty and types of data for the classifier and game that performed best on accuracy and F1 score. For all statistical tests, we used α=.05.

Results: The highest achieved mean accuracy (0.76) was achieved with the DT classifier that was trained on both sensor and game data obtained from playing the easiest and the hardest level of the roadrunner game. Significant differences in performance were found in the accuracy scores between data obtained from the roadrunner and maze games (DT, *P*=.03; KNN, *P*=.01; LR, *P*=.02; SVM, *P*=.04). No significant differences in performance were found in the accuracy scores between the best performing classifier and the other 3 classifiers for both the roadrunner game (DT vs KNN, *P*=.42; DT vs LR, *P*=.35; DT vs SVM, *P*=.08) and the maze game (DT vs KNN, *P*=.15; DT vs LR, *P*=.62; DT vs SVM, *P*=.26). The accuracy of only the best performing level of difficulty (combination of the easiest and hardest level) achieved with the DT classifier trained with sensor and game data obtained from the roadrunner game was significantly better than the combination of the easiest and middle level (*P*=.046).

Conclusions: The results of our study show that sensor-augmented toys can efficiently predict the fine MABC-2 scores for children in elementary school. Selecting the game type (focusing on speed or precision) and data type (sensor or game data) is more important for determining the performance than selecting the machine learning classifier or level of difficulty.

(J Med Internet Res 2021;23(4):e24237) doi:10.2196/24237

KEYWORDS

motor development; fine motor function; gamification; playful; motor skill assessment; Movement ABC (MABC); machine learning; motor function; motor skills; toys; children; game; movement assessment

Introduction

Background

Motor development is crucial in child development. Acquiring motor skills is not only essential for daily life functioning but also influences children's cognitive and social development [1]. Fine motor skills are a strong predictor of school results [2]. Motor skill development is not a fixed linear process. Every child has his/her unique learning curve and pace, and their motor skills develop by leaps and bounds [3]. Because of this unique and unpredictable motor development path, it is important to monitor children's motor development over time instead of assessing them once. That way, insight in the progress of their motor development can be given [3,4]. Children with fine motor development problems have difficulties with learning fine motor skills. They experience, for instance, problems with school tasks such as writing or cutting or daily life activities such as closing a zipper or tying shoelaces [4]. In total, 5%-10% of children in elementary school have developmental motor problems [5,6]. When monitoring children's motor development over time, these motor development problems can be recognized at an early stage. Consequently, appropriate diagnostic methods and required therapy could be started in time, which may diminish the effects of their motor development problems.

Both worldwide and in the Netherlands, the Movement Assessment Battery for Children Second Edition (MABC-2) is the major test for assessing children's motor development [5,7]. The MABC-2 consists of both tests for fine and gross motor skills. Finishing the fine motor skill part of the MABC-2 (fine MABC-2) takes approximately 15 minutes per child and requires a trained supervisor [7]. Elementary school would be a natural place to test children's motor skills since developmental motor problems affect cognitive development and school results and vice versa. Proficient fine motor skills are, for instance, essential for children to learn handwriting [8]. However, teachers in elementary school report that they do not have the required expertise and time to test all children, let alone to monitor them all over time.

Sensor-augmented toys and machine learning have been presented as possible solutions for the problems that teachers experience with the current assessment methods [9,10]. Both sensor data, regarding movements made with the toy, and game data, regarding events that occur in the game, can be collected while playing. Such a sensor-augmented toy can, for instance, measure the smoothness of movements made with the toy or how accurately a game was played. These data can be used to train machine learning algorithms in predicting children's motor skill levels. After training and testing those machine learning algorithms, they can be used to classify whether children have fine motor development problems or not. Thus, sensor-augmented toys can be used as an assessment tool for signaling fine motor development problems in children.

Using sensor-augmented toys for indicating fine motor development problems in children has many advantages. First, these toys do not require a trained supervisor and require less instruction time. Moreover, such toys provide more secure data collection compared to manually collected data. Furthermore, playing games can safely be considered to be more enjoyable for children than standard assessment methods. Last, since no trained supervisor is required and children can easily play with the toys in the classroom, the toys enable testing in a natural setting instead of a testing environment. By playing games in such a natural setting, the assessment can be kept implicit and, therefore, children are not aware of undergoing the assessment [11].

Related Work

Despite the advantages of evaluating children's fine motor skills with sensor-augmented toys and machine learning, limited research is done in this field. Gamification of assessment processes in other contexts such as cognitive assessments has been studied before [12]. The systematic review of Lumsden et al [12] shows that many gamified cognitive assessments have been validated successfully. Although gamification in the field of health and well-being is popular, most studies focus on promoting physical activity levels, mental health, and people with a chronic disease [13,14]. Only a few studied gamification in the context of motor skills and most of them involved training instead of assessment [15,16]. Moreover, most of those studies involved patients with motor problems that were primarily caused by medical conditions such as cerebral palsy or stroke. Those children are already seen by medical specialists who monitor their motor development. In contrast, our study involved children who may have a delay in their motor skill development but do not have such diseases.

To the best of our knowledge, only 3 studies involved smart toys for assessing children's fine motor skills [17-19]. Vega-Barbas et al [17] only performed a usability and feasibility test with smart toys that are potentially helpful for assessing motor skills, but they have not used it for assessment yet. The remaining 2 studies did use toys to evaluate children's fine motor skill levels, but both involved toddlers instead of elementary school children. Moreover, they did not build a classification model that might predict the outcome of current motor skill assessment tests. Rivera et al [18] studied the intraindividual variability. Guitiérrez García et al [19] did build a regression model, but this model has not been tested and used to classify the fine motor skill level based on the sensor data yet. In addition to the sensor data that both studies included, we will also include game data, that is, data about events that occur in the game, to study its additional value.

In preliminary research, we studied the possibilities to use sensor-augmented toys for fine motor skill assessment [9,10]. We studied whether a toy called the Futuro Cube could be used

to predict the outcome of the fine MABC-2 [9]. While a game was played with the toy, information regarding events occurring in the game was registered. In addition to these game data, sensor data were collected by measuring the movements of the toy with accelerometers inside the toy. These sensor data and game data were used as input for several supervised machine learning models, which then were used to classify the motor skill level of children. Our previous study showed that a machine learning model that uses sensor and game data of the Futuro Cube as input has the potential to classify the fine motor skill levels of children aged 7-8 years.

Objective of This Study

In this study, in which a larger number of children participated, we improved our toy and game compared to that used in our preliminary research. First, we explored whether additional sensor features would improve the results. Second, we studied a game in which children could choose their own pace in the game since such elements are also included in the fine MABC-2. Therefore, we added a gyroscope to the toy to collect rotational data in addition to acceleration data. Moreover, we designed an additional game that focusses on precision instead of speed. Based on these modifications, we will answer the following research question: *What is the influence of the classification model, game, type of data, and level of difficulty (LoD) on predicting the fine MABC-2 scores for children aged 6-9 years with playing games with the Futuro Cube?*

Methods

Recruitment

Children were recruited through their elementary school teachers. Since we required a sufficient number of participants having motor development problems for balanced class labels, we included 2 elementary schools in Amsterdam that were known for having a larger population of children with motor development problems. Children who were between the age of 6 and 9 years and who were in the 3rd or 4th year of elementary school were included. Fine motor development is important for handwriting education. In the Netherlands, handwriting education starts in the 3rd year of elementary school and is a very important part of the 4th year of elementary school. Teachers of those classes reported that they need to know whether children's fine motor development is proficient to start such education. Therefore, we chose not to include children in classes higher than the 4th year of elementary school. Pilot tests of our game showed that the explanation of the game was too hard for some children younger than 6 years. To make sure that all children understood how to play games with the toy, we chose not to include children younger than 6 years. Written informed consent was obtained from parents or legal guardians for participation of the child. A separate informed consent was acquired for publication of the pseudonymized raw data. In total, written informed consent for participation was given for 99 children and written informed consent for publication of the raw data was given for 49 children. A pseudonymized data set consisting of the sensor and game data of these children and their corresponding fine MABC-2 scores is available on request from the corresponding author. This study was performed according to the Declaration of Helsinki [20].

Procedures

Test Setup

Each child was tested for 25 minutes. The fine MABC-2 was taken in 15 minutes by a trained supervisor. Further, the child played 2 different games on the Futuro Cube. Each game started with a short instruction, followed by a warming-up phase in which the child was able to become familiar with the game. Half of the participants started with the fine MABC-2 and subsequently played with the toy. For the other half of the participants, the order was the other way around: first playing with the toy and subsequently performing the fine MABC-2.

Determining the Level of Fine Motor Skills

The subscale for the measurement of fine motor skills of the MABC-2 was used to determine the fine motor skill level of the children. The fine MABC-2 test consists of 3 subtests. In the first subtest, children had to place 12 pegs in a board with 12 holes. In the second subtest, children had to thread a lace back and forth through a lacing board with holes. Both the first and second subtests were time-sensitive. The child was told to perform the task as quickly as possible and the time to complete the task was denoted as the raw score of those subtests. In the last subtest, children had to draw a trail with a pencil. They had to draw a single line and were not allowed to cross the trail's boundaries. This subtest was not time-sensitive and the raw score consisted of the number of errors, that is, the number of times that the drawn line crossed the boundaries [7]. The raw scores of each subtest were summed to a raw total score. Based on the age of the participant, the raw total score was converted to a percentile score. This score, between 0 and 100, indicates the fine motor skill level of the participant compared to that of the children within the same age band. The higher the score, the better the fine motor performance of the participant compared to children of the same age. According to the MABC manual, a score in the 16th percentile or lower was defined as likely to have fine motor development problems. All scores in the 17th percentile or above were defined as not having fine motor development problems.

Toy and Games

The Futuro Cube, which is shown in Figure 1, is a commercial toy that was adapted for research [21]. The cube has 9 colored lights on each side. The accelerometer and gyroscope inside the cube track motion, sense rotation, and measure orientation. The cube is 52×52×52 cm with 9 colored light-emitting diodes (LEDs) on each side. Each square can be identified by a unique index number $i \in \{0, \ldots, 53\}$. This index number can be used to register the activation of an LED, including the color. The toy contains a tri-axial accelerometer with an acceleration sensitivity of ±8G and a tri-axial gyroscope with an angular rate sensitivity of 2000 dps. Based on these inertial measurement unit sensors, the orientation and the change of position can be recorded. Data collected with these sensors will be referred to as sensor data. The programming language PAWN2 was used for creating the games that are played with the toy. In the programmed script, we defined which information about events

that occur during a game should be saved, for example, the change of color of an LED during the game. Saved data about the game will from now on be referred to as game data. While playing, both the game data and the abovementioned sensor data were registered. Bluetooth low energy was used to wirelessly send all data in real time from the cube to a computer with a sample frequency of 110 Hz. In both games, a highlighted dot was moving on the cube's surface by activating the colored LEDs. In the first game, called the roadrunner game, the focus was on speed. The second game, called the maze game, focused on precision. In both games, no points were collected and neither visual nor auditory feedback was given about how well the game was played. In the roadrunner game, the velocity of the moving dot was predetermined and the child had to follow this speed. In the second game, however, the child was asked to move the dot as precisely as possible through the path without being tied to a certain pace.

Figure 1. Futuro cube.

In the roadrunner game, a green dot moved on the surface of the cube. The player was asked to rotate the cube in order to keep the spot on the top surface of the cube, which is shown in Figure 2. The dot moved with a certain velocity on the cube surface jumping from LED to LED. In case the spot was at the center LED of a side, it randomly turned left or right or kept moving forward. The velocity at which the spot moved was defined as the LoD. This game had 3 LoDs: LoD $\in$ {0,1,2}. The lower the level, the longer the spot remained at the same place. Thus, level 0 was the easiest level and level 2 was the hardest level. The time that the spot remained at the same index number was denoted as the delay in seconds. The LoDs correspond to a delay $d \in \{0.8, 0.6, 0.4\}$. Each level lasted for 30 seconds and occurred twice. Hereby, it has to be taken into account that each player started with the easiest level and 2 subsequent levels could not have the same LoD. The order of the LoDs was randomized and the game started with a warming up phase of 60 seconds to discover the game. Table 1 shows all 10 possible permutations along with their order of LoDs.

Figure 2. Schematic overview of the roadrunner game. A-D: show the way the cube should be rotated to keep the green dot in the correct position. E-H: show what happens in case the cube was not rotated at all.

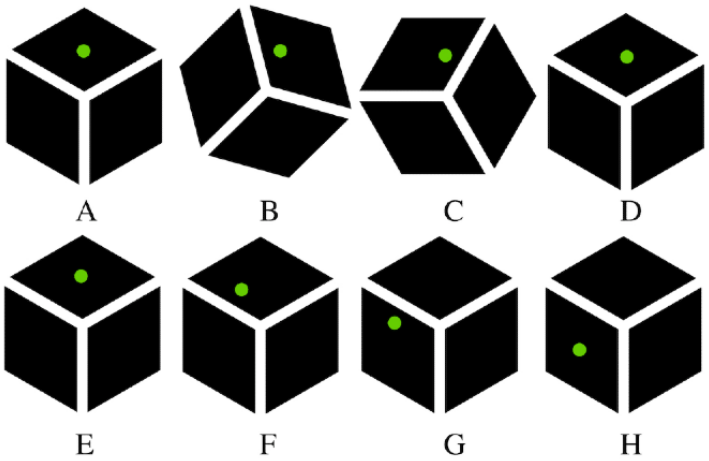


Table 1. All possible permutations for the roadrunner game.

Permutation	Order of levels of difficulty	Participants (n)
0	0, 1, 2, 0, 1, 2	10
1	0, 1, 2, 0, 2, 1	10
2	0, 1, 2, 1, 2, 0	10
3	0, 1, 2, 1, 0, 2	10
4	0, 2, 1, 0, 1, 2	10
5	0, 2, 1, 0, 2, 1	9
6	0, 2, 1, 2, 0, 1	10
7	0, 2, 1, 2, 1, 0	10
8	0, 1, 0, 2, 1, 2	7
9	0, 2, 0, 1, 2, 1	9

The goal of the maze game was to move a white dot through a maze of green dots, as is shown in Figure 3. The dot could be moved by rotating the cube. In case the player moved the white dot on a location that was green and thus not allowed to enter, the dot turned red. When the player moved the white dot back on the right track, the red dot returned green. The path that the players were asked to walk through with the white spot was created by displaying green LEDs on the cube at certain index numbers. The game had 2 different levels of difficulty, indicated by the index numbers that turned green and thus were not allowed to enter with the white dot: LoD $\in \{0, 1\}$. Level 0 was the easiest level. Here, the player only had to push the white dot around the corner of the toy in the middle of a side. In the hardest level, pushing around the corner of the toy could also have to take place at the edges of the cube. A schematic overview of both levels of the maze game is shown in Figure 4. The path in the maze was infinite and each level lasted for 60 seconds. Similar to the roadrunner game, each level occurred twice and the real levels were preceded to a warming up phase of 60 seconds. No permutations could be made to randomize the LoDs since there were only 2 LoDs; each game started with the easiest level and 2 subsequent levels could not have the same LoD.

Figure 3. Schematic overview of the maze game. A-C: show the way the cube should be rotated to correctly move the white dot through the maze. D: shows what happens in case the white dot was moved to a location wherein it was not allowed to enter.

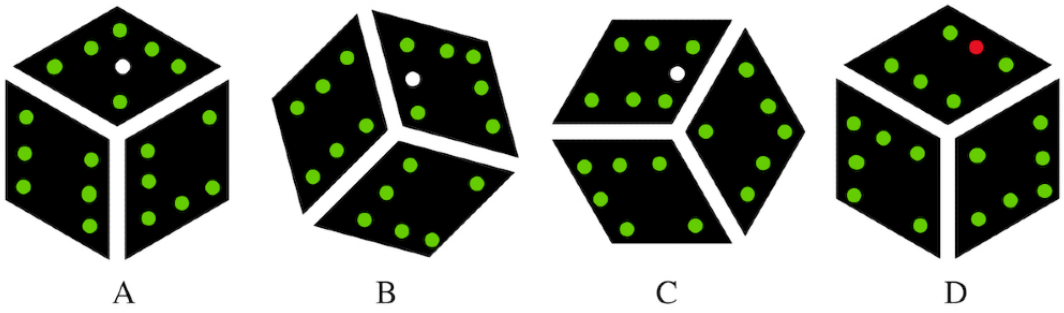
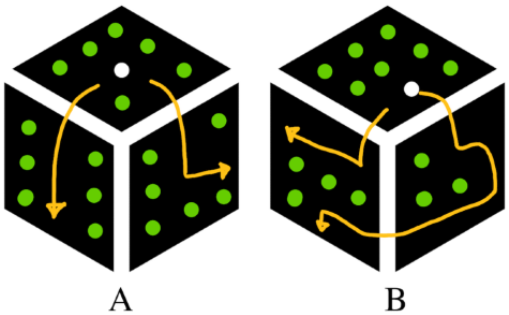


Figure 4. Schematic overview of the levels of the maze game. A: easiest level (level 0); B: hardest level (level 1).



Features

Before constructing features from the sensor data, we filtered the raw sensor data. Both the acceleration data and gyroscope data were filtered with a low-pass filter with a cut-off frequency of 4 Hz. The so-called feature jerk was calculated as the derivative of the acceleration and indicates the smoothness of the translational movements. To indicate the smoothness of the rotational movements, the derivative of the angular velocity was calculated. For each game, we have built 1 game feature.

The game feature for the roadrunner game is the cosine similarity. This feature indicates how accurately a player kept the green moving dot on the top surface of the cube. The cosine similarity indicates the similarity between the location of the dot and the location of the top of the cube. The cosine similarity has a range of –1 to 1, with –1 meaning opposite orientations and 1 meaning identical orientations. Thus, for the roadrunner game, a cosine similarity value of 1 indicates that the player kept the cube exactly in the preferred position regarding the location of the green dot. For the maze game, the game state feature is the maze correctness. This feature indicates the time that the white dot was on the correct path.

We added gender and age of the participant as general features to the sensor and game features. Table 2 shows all the features together with its meaning in the context, which game the feature applies to, and what type of feature it is. For each game, the data of the warming-up phase (ie, the first 60 seconds of each game) were removed since the warming up was intended to familiarize the participant with the toy and was not part of the assessment. The mean value of a variable over time (eg, mean acceleration, mean jerk) and standard deviation were calculated for each sensor and game feature. Thus, we constructed 8 sensor features, 2 game features, and 2 general features per game.

Table 2. Overview of all the features.

Feature name	Meaning in context	Game	Type of feature
a	Total acceleration (m/s^2)	Both	Sensor
ω	Total angular velocity (rad/s)	Both	Sensor
Jerk	Smoothness of the translational movements, derivative of a	Both	Sensor
α	Smoothness of the rotational movements, derivative of ω	Both	Sensor
Cosine similarity	Accuracy of keeping the green dot at the preferred position	Roadrunner	Game
Maze correctness	Time being on the correct path	Maze	Game
Gender	Gender	Both	General
Age	Age in years	Both	General

Classification Models

Four different supervised machine learning algorithms were compared: k-nearest neighbor (KNN), logistic regression (LR), decision tree (DT), and support vector machine (SVM). LR and DT were selected because they provide interpretable models, which is important for teachers who will use the toy in their classrooms. KNN and SVM were chosen because they are known to have good performance with nontextual data, of which SVM often performs well with relatively little data [22]. The labels that were used for training and testing the classification model were the binary outcomes of the fine MABC-2. A child who was likely to have fine motor skill development problems according to the fine MABC-2 was denoted as 1, while a child not having fine motor skill development problems according to the fine MABC-2 was labeled as 0. The performance of the classification model was analyzed with stratified 49-fold cross validation [23]. Since our data set is relatively small, we chose to perform cross validation to prevent overfitting. Because the data set included 49 children with label 1, we performed 49-fold cross validation to maximize the use of the available data. Ideally, the distribution of class labels is almost equal in the training and test set. Stratified cross-validation enables this ideal distribution while performing cross validation [24]. Since our data set consisted of approximately as many children with label 1 as children with label 0, each test set of a fold consisted of one child with label 1 and one child with label 0. The label of each child was used as a test set in at least one fold because we performed 49-fold cross validation and the data set contained 49 children with label 1. Since the data set included 46 children with label 0, the data of 3 children were reused in the test set of another fold to enable stratified cross-validation.

Comparing Performances

In the first analysis, the games and classifiers were compared. We trained and tested 4 different machine learning algorithms on all features of the roadrunner game and all features of the maze game. We used accuracy to indicate the percentage of the correctly classified cases. As an additional performance metric, we used the F1 score since it considers both precision and recall and these are both important in an assessment tool. Precision indicates the proportion of cases labeled positive that were actually correct, whereas recall indicates the proportion of actual positive cases that were labeled correctly. Wilcoxon tests were performed to show whether the best performing classifier performed significantly better than the other 3 classifiers. Moreover, Wilcoxon tests were performed to show the differences between the roadrunner game and the maze game for each classifier. For all statistical tests, we used α=.05. The game and classifier that performed best were used in the second analysis. Here, we compared the LoDs and the type of input features. For each possible combination of LoDs of the roadrunner game, we trained and tested the best performing classifier on the sensor features, the game features, and both sensor and game features. The general features such as age and gender were always included as classifier input. Wilcoxon tests were performed to show whether the best performing combination of the levels performed significantly better than the other combinations of levels. Furthermore, Wilcoxon tests were performed to show whether there were differences between the type of features for the best performing combination of levels.

Results

Participant Characteristics

In total, 95 children (52 girls and 43 boys) participated. Their mean age was 7.8 (SD 0.7) years. Based on the fine MABC-2 scores, 49 children showed problems with their fine motor skills, while 46 children did not have fine motor skill problems. Thus, the group of children having fine motor skill problems according to the fine MABC-2 score and the group of children not having fine motor skill problems were almost equal in number. Since we deliberately included urban elementary schools having a larger population of children with motor skill problems, this ratio is higher than the typical percentage of 5%-10% of the children having motor skill problems [5,6].

Effects of the Games and Classifiers

The DT classifier with only features of the roadrunner game as input performed best with a mean accuracy score of 0.68 and a mean F1 score of 0.65. The corresponding mean recall score was 0.74. The DT classifier also performed best for the maze game with a mean accuracy score of 0.52, a mean F1 score of 0.44, and a mean recall score of 0.51. For all classifiers, the highest mean accuracy and mean F1 scores were achieved with features of the roadrunner game used as input. An overview of all mean accuracy and mean F1 scores per game and machine learning algorithm is shown in Figure 5.

Although the DT classifier achieved higher accuracy and F1 scores than the other 3 classifiers for both the roadrunner and the maze games, no significant differences were found. All results of the performed Wilcoxon tests to study differences between the DT and the other classifiers are shown in Table 3.

Figure 5. Mean accuracy and mean F1 scores for the comparison of the classifiers and games. DT: decision tree; KNN: k-nearest neighbor; LR: logistic regression; SVM: support vector machine.

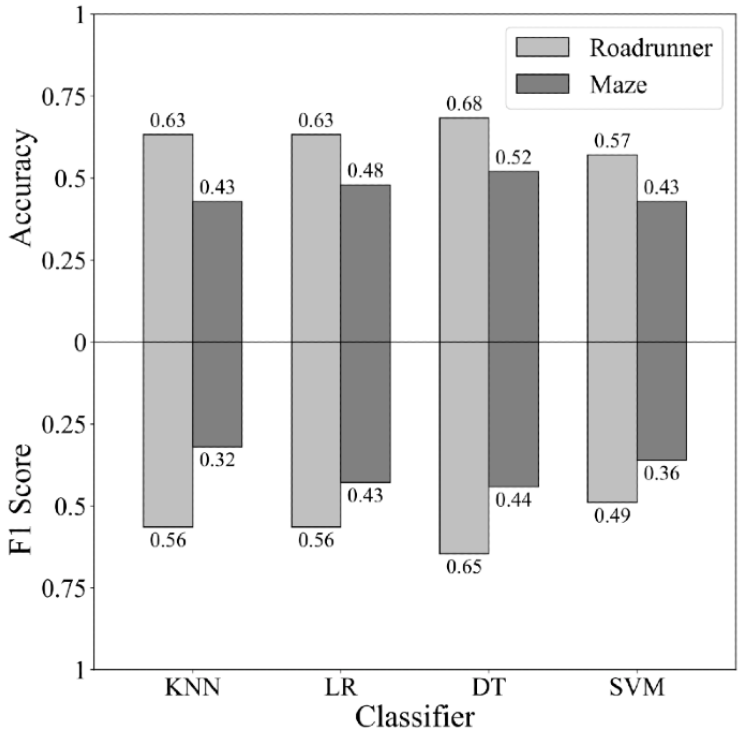


Table 3. Results of the Wilcoxon tests when the performance scores of the decision tree classifier were compared with those of the other classifiers.

Classifier	Accuracy (*P* value)		F1 score (*P* value)	
	Roadrunner game	Maze game	Roadrunner game	Maze game
k-nearest neighbor	.42	.15	.41	.14
Logistic regression	.35	.62	.18	.90
Support vector machine	.08	.26	.08	.23

Each classifier performed significantly better on accuracy with data obtained from playing the roadrunner game than with that obtained from playing the maze game (DT, *P*=.03; KNN, *P*=.01; LR, *P*=.02; SVM, *P*=.04). Except for the SVM classifier, each classifier also performed significantly better on the F1 score with data obtained from playing the roadrunner game than with data obtained from playing the maze game (DT, *P*=.02; KNN, *P*=.01; LR, *P*=.049). Table 4 shows all the results of the performed Wilcoxon tests to show differences between data obtained from playing the roadrunner game and the maze game.

Table 4. Results of the Wilcoxon tests when the performance scores of the roadrunner game were compared with those of the maze game for all classifiers.

Classifier	Accuracy (*P* value)	F1 score (*P* value)
Decision tree	.03[a]	.02[a]
k-nearest neighbor	.01[a]	.01[a]
Logistic regression	.02[a]	.049[a]
Support vector machine	.04[a]	.20

[a]Differences were statistically significant at $P<.05$.

Influence of the Game Levels and Features

The DT classifier with features of only the roadrunner game as input was used to compare the types of input features and combinations of levels since the combination of this classifier and game performed best in the first analysis. The highest mean accuracy, being 0.76, was achieved with a combination of data obtained from playing level 0 and level 2 and a combination of both sensor and game features. The corresponding mean F1 and recall scores were 0.67 and 0.71, respectively. The best mean F1 score, being 0.70, was achieved with the combination of level 1 and level 2 and only using game features. The corresponding mean accuracy and recall scores were 0.65 and 0.80, respectively. Figure 6 shows an overview of all mean accuracy and mean F1 scores per combination of levels and type of input features achieved with the DT classifier and data of the roadrunner game. Since the combination of data obtained from playing levels 0 and 2 with both sensor and game features achieved the highest mean accuracy, we compared levels 0 and 2 with the other combinations of levels. The combination of level 0 and level 2 only performed significantly better than the combination of level 0 and level 1 when both sensor and game features were used as input (P=.046). All results of the performed Wilcoxon tests to study differences between the combination of levels with both sensor and game features used as input are shown in Table 5.

Figure 6. Mean accuracy and F1 scores for the comparison of the levels and types of input features.

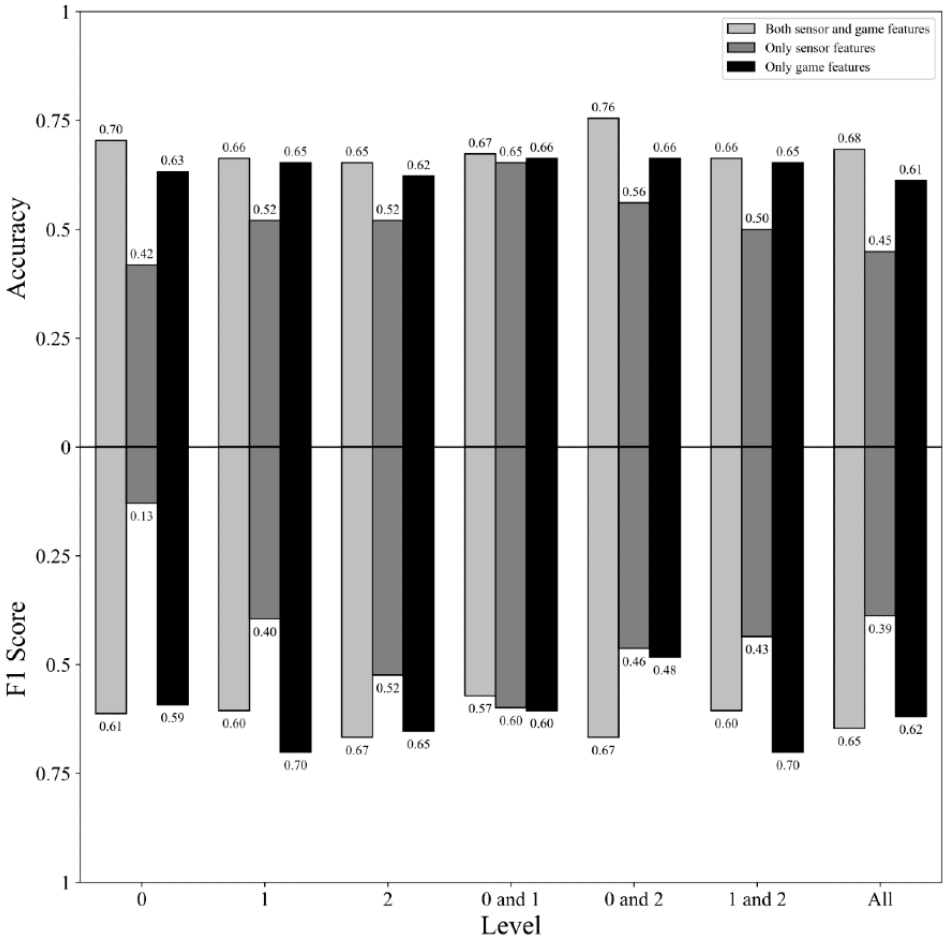


Table 5. Results of the Wilcoxon tests when the performance scores of game level 0 and level 2 were compared with those of the other levels.[a]

Game level	Accuracy (*P* value)	F1 score (*P* value)
Level 0	.38	.54
Level 1	.21	.52
Level 2	.11	.92
Level 0 + 1	.046[b]	.10
Level 1 + 2	.21	.52
Level 0 + 1 + 2	.20	.81

[a]In all cases, decision tree was used as the classifier and both sensor and game features of the roadrunner game were used as input.

[b]Differences were statistically significant at $P<.05$.

For all combinations of levels, the combination of both sensor and game features performed better regarding accuracy than only one of those features. When zooming in on the best performing combination of levels, that is, level 0 and level 2, we found a significant difference in the accuracy and F1 scores between using both types of features and using only sensor features (accuracy, $P=.001$; F1 score, $P=.01$). The results of the performed Wilcoxon tests to study differences between the type of input features for the combination of level 0 and level 2 are shown in Table 6.

Table 6. Results of the Wilcoxon tests when the performance scores of the types of input features were compared.[a]

Input feature	Accuracy (*P* value)	F1 score (*P* value)
Both features versus only sensor features	.001[b]	.01[b]
Both features versus only game features	.18	.06
Sensor versus game	.10	.91

[a]In all cases, decision tree was used as the classifier and both data from game level 0 and level 2 were used for the input features.

[b]Differences were statistically significant at $P<.05$.

Discussion

Principal Results

By comparing the classifiers and games, we learned that the game focusing on speed was more suitable for predicting the motor skill level than the game focusing on precision. Data obtained from playing the roadrunner game led to significantly better performances than data obtained from playing the maze game. Thus, adding the game focusing on precision did not improve our preliminary results. The important contribution of the roadrunner data to the classification performance may be explained by the fact that speed is an important component in the MABC-2 as well. Two out of 3 subtests of the fine MABC-2 are time-sensitive. These findings correspond to the results of Rivera et al [18], who showed that the time for completing a task was an important component for intraindividual variability with their tested sensor-augmented toy.

By comparing the types of data and the LoDs, we learned that the combination of both sensor and game features was the most suitable for predicting the motor skill level. For the combination of data obtained from playing level 0 and level 2, using both sensor and game features led to a significantly better performance than only using sensor features. Although the contribution of the sensor features to the performance was shown to be little, the addition of the gyroscope data led to improved results compared to our preliminary results [9]. The significant difference between using both sensor and game features and only using sensor features for the best performing combination of levels is interesting. This means that the game component of the assessment approach is not only beneficial for playfulness but it also plays an important role in the prediction of the motor skill level itself together with the sensor data. An interesting follow-up project could be to generate additional game features or design more speed-based games and study how they affect the prediction of the fine MABC-2 score. The fact that no significant differences were found between the DT classifier and the other 3 classifiers indicates that the selection of input features has more impact on the performance than the selection of the classifier. Although the DT classifier did not perform significantly better, it is preferred over the KNN and SVM algorithms since it gives insight into the classification process. This is an important characteristic for the teachers who will use the toy in their classrooms.

Strengths, Limitations, and Opportunities

The best achieved accuracy of 0.76 and F1 score of 0.70 on predicting the label of the fine MABC-2 are promising for assessing children's motor skills with sensor-augmented toys. Since the cube is easy to use in the classroom, it is relatively easy to collect data. Therefore, the current approach might not only be useful for one-time assessment but could also be used for monitoring. Although we predicted the outcome of the fine MABC-2 and this assessment indicates whether children have fine motor development problems or not, we cannot state that we can predict children's fine motor skill level with playing games with the Futuro Cube. To do so, we should repeat our research and replace the fine MABC-2 labels with expert view

labels of the motor skill level. Moreover, assessment with the Futuro Cube only indicates whether children might have fine motor development problems in general. However, fine motor development is complex and consists of several aspects. Important factors are, for instance, cognitive ability, anticipatory control, motor planning function, and spatial ability [4]. Since all of these factors are included in the tasks of the fine MABC-2, we included them in the games of the Futuro Cube as well. Thus, our toy only signals motor development problems in general but does not assess specific aspects of motor development. In case the assessment tool showed that a child was likely to have fine motor development problems, follow-up examination is required to investigate individual aspects or causes of fine motor development problems.

The current approach might not only be useful for assessment but might also be useful as a first screening tool. In that case, children who are not likely to have motor problems are already being filtered out with the results of the game. A valid and reliably fine motor skill assessment test such as the fine MABC-2 can be taken for the children who were likely to have fine motor skill problems based on the game results. That way, not all children have to take the fine MABC-2 test, and the false positives of the assessment with the toy can be filtered out afterwards.

Another promising opportunity of the Futuro Cube is using it for training instead of monitoring. Developing a valid and reliable assessment toy will take some time, but playing with the Futuro Cube might also be useful for training purposes. Children could train their fine motor skills while playing games with the toy. This could be valuable for both children with motor skill problems without having specific disorders as well as children with, for instance, cerebral palsy or fine motor problems after a stroke. For learning, fun is an important factor since it improves intrinsic motivation and focus [25]. It is shown that gamified training is highly engaging and boosts the motivation of players [12]. Therefore, such a playful way of training their motor skills would be a valuable addition to the current methods. When the toy is ready for assessment, it could also be used to monitor progress in therapy or rehabilitation of such children. Since data are wirelessly sent in real time to a computer, such training opportunities could be improved by making the game adaptive. The level could be fitted to the child's capacities, which improves the attention span and motivation. In this study, we focused on predicting the outcome of the fine MABC-2, but we did not include feasibility and usability in our approach. Although we did not study playfulness for children and usability for teachers in our approach, both children and teachers were very enthusiastic and their informal responses were, without exception, positive.

Conclusions

This study examined the possibilities of using sensor-augmented toys to assess children's fine motor skills. Such toys are less time-consuming and more playful and motivating than the current assessment methods. Compared to our preliminary research, we added the gyroscope for extra sensor data and an extra game that focused on precision instead of speed. With the best achieved accuracy of 0.76 and F1 score of 0.70, we showed that sensor-augmented toys can efficiently predict the outcome of the fine MABC-2 score. The selection of features is more important for the performance than the selection of the machine learning classifier. Classifiers that used input features obtained from playing the game focusing on speed performed significantly better that classifiers that used input features obtained from playing the game focusing on precision. Although our findings are a good start, further research is needed to develop a reliable and valid playful assessment tool. Possible improvements may be generating more game features, designing more speed-based games, and making the LoDs adaptive. Such adaptive games may also be valuable for training or rehabilitation purposes.

Acknowledgments

We thank all the children and teachers who participated in this study. Moreover, we thank Jan Krejsa for his valuable contribution to this study. This study is part of the research project Smart Play Sets, funded by COMMIT.

Conflicts of Interest

None declared.

References

1. Cools W, Martelaer KD, Samaey C, Andries C. Movement skill assessment of typically developing preschool children: a review of seven movement skill assessment tools. J Sports Sci Med 2009 Jun 01;8(2):154-168 [FREE Full text] [Medline: 24149522]
2. Grissmer D, Grimm KJ, Aiyer SM, Murrah WM, Steele JS. Fine motor skills and early comprehension of the world: two new school readiness indicators. Dev Psychol 2010 Sep;46(5):1008-1017. [doi: 10.1037/a0020104] [Medline: 20822219]
3. Gerber RJ, Wilks T, Erdie-Lalena C. Developmental milestones: motor development. Pediatr Rev 2010 Jul;31(7):267-277. [doi: 10.1542/pir.31-7-267] [Medline: 20595440]
4. Blank R, Barnett AL, Cairney J, Green D, Kirby A, Polatajko H, et al. International clinical practice recommendations on the definition, diagnosis, assessment, intervention, and psychosocial aspects of developmental coordination disorder. Dev Med Child Neurol 2019 Mar 22;61(3):242-285 [FREE Full text] [doi: 10.1111/dmcn.14132] [Medline: 30671947]

5. Geuze RH, Jongmans MJ, Schoemaker MM, Smits-Engelsman BC. Clinical and research diagnostic criteria for developmental coordination disorder: a review and discussion. Human Movement Science 2001 Mar;20(1-2):7-47. [doi: 10.1016/s0167-9457(01)00027-6]
6. Wilson PH, Ruddock S, Smits-Engelsman B, Polatajko H, Blank R. Understanding performance deficits in developmental coordination disorder: a meta-analysis of recent research. Dev Med Child Neurol 2013 Mar;55(3):217-228 [FREE Full text] [doi: 10.1111/j.1469-8749.2012.04436.x] [Medline: 23106668]
7. Henderson S, Sugden A, Barnett A. Movement Assessment Battery for Children - 2nd edition. London: Pearson; 2007.
8. Seo S. The effect of fine motor skills on handwriting legibility in preschool age children. J Phys Ther Sci 2018 Feb;30(2):324-327 [FREE Full text] [doi: 10.1589/jpts.30.324] [Medline: 29545705]
9. Sander J, de Schipper A, Brons A, Mironcika S, Toussaint H, Schouten B, et al. Detecting delays in motor skill development of children through data analysis of a smart play device. 2017 Presented at: Proceedings of the 11th EAI International Conference on Pervasive Computing Technologies for Healthcare; May; NY, USA p. 88-92. [doi: 10.1145/3154862.3154867]
10. Mironcika S, de Schipper A, Brons A, Toussaint H, Kröse B, Schouten B. Smart toys design opportunities for measuring children's fine motor skills development. 2018 Presented at: Proceedings of the Twelfth International Conference on Tangible, Embedded, and Embodied Interaction; March; NY, USA p. 349-356. [doi: 10.1145/3173225.3173256]
11. Cerrato A, Ferrara F, Ponticorvo M, Simona Sica L, Di Fernando A, Miglino O. DILIGO Assessment Tool: A Smart and Gamified Approach for Preschool Children Assessment. Cham: Springer; 2017:235-244.
12. Lumsden J, Edwards EA, Lawrence NS, Coyle D, Munafò MR. Gamification of Cognitive Assessment and Cognitive Training: A Systematic Review of Applications and Efficacy. JMIR Serious Games 2016 Jul 15;4(2):e11 [FREE Full text] [doi: 10.2196/games.5888] [Medline: 27421244]
13. Johnson D, Deterding S, Kuhn K, Staneva A, Stoyanov S, Hides L. Gamification for health and wellbeing: A systematic review of the literature. Internet Interv 2016 Nov;6:89-106 [FREE Full text] [doi: 10.1016/j.invent.2016.10.002] [Medline: 30135818]
14. Sardi L, Idri A, Fernández-Alemán J. A systematic review of gamification in e-Health. J Biomed Inform 2017 Jul;71:31-48 [FREE Full text] [doi: 10.1016/j.jbi.2017.05.011] [Medline: 28536062]
15. Ayed I, Ghazel A, Jaume-I-Capó A, Moyà-Alcover G, Varona J, Martínez-Bueso P. Vision-based serious games and virtual reality systems for motor rehabilitation: A review geared toward a research methodology. Int J Med Inform 2019 Nov;131:103909. [doi: 10.1016/j.ijmedinf.2019.06.016] [Medline: 31557701]
16. Page ZE, Barrington S, Edwards J, Barnett LM. Do active video games benefit the motor skill development of non-typically developing children and adolescents: A systematic review. J Sci Med Sport 2017 Dec;20(12):1087-1100. [doi: 10.1016/j.jsams.2017.05.001] [Medline: 28600111]
17. Vega-Barbas M, Pau I, Ferreira J, Lebis E, Seoane F. Utilizing Smart Textiles-Enabled Sensorized Toy and Playful Interactions for Assessment of Psychomotor Development on Children. Journal of Sensors 2015;2015:1-9. [doi: 10.1155/2015/898047]
18. Rivera D, García A, Ortega JE, Alarcos B, van der Meulen K, Velasco JR, et al. Intraindividual Variability Measurement of Fine Manual Motor Skills in Children Using an Electronic Pegboard: Cohort Study. JMIR Mhealth Uhealth 2019 Aug 28;7(8):e12434 [FREE Full text] [doi: 10.2196/12434] [Medline: 31464193]
19. Gutiérrez García M, Martín Ruiz M, Rivera D, Vadillo L, Valero Duboy M. A Smart Toy to Enhance the Decision-Making Process at Children's Psychomotor Delay Screenings: A Pilot Study. J Med Internet Res 2017 May 19;19(5):e171 [FREE Full text] [doi: 10.2196/jmir.7533] [Medline: 28526666]
20. Williams JR. The Declaration of Helsinki and public health. Bull World Health Organ 2008 Aug;86(8):650-652 [FREE Full text] [doi: 10.2471/blt.08.050955] [Medline: 18797627]
21. Krejsa J. Motion-controlled gaming cube. Rubik's Futuro cube. URL: https://www.futurocube.com/ [accessed 2021-04-05]
22. Choosing the right estimator. Scikit Learn. URL: https://scikit-learn.org/stable/tutorial/machine_learning_map/index.html [accessed 2021-04-05]
23. Pedregosa F, Varoquaux G, Gramfort A. Scikit-learn: machine learning in Python. Journal of Machine Learning Research. 2011. URL: https://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf [accessed 2021-04-05]
24. Breiman L, Friedman J, Olshen R, Stone CJ. Classification and Regression Trees. London, UK: Chapman and Hall/CRC; 1984.
25. Prensky M. Digital game-based learning. Comput. Entertain 2003 Oct;1(1):21-21. [doi: 10.1145/950566.950596]

Abbreviations

DT: decision tree
KNN: k-nearest neighbor
LoD: level of difficulty
LED: light-emitting diode
LR: logistic regression
MABC-2: Movement Assessment Battery for Children Second Edition

SVM: support vector machine

Edited by G Eysenbach; submitted 10.09.20; peer-reviewed by J Velasco, F Blumberg; comments to author 11.10.20; revised version received 10.11.20; accepted 15.01.21; published 22.04.21.

Please cite as:
Brons A, de Schipper A, Mironcika S, Toussaint H, Schouten B, Bakkes S, Kröse B
Assessing Children's Fine Motor Skills With Sensor-Augmented Toys: Machine Learning Approach
J Med Internet Res 2021;23(4):e24237
URL: https://www.jmir.org/2021/4/e24237
doi:10.2196/24237
PMID:33885371

Original Paper

Web-Based Smartphone Algorithm for Calculating Blood Pressure From Photoplethysmography Remotely in a General Adult Population: Validation Study

Paul Holyoke[1], PhD; Karthika Yogaratnam[1], MPH; Elizabeth Kalles[1], BSc

SE Research Centre, SE Health, Markham, ON, Canada

Corresponding Author:
Paul Holyoke, PhD
SE Research Centre
SE Health
90 Allstate Parkway
Markham, ON, L3R 6H3
Canada
Phone: 1 4168923791
Email: paulholyoke@sehc.com

Abstract

Background: Outside of a clinical setting, oscillometric devices make remote monitoring of blood pressure and virtual care more convenient and feasible. HeartBeat Technologies Ltd developed a novel approach to measuring blood pressure remotely after an initial blood pressure reading by a nurse using the conventional measurement method. Using a finger pulse oximeter, a photoplethysmogram wave is transmitted by Bluetooth to a smartphone or tablet. A smartphone app (MediBeat) transmits the photoplethysmogram to a server for analysis by a proprietary algorithm—the person's current blood pressure is sent back to the smartphone and to the individual's health care provider.

Objective: This study sought to determine whether the HeartBeat algorithm calculates blood pressure as accurately as required by the European Society of Hypertension International Protocol revision 2010 (ESH-IP2) for validation of blood pressure measuring devices.

Methods: ESH-IP2 requirements, modified to conform to a more recent international consensus statement, were followed. The ESH-IP2 establishes strict guidelines for the conduct and reporting of any validation of any device to measure blood pressure, including using the standard manual blood pressure instrument as a comparator and specific required accuracy levels for low, medium, and high ranges of blood pressure readings. The consensus statement requires a greater number of study participants for each of the blood pressure ranges. The validation of the accuracy of the algorithm was conducted with a Contec CMS50EW pulse oximeter and a Samsung Galaxy XCover 4 smartphone.

Results: The differences between the HeartBeat-calculated and the manually measured blood pressures of 62 study participants did not meet the ESH-IP2 standards for accuracy for either systolic or diastolic blood pressure measurements. There was no discernible pattern in the inaccuracies of the HeartBeat-calculated measurements.

Conclusions: The October 4, 2019 version of the HeartBeat algorithm, implemented in combination with the MediBeat app, a pulse oximeter, and an Android smartphone, was not sufficiently accurate for use in a general adult population.

Trial Registration: ClinicalTrials.gov NCT04082819; http://clinicaltrials.gov/ct2/show/NCT04082819

(J Med Internet Res 2021;23(4):e19187) doi:10.2196/19187

KEYWORDS

blood pressure measurement; remote monitoring; hypertension

Introduction

Accurately assessing blood pressure is necessary for the proper diagnosis of hypertension, implementing treatments and monitoring whether those treatments are working [1,2]. To measure blood pressure accurately, standardized measurement techniques, calibrated equipment, and valid interpretation of readings are necessary [2-4]. Aneroid sphygmomanometers that are properly calibrated and maintained are considered to be at

least as accurate as mercury-filled sphygmomanometers [5]. However, such instruments are not convenient for remote monitoring or virtual (remote) care because they can be difficult for a person to self-administer.

A novel approach to measuring blood pressure remotely has been developed (HeartBeat, HeartBeat Technologies Ltd). The measurement is taken after an initial blood pressure reading using the conventional measurement method, supplemented by specific characteristics of a person (age, gender, height, weight, and heart rate) to establish a baseline for the person. Unlike other devices [6-9], this approach uses a finger pulse oximeter that detects the changes in blood volume directly below the person's skin and indirectly measures oxygen saturation in the blood. The resulting photoplethysmogram (PPG) wave is transmitted by Bluetooth technology to a smartphone or tablet. The MediBeat (HeartBeat Technologies Ltd) smartphone or tablet app then transmits the PPG wave to a server where a proprietary algorithm analyzes the baseline measurement for the person and the PPG waveform to calculate the person's current blood pressure. The blood pressure measurement is then transmitted to the person's smartphone, and if applicable, to the person's health care provider's devices.

The purpose of this study was to test whether the accuracy of the HeartBeat algorithm, available as of October 4, 2019 and for a general adult population (ages 25 years and older), was sufficient to seek regulatory approval when compared with a standard aneroid sphygmomanometer with stethoscope for measuring blood pressure.

Methods

Protocol

The European Society of Hypertension International Protocol revision (ESH-IP2) for the validation of blood pressure measuring devices in adults [10] was followed, with modifications to account for additional guidance from an international consensus statement [11]. ESH-IP2 [10] requires strict adherence to its method and reporting requirements, which are reflected in this report: to "ensure a uniform distribution of test pressures across a representative range," 33 participants must be enrolled. Among the 33 participants, at least 10 must be male and 10 must be female; all should be at least 25 years of age; and between 10 and 12 participants must have blood pressure readings within each of the following recruitment ranges—systolic blood pressure: below 130 mmHg, between 130 and 180 mmHg, above 180 mmHg; diastolic blood pressure: below 80mmHg, between 80 and 130 mmHg, above 130 mmHg. All potential participants must be screened against these criteria and can be excluded if the gender criteria or age criteria are not met, or if including them in the study would exceed the maximum number of participants in any of the recruitment ranges.

An international consensus statement group [11] reviewed the adequacy of the number of participants in a variety of established protocols, including that of the ESH-IP2, and recommended that, to increase the power and accuracy of validation studies, at least 85 participants, rather than 33, should be included. To align with the consensus statement, the research team recalculated the ESH-IP2 requirement numbers to reflect the increased participant pool of 85. The recalculations are reflected in the tables and discussion in this report.

Devices

The pulse oximeter (Contec CMS50EW, Contec Medical Systems Co Ltd [12]) used to capture and transmit PPG is shown in Figure 1. The 8-bit Contec pulse oximeter is approved for use by Health Canada [13] and the US Food and Drug Administration [14]. The smartphone used in this study (upon which HeartBeat's MediBeat app was installed and used to transmit the data to HeartBeat servers and to receive and display the results) was a Samsung Galaxy XCover 4 (Figure 2).

Figure 1. Contec CMS50EW.

Figure 2. Samsung Galaxy XCover 4.

Recruitment

Potential participants were recruited by mass email and offline posters at a single-site multistory office building in Markham, Ontario, Canada, where they worked; participants provided informed consent (Multimedia Appendix 1) and were scheduled for a measurement appointment during their working day by a member of the research team.

Procedure

Three registered nurses who had training and expertise in blood pressure measurement carried out the measurements required for this study between October 4, 2019 to November 22, 2019. One nurse was the supervisor, and two nurses were the observers throughout the duration of the study. The nurses were trained as per ESH-IP2 directives and were instructed on how to follow the protocol guidelines for data collection. Blood pressure measurements were recorded to the nearest 2 mmHg.

Two new sphygmomanometers (Prosphyg 775 Model, American Diagnostics Corporation [15]), whose components were checked against one another and calibrated before the study, were used as the reference instruments by the observers. With respect to reliability and validity, the ESH-IP2 embeds an ongoing reliability mechanism requiring that any individual participants' blood pressure readings be excluded from the analysis if the difference in readings from the two observer nurses differ by 4 mmHg twice. The instruments were placed within one meter of the observers while they interacted with a participant, which allowed the observers to follow the instruments' dials at eye level from 40 mmHg to 180 mmHg. Multiple sizes of bladders (cuffs) were available for the sphygmomanometers to ensure that 80% to 100% of each participant's arm circumference could be encircled. The observers were also supplied with good quality nonelectronic stethoscopes with well-fitted earpieces.

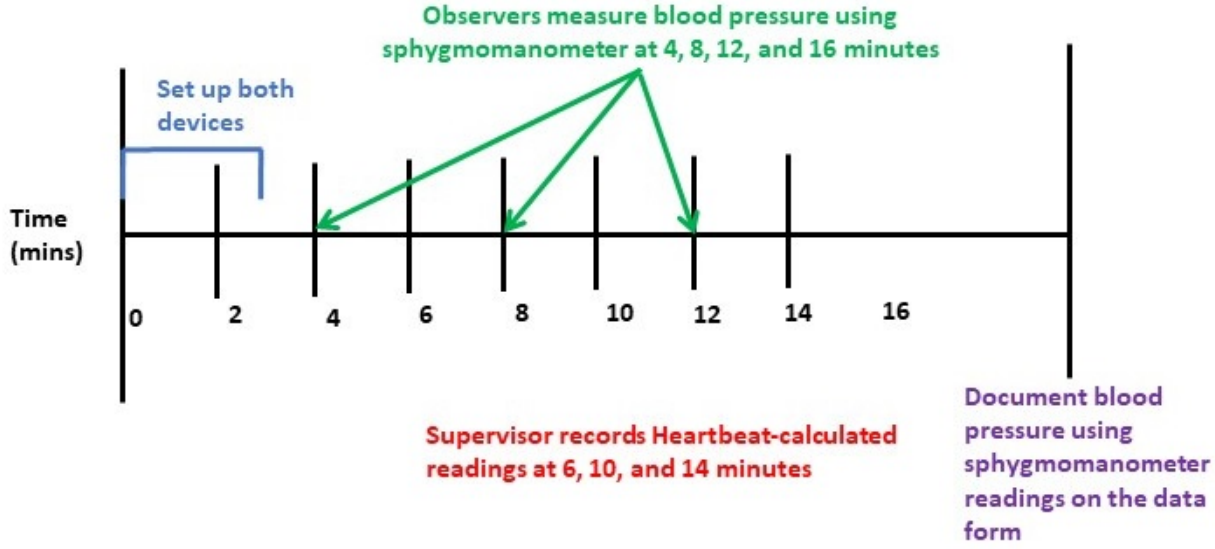
The observers were instructed that, in the event that an abnormal blood pressure reading was detected (for example, if the blood pressure was outside clinically normal range for a specific person), they were to remeasure the blood pressure using conventional means to determine if the abnormal reading was accurate. If the abnormal blood pressure measurement was accurate, the nurse supervisor was to act according professional standards. The abnormal blood pressure measurement should also be recorded on ESH-IP2 Form 2.

Participants' blood pressure measurements were taken in a room that was a comfortable temperature, without any noises or other influences that would have caused disturbances, such as telephones. The temperature in the room was the same as that throughout the building—20 °C to 24 °C. All participants participated in the study between the working hours of 9 AM and 4:30 PM. The pulse oximeter and the supervisor's smartphone were charged and tested each day prior to interactions with participants. The MediBeat app was calibrated for each participant as follows: (1) Reference data for the

individual (year of birth, height, weight, baseline manual blood pressure reading, and heart rate) were entered. (2) The participant rested for 10 minutes while seated comfortably with their legs uncrossed, back supported, and arm resting on the table at heart level. (3) The supervisor attached the finger pulse oximeter to the participant's index finger (on the opposite arm to the manual blood pressure cuff) and waited for a few seconds until a stable signal was received and the device displayed both the participant's pulse and SpO_2 (oxygen saturation) values. (4) Once the values were displayed, the supervisor initiated a new session on the smartphone to connect the phone to the Contec device via Bluetooth. This connection started the 1-minute calibration process, at which point each participant was asked to remain still until the calibration had been completed successfully.

Overseen by an independent supervisor, measurements were recorded by the two observers who were blinded from both the other observer's readings and from the reading calculated by the HeartBeat algorithm. Each participant's blood pressure was measured as illustrated in Figure 3. Each systolic blood pressure measurement and each diastolic blood pressure measurement calculated by the HeartBeat algorithm was compared with the nearest of the previous and next observer systolic and diastolic measurements, respectively, and the difference between each HeartBeat-calculated measurement and the nearest observer measurement was calculated. Every such difference was then classified in one of these error groups for each systolic and diastolic measurements: 0-5 mmHg, 6-10 mmHg, 11-15 mmHg, and >15 mmHg. The results were then compared with the number of tolerable differences allowed by the protocol at both the level of individual measurements (part 1) and at the level of the participants (part 2). The device gets an overall *pass* (an acceptable level of accuracy) if it meets 4 pass criteria.

Figure 3. Timeline of measurement of blood pressure for each participant.



Ethics

The study was reviewed and approved by the Southlake Regional Health Centre Research Ethics Board, Newmarket, Ontario, Canada (007-1920).

Results

A total of 105 individuals volunteered and were screened for the study, of whom 62 participants met the ESH-IP2 criteria, and 43 were excluded (Table 1; adapted version of the CONSORT flow diagram [16] in Figure 4). As required by the protocol, 5 participants out of the 105 were excluded because the differences in the measurements by the observer nurses exceeded 4 mmHg twice for the same participant, 34 were excluded participants were excluded because their blood pressure measurements fell within ranges for which the protocol quota had been fulfilled, and 4 participants were excluded because they were younger than 25 years of age. There was difficulty in recruiting hypertensive participants, that is, individuals with both systolic and diastolic blood pressures in the high ranges. This is reflected in the overall distribution (Figure 5 and Figure 6), in which most of the points fall below 125 mmHg and 90 mmHg, for systolic and diastolic blood pressures, respectively. By the time 105 participants had been recruited, only 2 met the criterion for high systolic blood pressure, and none met the criterion for diastolic blood pressure. The modified criteria would have required between 26 and 31 participants in the high blood pressure group to meet the protocol requirements. A calculation was performed to determine if it would be possible for the algorithm to achieve a pass if additional participants with high blood pressure were recruited, and there were a perfect match between their HeartBeat-calculated measurements and standard measurements. It was determined that such a result was not possible, and recruitment was discontinued. Therefore, the study concluded before the required number of participants was reached. The characteristics of participants are described in Tables 2 and 3.

Table 1. Participants screened, excluded, and recruited for the study.

Reason for exclusion	Excluded, n	Total, n
Participants screened		105
Participants excluded		43
Ranges complete[a]	34	
Range adjustment	0	
Arrhythmias	0	
Device failure	0	
Poor quality sounds	0	
Cuff size unavailable	0	
Observer disagreement[b]	5	
Distribution[c]	0	
Other reasons[c]	4	
Participants recruited		62

[a]Intake assessment placed them in ranges that had already been filled. The ESH-IP2 requires that the last participants to be recruited are the ones who are excluded.

[b]Discrepancy >4 mmHg twice for the same participant; therefore, these participants were, in accordance with the ESH-IP2, excluded from the study.

[c]Age <25 years or older, as required by the ESH-IP2.

Figure 4. Adapted version of the CONSORT flow diagram [16].

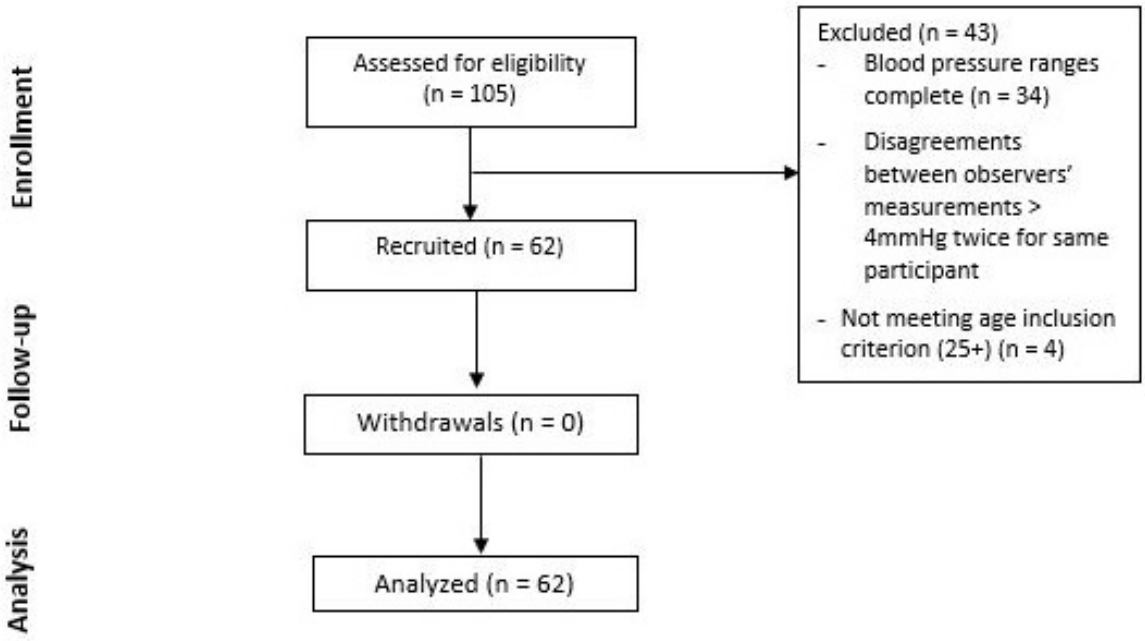


Figure 5. Bland-Altman plot of the differences between the Heartbeat-calculated measurements and the manual measurements of systolic blood pressure.

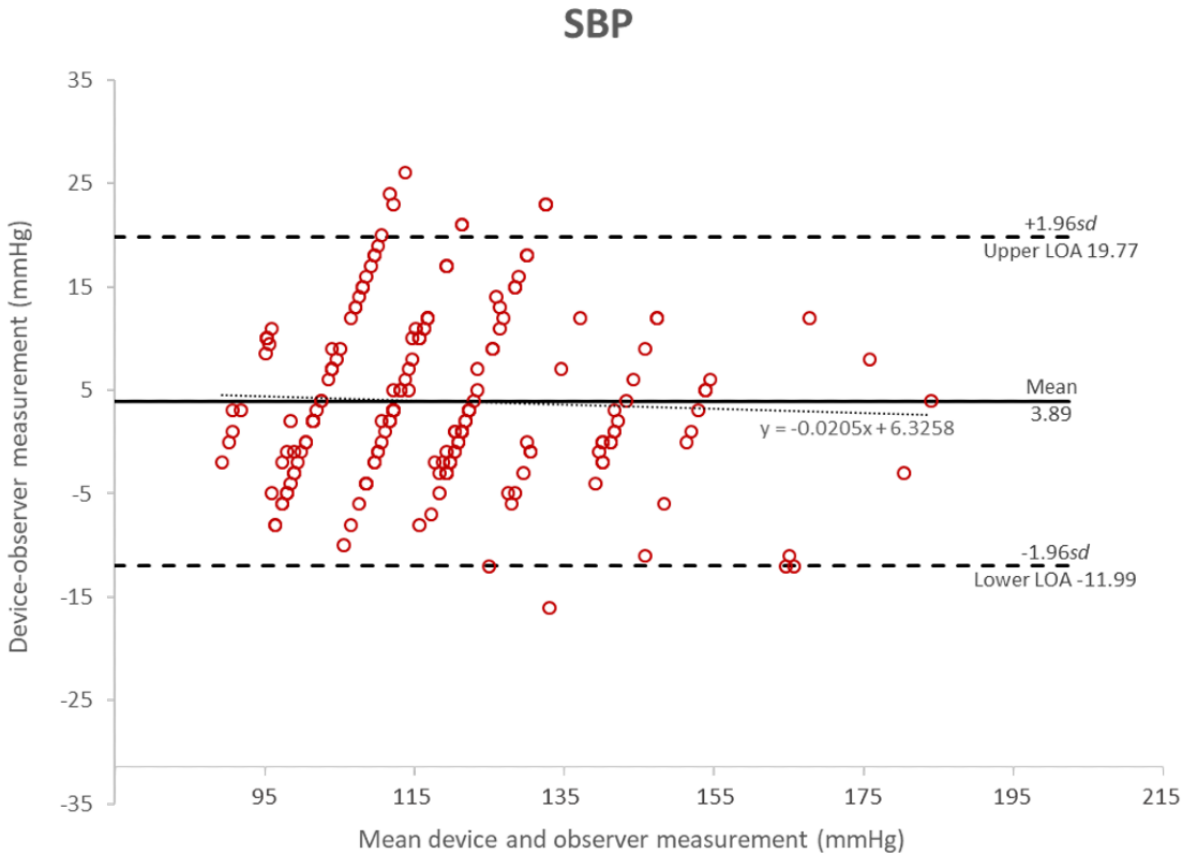


Figure 6. Bland-Altman plot of the differences between the Heartbeat-calculated measurements and the manual measurements of diastolic blood pressure.

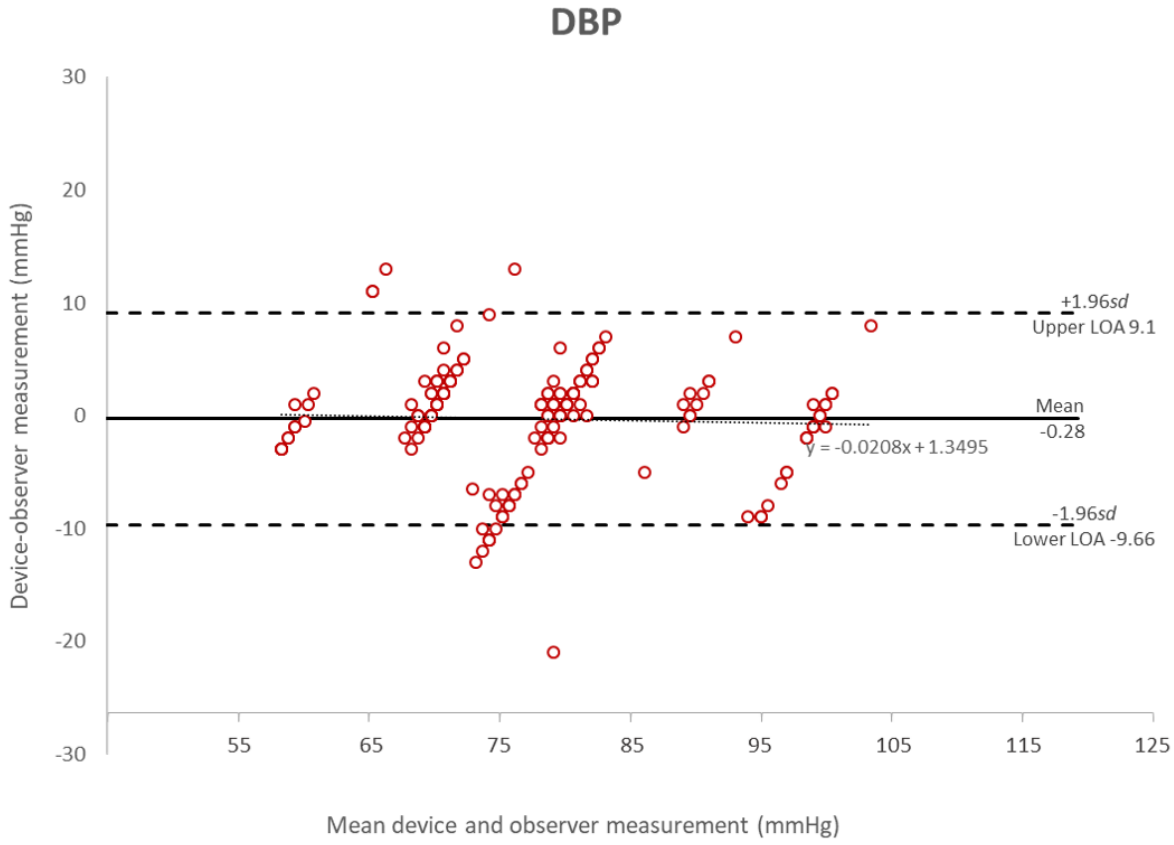


Table 2. Blood pressure and medication details for initially screened participants.

Category	Blood pressure (mmHg), range	Medication[a]	No medication	Total
Systolic blood pressure				
Low-low	<90	0	0	0
Low	90-129	5	40	45
Medium	130-160	7	9	15
High	161-180	1	1	2
High-high	>180	0	0	0
Diastolic blood pressure				
Low-low	<40	0	0	0
Low	40-79	3	30	33
Medium	80-100	10	19	29
High	101-130	0	0	0
High-high	>130	0	0	0

[a]Participants taking medication for hypertension.

Table 3. Participant details.

Category	Value
Sex, n	
Male	13
Female	49
Age (years)	
Mean (SD)	43.5 (11.2)
Range	25-66
Arm circumference (cm)	
Mean (SD)	29.5 (4.2)
Range	23-40
Cuff size used, n	
Small	0
Standard	61
Large	0
Missing data	1
Recruitment blood pressure (mmHg)	
Systolic blood pressure	
Mean (SD)	119.5 (18.9)
Range	90-180
Diastolic blood pressure	
Mean (SD)	77.6 (10.6)
Range	57-110

Table 4 and Table 5 present the blood pressure readings by the international consensus statement's predetermined levels. Table 4 shows the distribution of systolic and diastolic blood pressure readings across the recruited participants in the study. Table 5 shows the range of differences by observers 1 and 2.

Table 4. Blood pressure readings (3 per participant) after the initial screening measurement.

	Overall range (mmHg) (low-high)	Low[a], n	Medium[b], n	High[c], n	Maximum difference among highest and lowest totals in low, medium, and high ranges
Systolic blood pressure	90-180	150	30	6	144
Diastolic blood pressure	60-101	90	75	21	69

[a]Systolic ≤130 mmHg; diastolic ≤80 mmHg.

[b]Systolic 130 mmHg-160mmHg, diastolic 80 mmHg-100 mmHg.

[c]Systolic ≥160 mmHg, diastolic ≥100 mmHg.

Table 5. Observer differences.

	Range (low to high) (mmHg)	Mean (SD)	Repeated measurements
Systolic blood pressure[a]	–3 to 3	0.0081 (0.4)	—[b]
Diastolic blood pressure[a]	–4 to 3	0.0000 (0.6)	—[b]

[a]Measurements are calculated by subtracting observer 1 measurements from observer 2 measurements.

[b]There were no documented repeated measurements to calculate.

Table 6 shows the results in terms of each of the protocol requirements. According to the ESH-IP2 standards, for an acceptable level of accuracy, a device must pass all the protocol requirements in both parts 1 and 2. Note that in this study, the algorithm achieved a pass in only 1 category and a fail in 3 categories, for an overall fail.

Table 6. Validation results.

Protocol requirements	Range, n			Pass or fail
	<5 mmHg	<10 mmHg	<15 mmHg	
Part 1 (N=186[a])				
Required (2 of 3)	At least 137	At least 163	At least 180	
Achieved differences				
Systolic blood pressure[a]	120	147	169	Fail
Diastolic blood pressure[a]	159	180	186	Pass
Required (all)	At least 122	At least 152	At least 175	
Achieved differences				
Systolic blood pressure	120	147	169	
Diastolic blood pressure	159	180	186	
Part 2 (N=62[a])				
Required	At least 45			Fail
Achieved				
Systolic blood pressure	43	—[b]	—	
Diastolic blood pressure	10	—	—	
Allowed	At most 6			Fail
Achieved				
Systolic blood pressure	54	—	—	
Diastolic blood pressure	5	—	—	
Grade 3 (final result)—must pass all of parts 1 and 2				Fail

[a]These reflect the number of readings and participants included in the study when recruitment was stopped.

[b]Not applicable.

Figure 5 shows the Bland-Altman plots [17,18] of the differences in systolic blood pressure measurements between the MediBeat device and the manual observer measurements (*y*-axis) and the average of the 2 measurements (*x*-axis). Figure

6 shows the plots for the differences in diastolic measurements. In both cases, the averages of the 3 MediBeat measurements were plotted against their absolute differences from the manual measurements for both systolic blood pressure (bias 3.89, LOA –11.98 to 19.77) and diastolic blood pressure (bias –0.28, LOA –9.66 to 9.09). Although values fall within the 95% confidence interval for both Bland-Altman plots, there is no statistically significant correlation between the HeartBeat-calculated blood pressure measurements and the manual measurements (systolic blood pressure: r=0.04, P=.53; diastolic blood pressure: r=.04, P=.54). This indicates that the HeartBeat-calculated blood pressure measurements are not sufficiently accurate. There was no discernible pattern in the inaccuracies in the HeartBeat-calculated measures; therefore, the source of the inaccuracies is not known.

Discussion

The Bland-Altman plots show that the October 4, 2019 version of the HeartBeat algorithm, used in conjunction with a Contec CMS50EW pulse oximeter, MediBeat app, and a Samsung Galaxy XCover 4 smartphone, was not sufficiently accurate to meet ESH-IP2 standards for measuring the blood pressure of adults in the general population; therefore, as per the protocol, it is not recommended for personal or clinical use in this configuration.

Acknowledgments

We would like to acknowledge the nurses who conducted the measurements for this study: Favour Onyido, RN, Laura Agborchi, RN, and Valerie Mgbatogu, RN. We would also like to acknowledge the contributions of Brianna Croft, MMI, and Mary Lou Ackerman, MBA, who helped with the logistics of this study, and of Grace Lui, MSc, who helped with the early phase of this study.

Authors' Contributions

PH led the design of the study. KY and EK collected and analyzed the data. All authors contributed to the interpretation of the results and reviewed and edited the manuscript. All authors read and approved the final version of the manuscript.

Conflicts of Interest

Two Contec pulse oximeters were supplied for the purposes of the study by HeartBeat Technologies Ltd. None of the authors have any association with, or have received any personal benefit from, HeartBeat Technologies Ltd.

Multimedia Appendix 1
Information given during recruitment.
[PDF File (Adobe PDF File), 406 KB - jmir_v23i4e19187_app1.pdf]

References

1. Lindsay P, Connor Gorber S, Joffres M, Birtwhistle R, McKay D, Cloutier L, Canadian Task Force on Preventive Health Care (CTFPHC). Recommendations on screening for high blood pressure in Canadian adults. Can Fam Physician 2013 Sep;59(9):927-33, e393 [FREE Full text] [Medline: 24029505]
2. Kaczorowski J, Chambers LW, Dolovich L, Paterson JM, Karwalajtys T, Gierman T, et al. Improving cardiovascular health at population level: 39 community cluster randomised trial of Cardiovascular Health Awareness Program (CHAP). BMJ 2011 Feb 07;342:d442 [FREE Full text] [doi: 10.1136/bmj.d442] [Medline: 21300712]
3. Nerenberg KA, Zarnke KB, Leung AA, Dasgupta K, Butalia S, McBrien K, Hypertension Canada. Hypertension Canada's 2018 guidelines for diagnosis, risk assessment, prevention, and treatment of hypertension in adults and children. Can J Cardiol 2018 May;34(5):506-525. [doi: 10.1016/j.cjca.2018.02.022] [Medline: 29731013]
4. McAlister FA, Straus SE. Evidence based treatment of hypertension. Measurement of blood pressure: an evidence based review. BMJ 2001 Apr 14;322(7291):908-911 [FREE Full text] [doi: 10.1136/bmj.322.7291.908] [Medline: 11302909]
5. Mercury versus nonmercury sphygmomanometers: a review of the comparative clinical effectiveness and guidelines. Canadian Agency for Drugs and Technologies in Health. 2012. URL: https://www.cadth.ca/mercury-versus-non-mercury-sphygmomanometers-review-comparative-clinical-effectiveness-and [accessed 2018-06-25]
6. Mena LJ, Félix VG, Ostos R, González AJ, Martínez-Peláez R, Melgarejo JD, et al. Mobile personal health care system for noninvasive, pervasive, and continuous blood pressure monitoring: development and usability study. JMIR Mhealth Uhealth 2020 Jul 20;8(7):e18012 [FREE Full text] [doi: 10.2196/18012] [Medline: 32459642]
7. Plante TB, Urrea B, MacFarlane ZT, Blumenthal RS, Miller ER, Appel LJ, et al. Validation of the instant blood pressure smartphone app. JAMA Intern Med 2016 May 01;176(5):700-702 [FREE Full text] [doi: 10.1001/jamainternmed.2016.0157] [Medline: 26938174]
8. Plante TB, O'Kelly AC, Urrea B, Macfarlane ZT, Appel LJ, Miller III ER, et al. Auralife instant blood pressure app in measuring resting heart rate: validation study. JMIR Biomed Eng 2018 Nov 21;3(1):e11057. [doi: 10.2196/11057]

9. Mazoteras-Pardo V, Becerro-De-Bengoa-Vallejo R, Losa-Iglesias ME, López-López D, Palomo-López P, Rodríguez-Sanz D, et al. Validation in the general population of the iHealth track blood pressure monitor for self-measurement according to the European Society of Hypertension International Protocol Revision 2010: descriptive investigation. JMIR Mhealth Uhealth 2019 Mar 19;7(3):e13137 [FREE Full text] [doi: 10.2196/13137] [Medline: 30888331]
10. O'Brien E, Atkins N, Stergiou G, Karpettas N, Parati G, Asmar R, Working Group on Blood Pressure Monitoring of the European Society of Hypertension. European Society of Hypertension International Protocol revision 2010 for the validation of blood pressure measuring devices in adults. Blood Press Monit 2010 Feb;15(1):23-38. [doi: 10.1097/MBP.0b013e3283360e98] [Medline: 20110786]
11. Stergiou G, Alpert B, Mieke S, Asmar R, Atkins N, Eckert S, et al. A universal standard for the validation of blood pressure measuring devices: Association for the Advancement of Medical Instrumentation/European Society of Hypertension/International Organization for Standardization (AAMI/ESH/ISO) collaboration statement. J Hypertens 2018 Mar;36(3):472-478 [FREE Full text] [doi: 10.1097/HJH.0000000000001634] [Medline: 29384983]
12. CMS50EW Pulse Oximeter. Contec Medical Systems Co Ltd. 2020. URL: https://www.contecmed.com/productinfo/602619.html [accessed 2021-04-15]
13. Licence no.: 90629. Health Canada. URL: https://health-products.canada.ca/mdall-limh/information.do?companyId_idCompanie=130273&lang=eng [accessed 2021-03-22]
14. K141362. U.S. Food and Drug Administration. URL: https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpmn/pmn.cfm?id=K141362 [accessed 2021-03-22]
15. K962655. U.S. Food and Drug Administration. URL: https://www.accessdata.fda.gov/SCRIPTS/cdrh/devicesatfda/index.cfm?db=pmn&id=K962655 [accessed 2021-03-22]
16. Schulz KF, Altman DG, Moher D, CONSORT Group. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. PLoS Med 2010 Mar 24;7(3):e1000251 [FREE Full text] [doi: 10.1371/journal.pmed.1000251] [Medline: 20352064]
17. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986 Mar 08;1(8476):307-310. [Medline: 2868172]
18. Bland JM, Altman DG. Measuring agreement in method comparison studies. Stat Methods Med Res 1999 Jun;8(2):135-160. [doi: 10.1177/096228029900800204] [Medline: 10501650]

Abbreviations

ESH-IP2: European Society of Hypertension International Protocol revision 2010
LOA: limits of agreement
PPG: photoplethysmogram
SpO_2: peripheral capillary oxygen saturation

Edited by G Eysenbach; submitted 07.04.20; peer-reviewed by C Calvo-Lobo, K Brunt, T McGough; comments to author 12.06.20; revised version received 23.10.20; accepted 14.03.21; published 23.04.21.

Please cite as:
Holyoke P, Yogaratnam K, Kalles E
Web-Based Smartphone Algorithm for Calculating Blood Pressure From Photoplethysmography Remotely in a General Adult Population: Validation Study
J Med Internet Res 2021;23(4):e19187
URL: https://www.jmir.org/2021/4/e19187
doi:10.2196/19187
PMID:33890856

Original Paper

A Technology Acceptance Model for Deploying Masks to Combat the COVID-19 Pandemic in Taiwan (My Health Bank): Web-Based Cross-sectional Survey Study

Wen-Hsun Tsai[1,2], MS; Yi-Syuan Wu[3], MS; Chia-Shiang Cheng[3], MS; Ming-Hao Kuo[1], MS; Yu-Tien Chang[4], PhD; Fu-Kang Hu[1,5], MS; Chien-An Sun[6*], PhD; Chi-Wen Chang[7,8*], PhD; Ta-Chien Chan[9*], PhD; Chao-Wen Chen[10*], MD; Chia-Cheng Lee[5,11*], MD; Chi-Ming Chu[1,3,4,6,12,13,14,15*], PhD

[1]Graduate Institute of Medical Sciences, National Defense Medical Center, Taipei, Taiwan
[2]Medical Administration Office, Beitou Branch, Tri-Service General Hospital, Taipei, Taiwan
[3]Graduate Institute of Life Sciences, National Defense Medical Center, Taipei, Taiwan
[4]School of Public Health, National Defense Medical Center, Taipei, Taiwan
[5]Medical Informatics Office, Tri-Service General Hospital, Taipei, Taiwan
[6]Big Data Research Center, College of Medicine, Fu-Jen Catholic University, New Taipei City, Taiwan
[7]School of Nursing, College of Medicine, Chang-Gung University, Taoyuan, Taiwan
[8]Division of Pediatric Endocrinology & Genetics, Department of Pediatrics, Chang-Gung Memorial Hospital, Taoyuan, Taiwan
[9]Research Center for Humanities and Social Sciences, Academia Sinica, Taipei, Taiwan
[10]Trauma and Critical Care Service Department of Surgery, Kaohsiung Medical University Hospital, Kaohsiung, Taiwan
[11]Division of Colorectal Surgery, Department of Surgery, Tri-Service General Hospital, Taipei, Taiwan
[12]Department of Public Health, China Medical University, Taichung, Taiwan
[13]Department of Public Health, College of Health Sciences, Kaohsiung Medical University, Kaohsiung, Taiwan
[14]Department of Healthcare Administration and Medical Informatics, College of Health Sciences, Kaohsiung Medical University, Kaohsiung, Taiwan
[15]Department of Medical Research, Kaohsiung Medical University Hospital, Kaohsiung, Taiwan
[*]these authors contributed equally

Corresponding Author:
Chi-Ming Chu, PhD
School of Public Health
National Defense Medical Center
4F, No161, Min-Chun E Rd, Sec 6, Neihu Dist
Taipei
Taiwan
Phone: 886 2 8792 3100 ext 18473
Email: chuchiming@web.de

Abstract

Background: The successful completion of medical practices often relies on information collection and analysis. Government agencies and medical institutions have encouraged people to use medical information technology (MIT) to manage their conditions and promote personal health. In 2014, Taiwan established the first electronic personal health record (PHR) platform, My Health Bank (MHB), which allows people to access and manage their PHRs at any time. In the face of the COVID-19 pandemic in 2020, Taiwan has used MIT to effectively prevent the spread of COVID-19 and undertaken various prevention measures before the onset of the outbreak. Using MHB to purchase masks in an efficient and orderly way and thoroughly implementing personal protection efforts is highly important to contain disease spread.

Objective: This study aims to understand people's intention to use the electronic PHR platform MHB and to investigate the factors affecting their intention to use this platform.

Methods: From March 31 to April 9, 2014, in a promotion via email and Facebook, participants were asked to fill out a structured questionnaire after watching an introductory video about MHB on YouTube. The questionnaire included seven dimensions: perceived usefulness, perceived ease of use, health literacy, privacy and security, computer self-efficacy, attitude toward use, and behavioral intention to use. Each question was measured on a 5-point Likert scale ranging from "strongly disagree" (1 point) to

"strongly agree" (5 points). Descriptive statistics and structural equation analysis were performed using SPSS 21 and AMOS 21 software.

Results: A total of 350 valid questionnaire responses were collected (female: 219/350, 62.6%; age: 21-30 years: 238/350, 68.0%; university-level education: 228/350, 65.1%; occupation as student: 195/350, 56.6%; average monthly income <NT $30,000 [<US $1054.89]: 230/350, 65.7%; residence in northern Taiwan: 236/350, 67.4%; and health status perceived as "good": 171/350, 48.9%). Five indicators, including chi-square test (X^2_{310}=2.63), goodness-of-fit index (0.85), adjusted goodness-of-fit index (0.81), comparative fit index (0.91), and root mean square error of approximation (0.07), were calculated. The results indicated a good fit. Further analysis indicated that the most important factor affecting respondents' behavioral intention to use MHB was their attitude toward use (0.78), followed by perceived ease of use (0.65), perceived usefulness (0.41), health literacy (0.10), and privacy and security (0.07).

Conclusions: From the perspective of the populace, this study explored the factors affecting the use of MHB and constructed an interpretation model with a strong goodness of fit. The results of our analysis are consistent with the technology acceptance model. Through the diverse value-added services of MHB, Taiwan's experience in pandemic prevention with smart technology can facilitate future responses to unknown, emerging infectious diseases.

(J Med Internet Res 2021;23(4):e27069) doi:10.2196/27069

KEYWORDS

personal health record; electronic medical record; my health bank; technology acceptance model; structural equation model; electronic health record; COVID-19; protection; survey; model; intention; usage; literacy; privacy; security

Introduction

Background

With the rapid development of various information and communication technologies (ICTs), medical institutions now use a variety of digital medical information tools to provide the information needed by patients and medical care providers and allow physicians and patients more time for communication and discussion to make correct medical decisions. Information tools are now regarded as the mainstream way to provide clinical care [1-3]. The digitization of personal health records (PHRs) is a particularly important tool for realizing the goal of patient-centered care [3-9]. PHRs can be collected in different forms for various health management purposes [10]. Compared with electronic medical records (EMRs), PHRs contain other health-related information in addition to medical records, such as social status, family history, and living environment [7,11-15]. Therefore, PHRs were built on the basis of EMRs. As the use of EMRs increases, the use of PHRs will also increase [11].

EMRs are mainly stored in hospital databases and can be accessed by different types of medical personnel but are not easily exchanged between hospitals. In 2004, Taiwan began to develop the basic format of EMRs to promote the digitization of health information. In 2007, Executive Yuan initiated the National Health Informatics Project to promote the development of a health care information infrastructure and create a development environment for health information [16]. In 2008, the Clinical Document Architecture was defined according to the Health Level 7 standard, and the localized EMR format—the Taiwan Electronic Medical Record Template—was created using 108 individual EMR templates, which provided an important foundation for the interhospital exchange of EMRs [7,17,18]. The Taiwanese governments have encouraged medical institutions to use EMRs through policies, regulations, and financial incentives for health care insurance.

Most studies on EMRs have been from the perspective of medical personnel. In a patient-centered paradigm, medical records would belong to the individual patient. Through EMRs, patients can have more control over their medical information. Zarcadoolas et al [5] used a focus group interview to explore the use of patient portals, which are health care–related web applications. Their results showed that most respondents were interested in accessing their own medical records and believed that such access was important in improving health literacy and promoting their own health and that of their family members. In a national study in the Netherlands, representatives of medical centers jointly discussed future medical information policies, among which "patient participation and empowerment" was given the highest score [19]. Therefore, for governments, the patients' acceptance of and intention to use electronic PHRs could influence subsequent policy formulation and must be considered. Owing to the advantages of EMRs, such as reducing medical costs, promoting care quality, and enhancing medical efficacy, many countries have begun to plan and develop EMRs with different formats and applications [20-24]. However, in most cases, patients cannot access and manage their PHR at any time. In view of this, the Ministry of Health and Welfare of Taiwan developed a web-based health information enquiry system, My Health Bank (MHB), in 2014 to allow people to access their personal health information from a computer or smartphone at any time or place [7,24-26].

The first edition of MHB could be used to check outpatient records, emergency records, inpatient records, dental records, traditional Chinese medicine records, disease diagnosis, drug usage, medical expenses, pathological reports, x-ray examinations, allergy histories, and vaccinations, but the interactivity and advanced query capabilities were limited. It also had no images. In other words, it only provided data, not information or knowledge [7,25,26]. Patient portals that present content in simple and easy-to-understand text or images are more in line with people's internet usage habits; in addition, a

user-friendly interface would help improve the utilization rate [5,26-28].

MHB was thus revised in July 2016 with the addition of functions for visualization of medical information, disease management service, and self-health management, as shown in Figures 1-3 [26]. These new functions greatly improved the functionality of MHB and provided diversified health management services. In response to the need for functions specific to COVID-19 prevention, new functions such as preordering masks, maps of masks, and assistance with getting masks were added to MHB in March 2020, as shown in Figure 4 [29]. When the COVID-19 pandemic became severe, people in Taiwan could purchase masks through the MHB app, which not only helped them effectively implement personal protective measures but also helped stabilize public sentiments.

Figure 1. Medical data visualization (screenshot from My Health Bank): People can query information about outpatients (emergency department and hospital) and medical diagnoses over the past 3 years. Statistical results are presented graphically.

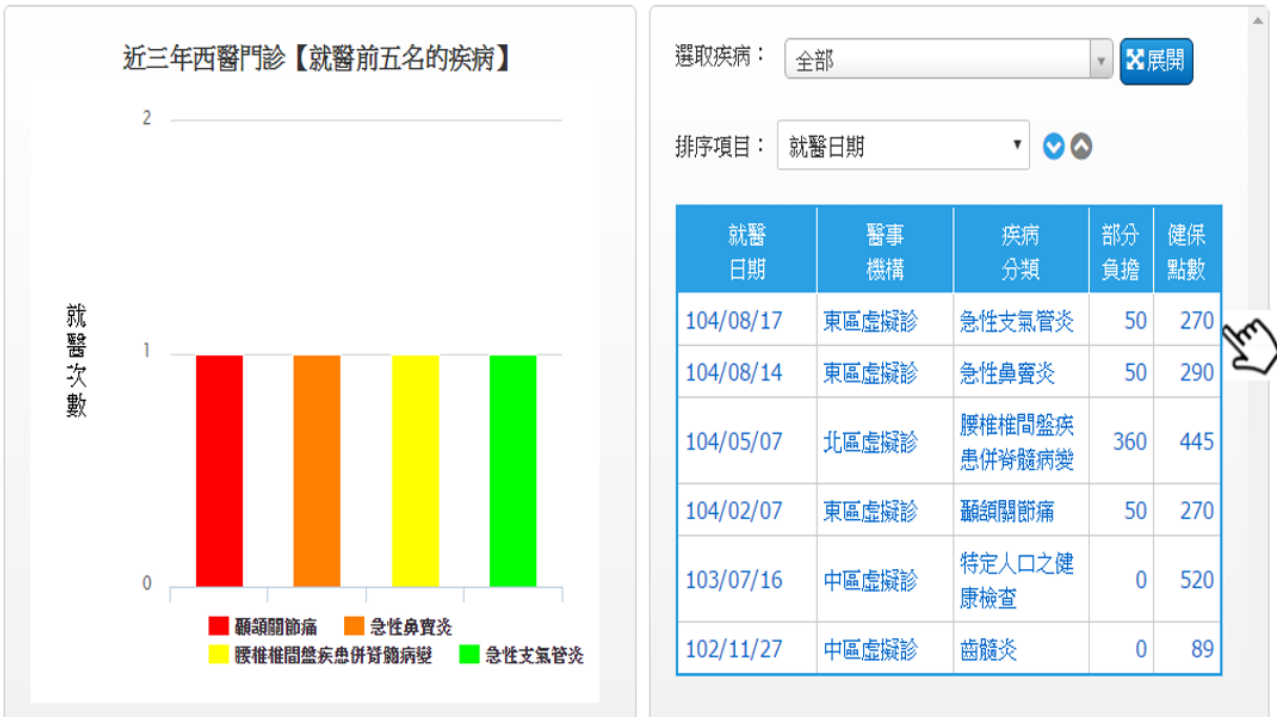
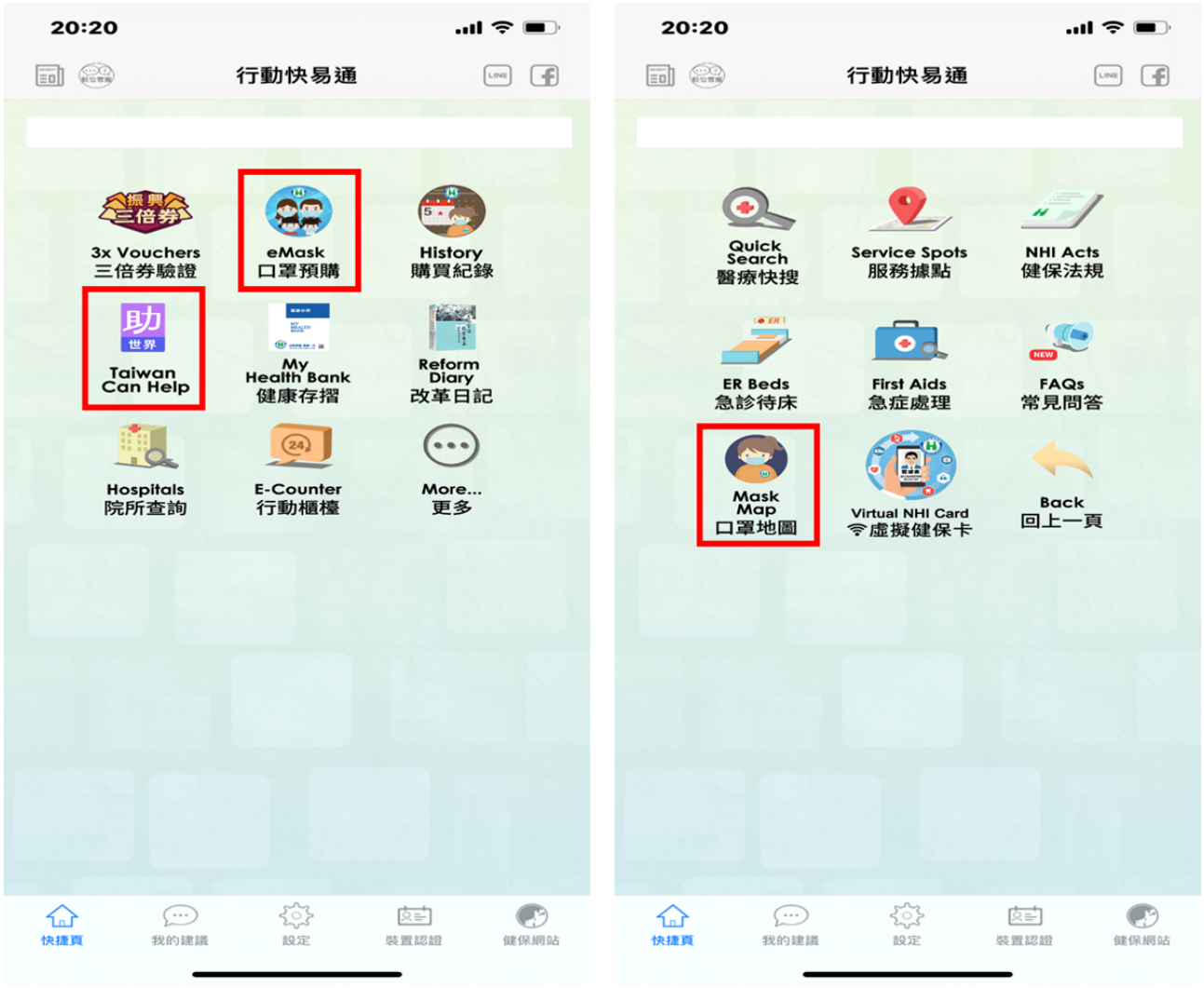


Figure 2. Disease management services (screenshot from My Health Bank): After the user enters their basic information, the system predict the risk of major diseases (eg, hepatic cancer and end-stage renal disease) and then provide links to external websites for further health information queries.

Figure 3. Self-health management (screenshot from My Health Bank): Physiological measurement data, such as height, weight, and blood pressure, are entered and historical trends are monitored. In addition, the system proactively reminds the user about dental cleaning, continuous prescription collection, and health check-ups.

Figure 4. Preordering masks (screenshot from My Health Bank mobile app): Users can log in with their ID or mobile phone number to buy masks, check the stock of masks in nearby drugstores, or even donate their mask quotas to other countries.



Theoretical Foundation

With the development of informatization of the medical industry, the acceptance and use of relevant information technology by clinical medical staff and patients have gradually received attention. The research model used by most researchers is the technology acceptance model (TAM). Although it was not initially developed for use in the medical industry [30,31], this model has become an important theoretical model for studying medical information usage behavior with the expansion of research on the medical industry in recent years [32-40]. In the original TAM, Davis et al [30,31] used three variables: "perceived usefulness," "perceived ease of use," and "attitude toward use" to explain and predict the behavioral intention of users. Because health behaviors are too complex to be explained by a single theory, many researchers use TAM as a foundation in combination with other theories or references to construct theoretical models with better explanatory capabilities in the form of added variables [31,33,36,37,40-46].

According to a literature review by Rahimi et al [45], the original TAM has been extended to suit the dynamic health service environment. TAM has been used to explore three categories of application areas of ICT in the medical service industry: "telemedicine," "electronic health records," and "mobile applications." Researchers have added different variables to the original TAM model according to different ICT application fields, thereby enhancing the explanatory power of the extended model. Health literacy is a predictive factor that has been widely studied in research on health-related behaviors and is also a key factor for the appropriate selection and use of health information [42,47]. Relevant studies have also found that the application of health literacy in TAM is related to perceived usefulness, perceived ease of use, and behavioral intention. A high level of health literacy can increase people's willingness to adopt new health information technology [12,42,48,49].

The health system is a complex social system composed of stakeholders with different backgrounds, experiences, and values [36]. ICT applications allow the system to run smoothly. The storage, retrieval, transmission, and sharing of medical information are closely related to the operation of computer software and hardware [31]. Compeau and Higgins defined "computer self-efficacy" as an individual's ability to use information technology, which plays an important role in shaping personal perception and usage behavior. Computer self-efficacy also affects an individual's perceived ease of use.

Individuals with high computer self-efficacy use computers more often and have less computer anxiety [44,46,50-52]. However, in the process of using medical information technology (MIT), new issues of security and privacy may arise [41]. To solve these security and privacy issues, many countries have not only enacted laws and regulations to protect their citizens' health data [53] but also studied and explored the impact of security and privacy on the use of mobile health (mHealth) systems [41,45,54,55].

This study explores the intention of the Taiwanese people to use MHB and the factors influencing this intention, thus providing a reference for governments to consult when promoting electronic PHRs in the future.

Methods

Study Design

This was a cross-sectional study. To investigate participants' intention to use MHB and the influencing factors, participants were asked to watch a brief introductory video about MHB on YouTube and then fill out a structured questionnaire on Google Forms. The content of the video was how to use MHB, as shown in Figure 5. After a patient goes to hospital A for medical treatment, he or she can log into MHB with their account number, password, or natural person certificate to access five items: basic personal information, hospital visit records, examination records, PHRs, and personal insurance status. This information can be shared with other medical personnel, family members, and insurance companies.

Based on the above study objectives and a literature review, we added three extended variables to the original technology acceptance model: "Health literacy," "Privacy and security," and "Computer self-efficacy." The research framework and hypotheses are illustrated in Figure 6 and described below:

- Hypothesis 1: Perceived ease of use has a positive effect on perceived usefulness.
- Hypothesis 2: Perceived usefulness has a positive effect on attitude toward use.
- Hypothesis 3: Perceived ease of use has a positive effect on attitude toward use.
- Hypothesis 4: Perceived usefulness has a positive effect on behavioral intention to use.
- Hypothesis 5: Health literacy has a positive effect on behavioral intention to use.
- Hypothesis 6: Privacy and security have a positive effect on behavioral intention to use.
- Hypothesis 7: Computer self-efficacy has a positive effect on behavioral intention to use.
- Hypothesis 8: Attitude toward use has a positive effect on behavioral intention to use.

Figure 5. Schematic of My Health Bank.

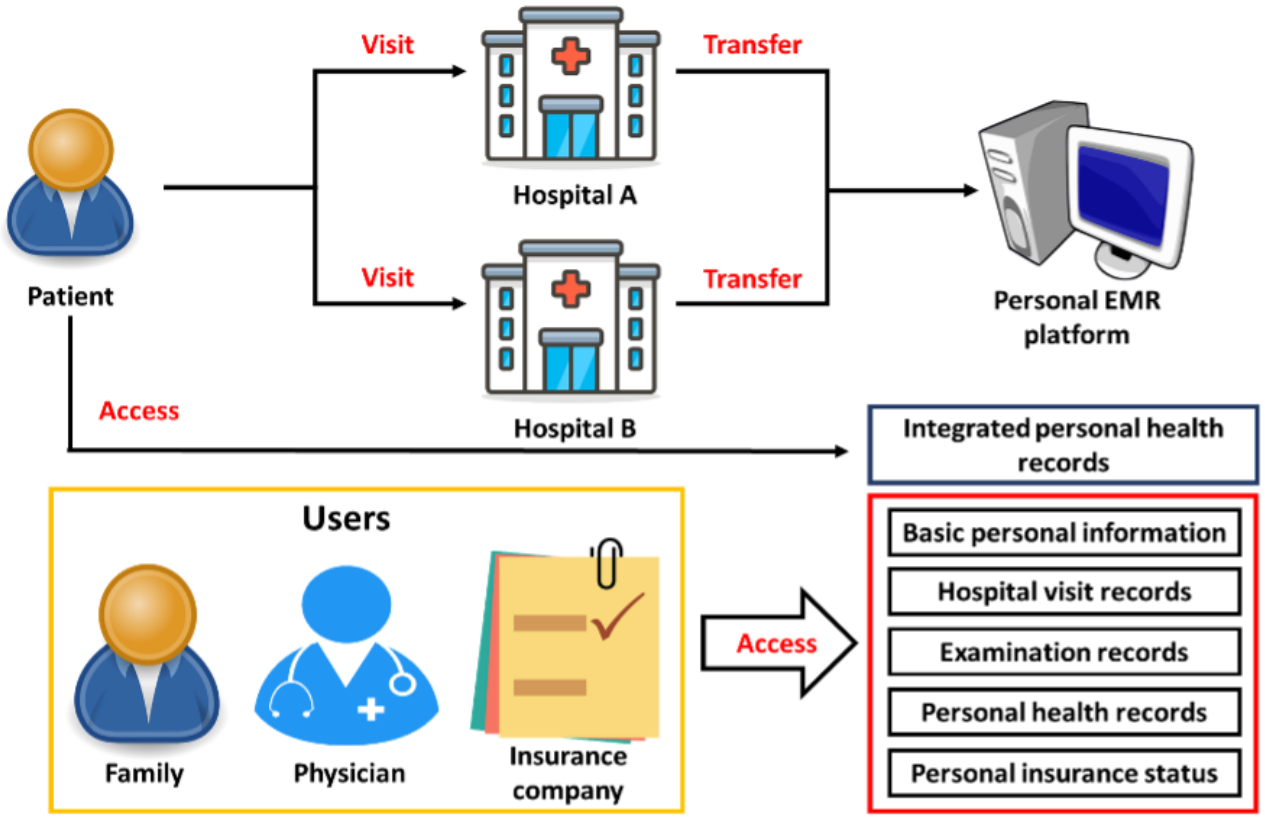
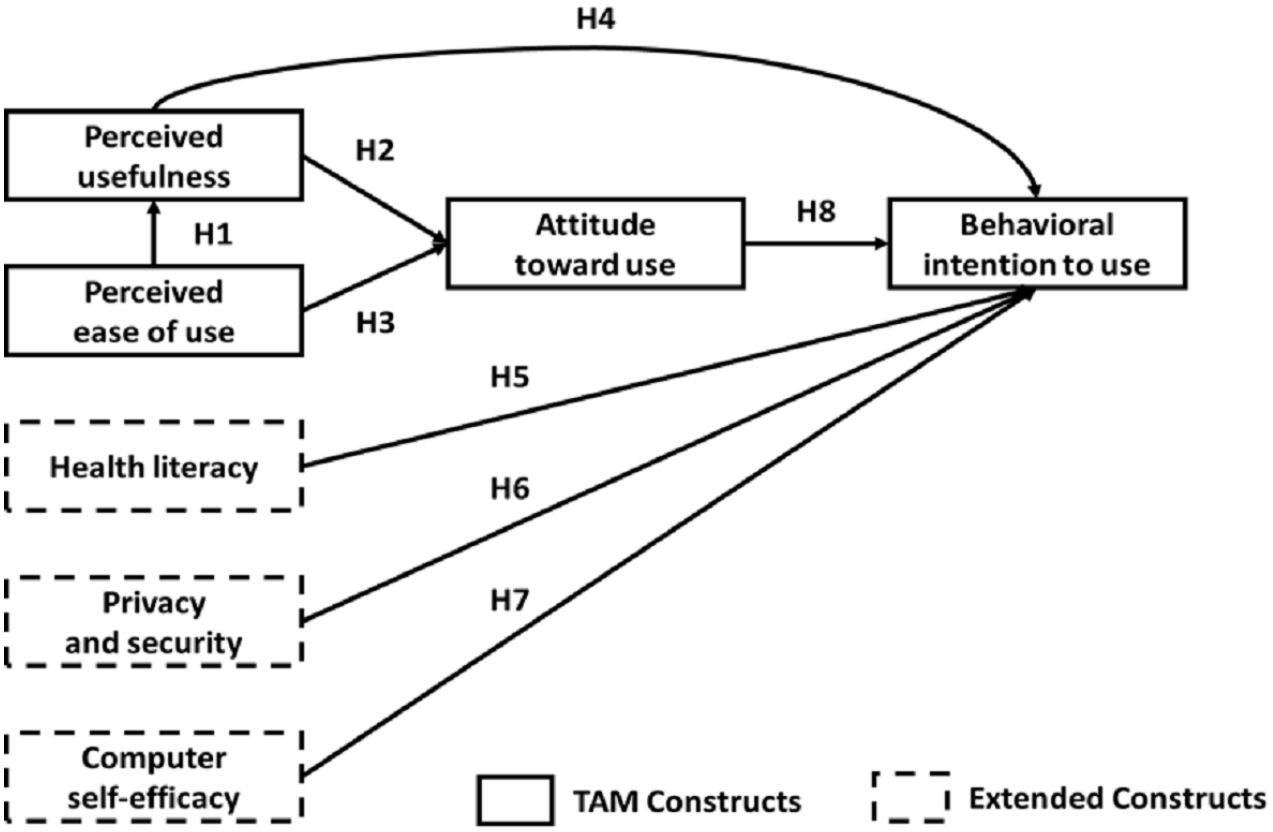


Figure 6. Research framework and hypotheses (extension of the technology acceptance model). EMR: electronic medical record.



Research Tools

Table 1 shows the questionnaire comprising 33 questions that covered seven dimensions: perceived usefulness, perceived ease of use, health literacy, privacy and security, computer self-efficacy, attitude toward use, and behavioral intention to use. Each question was answered on a 5-point Likert scale ("Strongly agree," 5 points; "Agree," 4 points; "Neutral," 3 points; "Disagree," 2 points; and "Strongly disagree," 1 point).

Before the formal survey, four experts in relevant fields were invited to assess the importance, applicability, and clarity of each question in the questionnaire to confirm that the questionnaire had good external validity. In addition, a pretest was conducted with 30 people to confirm that respondents could understand the questions and answer them clearly.

Table 1. Questionnaire design.

Dimension	Definition	Questions
Perceived usefulness	The usefulness of MHB[a] in managing the PHR[b]	5
Perceived ease of use	The ease of use of MHB in managing the PHR	5
Health literacy	When using MHB, the required ability to process and understand this information service in order to make appropriate medical and health decisions	3
Privacy and security	The awareness of privacy and security issues when using MHB	5
Computer self-efficacy	The basic computer skills required when using MHB	5
Attitude toward use	Public evaluation of MHB	4
Behavioral intention to use	The intention to use MHB	6

[a]MHB: My Health Bank.

[b]PHR: personal health record.

Sampling and Exclusion Criteria

From March 31 to April 9, 2014, in a promotion via email and Facebook, internet users were randomly selected to watch the introductory video about MHB and then complete the web-based questionnaire. The size of the valid sample was determined according to the sampling guidelines proposed by Magnani et al [56,57]. The calculated minimum number of valid samples needed for this study was 323. During the collection process, 614 people visited the website, 355 of whom completed the

questionnaire (response rate 57.8%). After excluding duplicate and invalid questionnaire responses, 350 questionnaire responses were collected, which exceeded the minimum number of valid samples needed for this study.

Data Analysis

The data in this study were sorted in Microsoft Excel and then analyzed with SPSS 21 (IBM Corp). The significance level for statistical analysis was set at 5%, that is, $P<.05$. Descriptive statistics, including the mean, SD, median, frequency distribution, and percentage, were used to observe the data distribution of the responses.

Inferential statistics were used to understand the correlations between the characteristics of the respondents and perceived usefulness, perceived ease of use, health literacy, privacy and security, computer self-efficacy, attitude toward use, and behavioral intention to use. AMOS 21 (IBM Corp) was used to test the validity of the dimensions and construct the research model.

Results

Respondent Characteristics

A total of 350 valid questionnaires were collected in this study, and the distribution of the respondents' characteristics is shown in Table 2. Most respondents were female (219/350, 62.6%). The age grouping interval was 10 years, and the majority of the respondents (238/350, 68%) were in the age group of 21-30 years. The majority of the respondents had a university education (228/350, 65.1%), were students (195/350, 55.7%), had an average monthly income lower than NT $30,000 (US $1054.89; 230/350, 65.7%), resided in northern Taiwan (236/350, 67.4%), and had a self-perception of *good* health (171/350, 48.9%).

Table 2. Characteristics of the respondents (N=350).

Characteristic	Value, n (%)
Sex	
Male	131 (37.4)
Female	219 (62.6)
Age (years)	
<20	46 (13.1)
21-30	238 (68)
31-40	56 (16.)
>40	10 (2.9)
Highest education level	
High school degree	8 (2.3)
University degree	228 (65.1)
Graduate school or above	114 (32.6)
Employment	
Student	195 (55.7)
Services	61 (17.4)
Manufacturing	29 (8.3)
Financial industry	7 (2)
Military and police education	47 (13.4)
Unemployed	11(3.1)
Monthly income in NT $ (US $)	
<30,000 (<1054.89)	230 (65.7)
30,001-50,000 (1054.92-1758.15)	94 (26.9)
50,001-70,000 (1758.18-2461.41)	21 (6)
>70, 001 (>2461.44)	5 (1.5)
Living area	
Northern Taiwan	236 (67.4)
Central Taiwan	49 (14)
Southern Taiwan	55 (15.7)
Eastern Taiwan	7 (2)
Offshore islands	3 (0.9)
Health status	
Excellent	28 (8)
Good	171 (48.9)
Normal	132 (37.7)
Poor	19 (5.4)

Measurement Model

This study evaluated the measurement model with internal reliability, convergent validity, and discriminant validity [58]. Internal reliability was evaluated by Cronbach alpha and composite reliability. A value higher than .70 can be regarded as an acceptable level of internal consistency [59]. If the “average variance extracted” (AVE) for a dimension is higher than 0.50, the model is deemed to have reached acceptable convergent validity. Table 3 shows the composite reliability, Cronbach α, and AVE obtained in this study.

Discriminant validity was evaluated by calculating the square root of AVE and cross-loading matrix. When the square root of the AVE is greater than the corresponding correlation, the discriminant validity of the data is confirmed. The calculated values are shown in Table 4.

Table 3. Cronbach alpha, composite reliability, and average variance extracted.

Dimension	Cronbach α	Composite reliability	Average variance extracted
PU[a]	.85	0.82	0.53
PE[b]	.83	0.81	0.52
HL[c]	.82	0.70	0.54
PS[d]	.95	0.90	0.75
CE[e]	.78	0.80	0.50
AU[f]	.86	0.84	0.57
BI[g]	.92	0.93	0.69

[a]PU: perceived usefulness.

[b]PE: perceived ease of use.

[c]HL: health literacy.

[d]PS: privacy and security.

[e]CE: computer self-efficacy.

[f]AU: attitude toward use.

[g]BI: behavioral intention to use.

Table 4. Correlation matrix and square root of the average variance extracted. Note: The elements on the diagonal represent the square root of the average variance extracted, and off-diagonal elements represent the correlations between the constructs.

Dimension	PU[a]	PE[b]	HL[c]	PS[d]	CE[e]	AU[f]	BI[g]
PU	0.73	0.72	0.24	0.10	0.59	0.71	0.62
PE	0.72	0.72	0.45	0.24	0.69	0.71	0.72
HL	0.24	0.45	0.74	0.36	0.48	0.26	0.33
PS	0.10	0.24	0.36	0.87	0.33	0.30	0.19
CE	0.59	0.69	0.48	0.33	0.71	0.70	0.65
AU	0.71	0.71	0.26	0.30	0.70	0.75	0.74
BI	0.62	0.72	0.33	0.19	0.65	0.74	0.83

[a]PU: perceived usefulness.

[b]PE: perceived ease of use.

[c]HL: health literacy.

[d]PS: privacy and security.

[e]CE: computer self-efficacy.

[f]AU: attitude toward use.

[g]BI: behavioral intention to use.

Structural Model

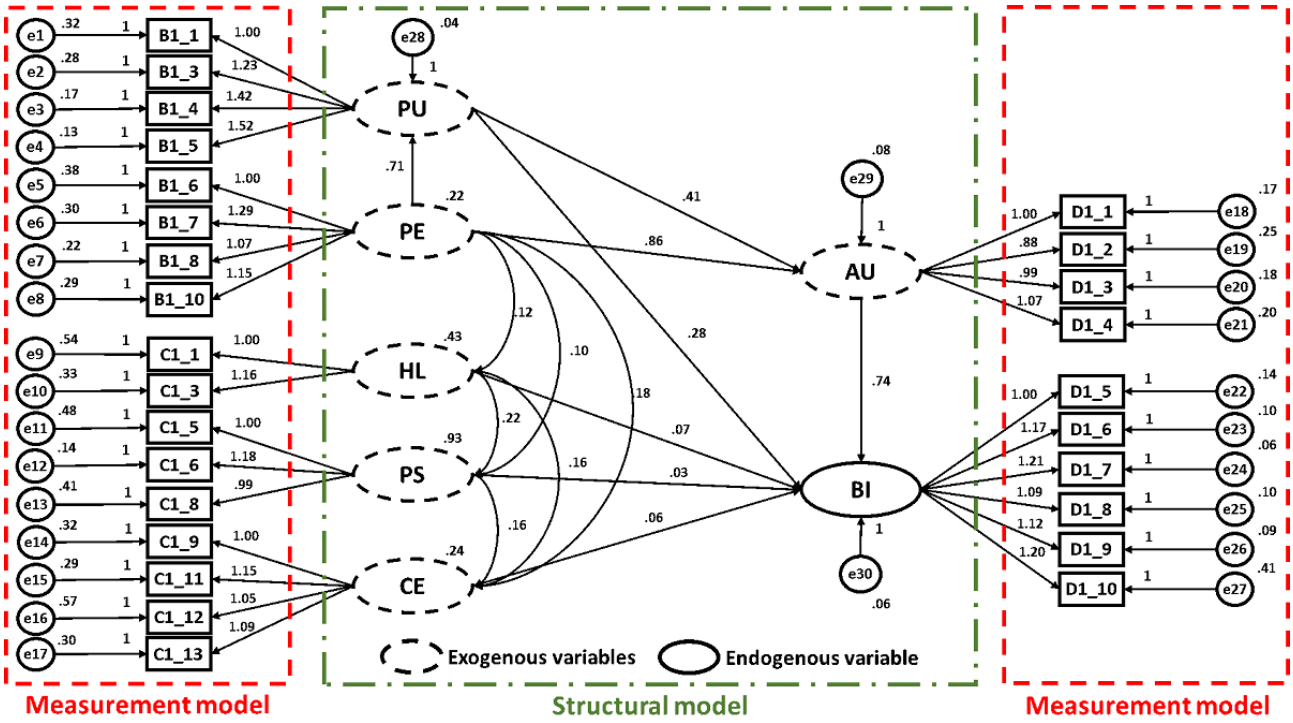
Structural equation modeling is a multivariate statistical technique combining factor analysis and path analysis [60]. In this study, we used structural equation modeling to investigate the relevant factors affecting participants' use of MHB and the importance of each factor [34]. Five indicators were used to assess the overall fit of the model: X^2 (*df*), goodness-of-fit index (GFI), adjusted goodness-of-fit index (AGFI), comparative fit index (CFI) and root mean square error of approximation (RMSEA). As shown in Table 5, X^2_{310}=2.63, which is less than the cutoff of 5 [61]; GFI=0.85 and AGFI=0.81, both of which are higher than the standard value of 0.8 [62,63]; CFI=0.91, which is close to 1 [64]; and RMSEA=0.07 [65]. The above results indicate that the overall fit and fitness of the model are good.

Table 5. Results of overall model fit.

Index	Criteria	Result	Mode fit
Chi-square (*df*)	<5	2.63 (310)	Fit
Goodness-of-fit index	>0.8	0.85	Fit
Adjusted goodness-of-fit index	>0.8	0.81	Fit
Comparative fit index	Close to 1	0.91	Fit
Root mean square error of approximation	• Acceptable: 0.08-0.10 • Good: 0.05-0.08 • Excellent: <0.05	0.07	Good

Structural equation modeling was used to evaluate whether the causal path relationships between the hypotheses proposed in this study were valid. The path diagram is shown in Figure 7, and the standardized estimation results are shown in Table 6. Perceived ease of use explained 74.7% of the variance in perceived usefulness. Perceived usefulness and perceived ease of use explained 67.4% of the variance in attitude toward use, and the other dimensions explained 72.6% of the variance in behavioral intention to use. The above results are the direct effects between the dimensions. To further understand the effects between dimensions other than direct effects, indirect effects were included in the analyses to calculate the total effects. Table 7 shows the standardized direct effects, indirect effects, and total effects of the dimensions.

Figure 7. Path diagram of the research framework. Measured items are illustrated in rectangles (eg, B1_1); latent variables are illustrated in ovals (eg, PU); smaller circles illustrate the error of measurement (eg, e1); associations are illustrated by arrows that indicate the direction of prediction. Factor loadings are noted next to the items. Coefficients are noted for each association (ie, directional arrow). AU: attitude toward use; BI: behavioral intention to use; CE: computer self-efficacy; HL: health literacy; PE: perceived ease of use; PS: privacy and security; PU: perceived usefulness.



In terms of behavioral intention to use, the direct effect of attitude toward use (0.78) was relatively large, whereas perceived ease of use (0.65) and perceived usefulness (0.13) had significant indirect effects on behavioral intention to use. Perceived ease of use had no direct effect on behavioral intention to use, and its indirect effect was much stronger than the direct effects of the other dimensions. Overall, the main factor affecting behavioral intention to use was the attitude toward use, followed by perceived ease of use, perceived usefulness, health literacy, and privacy and security. The test results of all hypotheses were statistically significant except for Hypothesis 5, Hypothesis 6, and Hypothesis 7.

Table 6. Results of path analysis.

Hypothesis	Path coefficient	*t* test (*df*)	*P* value	R^2
H1: Perceived ease of use has a positive effect on perceived usefulness.	0.71	8.24 (310)	<.001	0.747
H2: Perceived usefulness has a positive effect on attitude toward use.	0.41	3.12 (310)	<.001	0.674
H3: Perceived ease of use has a positive effect on attitude toward use.	0.86	5.01 (310)	<.001	__[a]
H4: Perceived usefulness has a positive effect on behavioral intention to use.	0.28	3.49 (310)	<.001	0.726
H5: Health literacy has a positive effect on behavioral intention to use.	0.07	1.83 (310)	.067	—
H6: Privacy and security has a positive effect on behavioral intention to use.	0.03	1.58 (310)	.115	—
H7: Computer self-efficacy has a positive effect on behavioral intention to use.	0.06	0.73 (310)	.468	—
H8: Attitude toward use has a positive effect on behavioral intention to use.	0.74	8.18 (310)	<.001	—

[a]Not available.

Table 7. Direct effects, indirect effects, and total effects.

Dimension and effect	PU[a]	PE[b]	HL[c]	PS[d]	CE[e]	AU[f]	BI[g]
PU							
Direct effect	__[h]	0.86	—	—	—	—	—
Indirect effect	—	—	—	—	—	—	—
Total effect	—	0.86	—	—	—	—	—
AU							
Direct effect	0.28	0.81	—	—	—	—	—
Indirect effect	—	0.02	—	—	—	—	—
Total effect	0.28	0.82	—	—	—	—	—
BI							
Direct effect	0.28	—	0.10	0.07	0.06	0.78	—
Indirect effect	0.13	0.65	—	—	—	—	—
Total effect	0.41	0.65	0.10	0.07	0.06	0.78	—

[a]PU: perceived usefulness.
[b]PE: perceived ease of use.
[c]HL: health literacy.
[d]PS: privacy and security.
[e]CE: computer self-efficacy.
[f]AU: attitude toward use.
[g]BI: behavioral intention to use.
[h]Not available.

Discussion

Principal Findings

Taiwan has had no local cases of COVID-19 since April 12, 2020, and there has been no second wave of the pandemic in Taiwan, which is mainly due to the effective use of MIT to build a strong pandemic prevention network. For example, an integrated circuit card is used as an insurance certificate to give full play to the functions of smart medical cards. Physicians can log into the health care medical information cloud query system through a health insurance card to quickly obtain the medical information, travel history, and contact history of patients without revealing personal information. During the COVID-19 pandemic, assisting frontline medical personnel in assessing disease risk and taking corresponding infection control measures has been greatly helpful for prevention of disease spread [66]. In addition, in accordance with the policy of real-name mask purchasing, people who have completed identity authentication or mobile phone authentication can use MHB to prepurchase masks and know how many masks are still available to them from the mask purchase map, so they can purchase personal protective equipment more efficiently. The aforementioned innovative information technologies, together with the National Health Insurance and the Central Epidemic Command Center, which conducts daily live broadcasts to explain the pandemic situation and future pandemic prevention policies, has calmed

the unease and anxiety of the people and made pandemic control in Taiwan effective [67].

During the promotion of an innovative medical information service, users might experience the three stages of awareness, want, and adoption. Liang [2] found that there is a digital divide between these three stages. This digital divide differs significantly not only between demographic groups but also depending on personal computer ownership and internet use habits. Therefore, the acceptance of an innovative medical information service by users could play a key role in user attitude and experience [7].

This study found that perceived ease of use had a significant positive effect on perceived usefulness, indicating that the respondents believed that if MHB was easy to use, its effects could be significant. Perceived usefulness and perceived ease of use had significant positive impacts on the attitude toward use, indicating that the respondents believed that if MHB had practical benefits in managing personalized EMRs and was easy to use, they would have a positive attitude toward the use of MHB. Perceived usefulness had a significant positive impact on behavioral intention to use, which means that the respondents believed that if MHB helped them understand their own health conditions, their intention to use MHB would increase. Attitude toward use had a positive effect on behavioral intention to use, indicating that the better the functions of respondent's health records are managed, the higher their intention to use MHB.

Financial and resource constraints on the health care system are increasing due to the aging population and changing disease patterns [68]. To improve the service efficiency of medical institutions and the participation of patients, medical institutions are encouraging patients to make appropriate medical care decisions and undertake health promotion activities through electronic PHRs to control their own health status and achieve the goal of patient-centered care [9,27,69]. Therefore, increasing patients' intention to use electronic PHRs is particularly important. The TAM proposed by Davis et al [30] indicated that perceived ease of use affects perceived usefulness and, thus, intention to use. The results of this study showed that respondents strongly agreed (5="strongly agree," 1="strongly disagree) with the usefulness (mean score 4.42, SD 0.53) and ease of use of MHB (mean score 4.28, SD 0.57), and these two technical or cognitive factors had a significant positive effect on the respondents' attitude toward use and behavioral intention to use concerning MHB. Other studies have shown that both the general public and clinicians are more willing to use a medical information tool if the interface of the tool is designed to be easy to use and can fully realize the benefits of care [18,27,70-72]. MHB has become an extension of the medical treatment process and an important management tool to improve health knowledge and promote the health of people and their loved ones [5].

According to a web-based survey on health issues, 70% of people and 65% of physicians believe that patients can download and manage their personalized digital health information [73]. When people use digital health information tools, in addition to a certain ability to use information devices, they must also be able to understand the health information itself. In theory, with advances in information technology, the effectiveness of various health care services may also increase, but people with low computer self-efficacy and health literacy may not be able to make full use of the information technology, which could further deepen the digital divide [1,4,5,19,28,74]. Therefore, to effectively use personalized digital health information, people should have sufficient health literacy and computer self-efficacy. The World Health Organization defines health literacy as "the achievement of a level of knowledge, personal skills and confidence to take action to improve personal and community health by changing personal lifestyles and living conditions" [75]. In summary, health literacy refers to how people acquire, understand, use, and communicate health-related information [42,49]. Computer self-efficacy is self-judgment and self-confidence in the ability to use a computer and includes both computer operation skills (such as formatting hard disks) and the ability to combine these skills to perform different tasks [52]. In this study, health literacy and computer self-efficacy did not have a significant positive impact on behavioral intention to use MHB, indicating that the respondents' intention to use MHB would not be affected by their own health literacy and computer self-efficacy. Due to advances in information technology and higher education levels, the health literacy and computer self-efficacy of young people are higher than those of older people [5,76-78]. The respondents in this study were mostly young internet users who likely had confidence in understanding, acquiring, and applying relevant health information. Therefore, their levels of health literacy and computer self-efficacy did not affect their intention to use MHB. Nevertheless, the impact of health literacy and computer self-efficacy on the use of medical information tools cannot be ignored.

The promotion and use of various MITs are aimed at improving the quality of care and patient safety, enabling correct medical decision-making, reducing medical costs, improving the accessibility of medical services, and promoting service efficiency [1,2,7-9,18,27,69,74,79-81]. In the context of using these MITs, privacy and security are relatively important issues. Studies have shown that regardless of country or national conditions, people's views on the privacy and security of health information systems are consistent [23,24,73]. Since the discussion of medical privacy is a relatively abstract concept, this study evaluated privacy issues in PHRs from three aspects: information privacy, psychological privacy, and social privacy [82]. Privacy and security had no influence on the behavioral intention to use MHB. Due to the increasing popularity of social media (eg, Facebook, blogs, and Twitter), people care more about the speed of information transmission than about how to protect their privacy [83]; in addition, respondents who have not actually used MHB may have different views on privacy and security. Although our results on privacy and security did not reach statistical significance, the scores of the privacy and security dimension showed that most respondents chose the option "disagree" (mean score 2.45, SD 1.07), indicating that the respondents were still concerned about data privacy and security issues. In the future establishment and promotion of digital PHRs, privacy and security should be given continuous attention [4,6,23,84,85].

In addition, the results of this study show that attitude toward use was the main factor affecting behavioral intention to use (total effect: 0.78), and the Likert scores of the respondents on attitude toward use (mean 4.22, SD 0.56) and behavioral intention to use (mean 4.20, SD 0.56) were positive. The results of our path analysis showed that attitude toward use had a positive effect on behavioral intention to use. Therefore, our results confirm that when the attitude toward use is more positive, the intention to use is higher. To enhance the public's views on the ease of use and usefulness of MHB and to strengthen the positive assessment of MHB, the Ministry of Health and Welfare should continue to design diversified value-added services for MHB.

Since March 2020, with the policy of real-name mask purchases, MHB has combined the functions of mask preorders, mask purchase maps, and mask donation. The Taiwanese people can conveniently purchase masks during the COVID-19 pandemic to achieve effective COVID-19 prevention [29,86]. Combined with its excellent public health facilities, Taiwan took advantage of its advanced deployment of MIT to prevent and control the COVID-19 pandemic at the early stage. In contrast to the severe ongoing global COVID-19 pandemic, Taiwan has seen a respite from the impact of the disease [67,87]. In 2017, approximately 590,000 MHB accounts were registered and used, accounting for approximately 2.7% of the total Taiwanese population [24]. After adding a variety of diverse services in MHB, according to Apple, there were more than 7.3 million downloads of MHB in 2020, and the number of users significantly increased by 31.7% [88], which shows that as long as people are satisfied with the real benefits and value brought by an information tool, their intention to use it will increase, and the actual use of the tool will increase accordingly.

Limitations

The MHB discussed in this study is a newly promoted health information service in Taiwan, and relatively few people have used it before. Therefore, people might have a limited understanding of the service, content, and functions of MHB. To reduce possible errors in our study, before filling out the questionnaire, the respondents were asked to watch an introductory video about MHB to ensure that they had a certain level of knowledge about the service.

In this study, a structured web-based questionnaire was used to survey the people, so the results may not cover the entire parent population, and the sample representativeness may be limited. Most of the respondents were young female internet users and lived in northern Taiwan. Therefore, future studies should collect more samples to increase the representativeness of the results. In addition, respondents may be affected by the surrounding environment, emotions, and other uncertain factors, resulting in measurement errors, which can be corrected using appropriate statistical methods.

In this study, data were collected by means of a cross-sectional survey, a method that is fast, easy, and inexpensive to perform and the results of which are helpful for further complex research [89]. A study that used longitudinal methods to assess the 14-year use of clinical information systems found that the acceptance of clinical information systems increased over time. In addition, the factors affecting acceptance varied with time [38]. Accordingly, with the development of the life cycle of ICT applications, the relationship between variables in TAM may also change [45]. Therefore, future studies may also consider regularly collecting data to investigate the effect of time factors on the use of MHB. Although it is difficult to extrapolate the results, this pioneering study could still be used as a reference for the development of electronic PHR platforms in Taiwan.

Conclusions

This study used the TAM and a structured web-based questionnaire to investigate the Taiwanese people’s intention to use MHB. Validation by structural equation modeling was performed to obtain an interpretation model with good fitness, and the results showed that perceived ease of use has a significant positive impact on perceived usefulness; perceived ease of use and perceived usefulness both have a significant positive influence on attitude toward use; perceived usefulness has a significant positive impact on behavioral intention to use; and attitude toward use has a significant positive effect on behavioral intention to use. Health literacy, computer self-efficacy, and privacy and security have nonsignificant effects on behavioral intention to use. Further exploration of the effectiveness of each dimension indicates that attitude toward use is the most important dimension affecting respondents’ use of MHB. Even though health literacy, computer self-efficacy, and privacy and security have no significant impact on intention to use, they are still considered important influencing factors in relevant studies. In the future, different research designs can be used for further exploration.

The purpose of establishing MHB is to return the management of personal health information to patients so that they can check their past medical experience and health records at any time to strengthen their self-health care ability. In addition, MHB also allows patients to have autonomy over their personal digital health records, which truly realizes the idea of empowering the people. The planning of a nation's medical policy should be adjusted with the development of the latest ICTs and must be able to accommodate all kinds of complex scientific and humanity issues. Policy implementation not only enables disease management from the perspective of individuals but also promotes health value from the perspective of groups. Taiwan's successful experience in preventing the COVID-19 pandemic by using various MITs is conducive to the sustainable development of more diversified value-added services for MHB in the future and the creation of different medical service modes. The model developed in this study could not only be applied to the adoption of other similar PHR platforms but also be used as a reference for other countries to formulate medical information policies to provide management insights, thereby increasing the use of patients' PHRs.

Acknowledgments

The authors would like to thank Tang Feng, Minister without Portfolio of Taiwan, for using My Health Bank to strengthen the prevention of the COVID-19 epidemic and inspire this study.

Conflicts of Interest

None declared.

References

1. Lyles CR, Fruchterman J, Youdelman M, Schillinger D. Legal, practical, and ethical considerations for making online patient portals accessible for all. Am J Public Health 2017 Oct;107(10):1608-1611. [doi: 10.2105/AJPH.2017.303933] [Medline: 28817324]
2. Liang TH. Matrix analysis of the digital divide in eHealth services using awareness, want, and adoption gap. J Med Internet Res 2012 Feb 13;14(1):e11 [FREE Full text] [doi: 10.2196/jmir.1670] [Medline: 22329958]
3. Wolcott V, Agarwal R, Nelson DA. Is provider secure messaging associated with patient messaging behavior? Evidence from the US army. J Med Internet Res 2017 Apr 06;19(4):e103 [FREE Full text] [doi: 10.2196/jmir.6804] [Medline: 28385681]
4. Greenberg AJ, Falisi AL, Finney Rutten LJ, Chou WS, Patel V, Moser RP, et al. Access to electronic personal health records among patients with multiple chronic conditions: a secondary data analysis. J Med Internet Res 2017 Jun 02;19(6):e188 [FREE Full text] [doi: 10.2196/jmir.7417] [Medline: 28576755]
5. Zarcadoolas C, Vaughon WL, Czaja SJ, Levy J, Rockoff ML. Consumers' perceptions of patient-accessible electronic medical records. J Med Internet Res 2013 Aug 26;15(8):e168 [FREE Full text] [doi: 10.2196/jmir.2507] [Medline: 23978618]
6. Kim J, Bates DW. Analysis of the definition and utility of personal health records using Q methodology. J Med Internet Res 2011 Nov 29;13(4):e105 [FREE Full text] [doi: 10.2196/jmir.1781] [Medline: 22126860]
7. Rau HH, Wu YS, Chu CM, Wang FC, Hsu MH, Chang CW, et al. Importance-performance analysis of personal health records in Taiwan: a web-based survey. J Med Internet Res 2017 Apr 27;19(4):e131 [FREE Full text] [doi: 10.2196/jmir.7065] [Medline: 28450273]
8. Lämsä E, Timonen J, Mäntyselkä P, Ahonen R. Pharmacy customers' experiences with the national online service for viewing electronic prescriptions in Finland. Int J Med Inform 2017 Jan;97:221-228. [doi: 10.1016/j.ijmedinf.2016.10.014] [Medline: 27919380]
9. Reti SR, Feldman HJ, Ross SE, Safran C. Improving personal health records for patient-centered care. J Am Med Inform Assoc 2010;17(2):192-195 [FREE Full text] [doi: 10.1136/jamia.2009.000927] [Medline: 20190063]
10. Bonander J, Gates S. Public health in an era of personal health records: opportunities for innovation and new partnerships. J Med Internet Res 2010 Aug 10;12(3):e33 [FREE Full text] [doi: 10.2196/jmir.1346] [Medline: 20699216]
11. Archer N, Fevrier-Thomas U, Lokker C, McKibbon KA, Straus SE. Personal health records: a scoping review. J Am Med Inform Assoc 2011;18(4):515-522 [FREE Full text] [doi: 10.1136/amiajnl-2011-000105] [Medline: 21672914]
12. Park Y, Yoon HJ. Understanding personal health record and facilitating its market. Healthc Inform Res 2020 Jul;26(3):248-250 [FREE Full text] [doi: 10.4258/hir.2020.26.3.248] [Medline: 32819044]
13. Godbole NS, Lamb J. The triple challenge for the healthcare industry: Sustainability, privacy, and cloud-centric regulatory compliance. : IEEE; 2013 Presented at: 2013 10th International Conference and Expo on Emerging Technologies for a Smarter World (CEWIT); October, 21-22, 2013; Melville, NY p. 21-22. [doi: 10.1109/cewit.2013.6713768]
14. Gaynor M, Yu F, Andrus CH, Bradner S, Rawn J. A general framework for interoperability with applications to healthcare. Health Policy and Technology 2014 Mar;3(1):3-12. [doi: 10.1016/j.hlpt.2013.09.004]
15. Giokas D. EMR, EHR and PHR – and Now aEMR and H/HIS – What's with These Systems? Canada Health Infoway. 2016 May 4. URL: https://tinyurl.com/2tp7pudz [accessed 2020-10-25]
16. Ministry of Health and Welfare. Introduction of National Health Informatics Project. Electronic Medical Record Promotion Area. URL: https://emr.mohw.gov.tw/emr/introduction.aspx [accessed 2020-07-15]
17. Jian WS, Hsu CY, Hao TH, Wen HC, Hsu MH, Lee YL, et al. Building a portable data and information interoperability infrastructure-framework for a standard Taiwan Electronic Medical Record Template. Comput Methods Programs Biomed 2007 Nov;88(2):102-111. [doi: 10.1016/j.cmpb.2007.07.014] [Medline: 17936402]
18. Iqbal U, Ho CH, Li YC, Nguyen PA, Jian WS, Wen HC. The relationship between usage intention and adoption of electronic health records at primary care clinics. Comput Methods Programs Biomed 2013 Dec;112(3):731-737. [doi: 10.1016/j.cmpb.2013.09.001] [Medline: 24091088]
19. Rauwerdink A, Kasteleyn MJ, Haafkens JA, Chavannes NH, Schijven MP, steering committee of the Citrien fund program eHealth. A national eHealth vision developed by University Medical Centres: A concept mapping study. Int J Med Inform 2020 Jan;133:104032. [doi: 10.1016/j.ijmedinf.2019.104032] [Medline: 31778886]
20. Patient Access to Health Records. HealthIT.gov. URL: https://www.healthit.gov/topic/patient-access-health-records/patient-access-health-records [accessed 2020-10-25]

21. What Is Digital Health. Canada Health Infoway. URL: https://www.infoway-inforoute.ca/en/what-we-do/benefits-of-digital-health/what-is-digital-health [accessed 2020-10-25]
22. What is Digital Health. Australian Digital Health Agency. URL: https://www.digitalhealth.gov.au/what-is-digital-health [accessed 2020-10-25]
23. Nøhr C, Parv L, Kink P, Cummings E, Almond H, Nørgaard JR, et al. Nationwide citizen access to their health data: analysing and comparing experiences in Denmark, Estonia and Australia. BMC Health Serv Res 2017 Aug 07;17(1):534 [FREE Full text] [doi: 10.1186/s12913-017-2482-y] [Medline: 28784173]
24. Yeh MJ, Saltman RB. Creating online personal medical accounts: Recent experience in two developed countries. Health Policy and Technology 2019 Jun;8(2):171-178. [doi: 10.1016/j.hlpt.2019.05.004]
25. Huang SK, Pan YT, Chen MS. My Health Bank 2.0-making a patron saint for people's health. J Formos Med Assoc 2017 Feb;116(2):69-71 [FREE Full text] [doi: 10.1016/j.jfma.2016.05.008] [Medline: 27421173]
26. My Health Bank User's Guidance. Webpage in Chinese. National Health Insurance. URL: https://myhealthbank.nhi.gov.tw/IHKE0002/ihke0002s18.aspx#/#m-info [accessed 2020-11-29]
27. Ancker JS, Hafeez B, Kaushal R. Socioeconomic disparities in adoption of personal health records over time. Am J Manag Care 2016 Aug;22(8):539-540 [FREE Full text] [Medline: 27541700]
28. Coughlin SS, Stewart JL, Young L, Heboyan V, De Leo G. Health literacy and patient web portals. Int J Med Inform 2018 May;113:43-48. [doi: 10.1016/j.ijmedinf.2018.02.009] [Medline: 29602432]
29. Mask Real-Name System 2.0 Officially Launched, My Health Bank Fully Supports Epidemic Prevention. eBook in Chinese. Ministry of Health and Welfare, Taiwan. Taipei, Taiwan; 2020 Mar. URL: http://www.mohwpaper.tw/adv3/maz24/ebook/mobile/index.html [accessed 2021-04-16]
30. Davis FD, Bagozzi RP, Warshaw PR. User acceptance of computer technology: a comparison of two theoretical models. Management Science 1989 Aug;35(8):982-1003. [doi: 10.1287/mnsc.35.8.982]
31. Holden RJ, Karsh BT. The technology acceptance model: its past and its future in health care. J Biomed Inform 2010 Feb;43(1):159-172 [FREE Full text] [doi: 10.1016/j.jbi.2009.07.002] [Medline: 19615467]
32. Tavares J, Oliveira T. New integrated model approach to understand the factors that drive electronic health record portal adoption: cross-sectional national survey. J Med Internet Res 2018 Nov 19;20(11):e11032 [FREE Full text] [doi: 10.2196/11032] [Medline: 30455169]
33. Ketikidis P, Dimitrovski T, Lazuras L, Bath PA. Acceptance of health information technology in health professionals: an application of the revised technology acceptance model. Health Informatics J 2012 Jun;18(2):124-134 [FREE Full text] [doi: 10.1177/1460458211435425] [Medline: 22733680]
34. Wilson EV, Lankton NK. Modeling patients' acceptance of provider-delivered e-health. J Am Med Inform Assoc 2004;11(4):241-248 [FREE Full text] [doi: 10.1197/jamia.M1475] [Medline: 15064290]
35. Tavares J, Oliveira T. Electronic health record patient portal adoption by health care consumers: an acceptance model and survey. J Med Internet Res 2016 Mar 02;18(3):e49 [FREE Full text] [doi: 10.2196/jmir.5069] [Medline: 26935646]
36. Hoque MR, Bao Y, Sorwar G. Investigating factors influencing the adoption of e-Health in developing countries: A patient's perspective. Inform Health Soc Care 2017 Jan;42(1):1-17. [doi: 10.3109/17538157.2015.1075541] [Medline: 26865037]
37. Kim J, Park HA. Development of a health information technology acceptance model using consumers' health behavior intention. J Med Internet Res 2012 Oct 01;14(5):e133 [FREE Full text] [doi: 10.2196/jmir.2143] [Medline: 23026508]
38. Hadji B, Martin G, Dupuis I, Campoy E, Degoulet P. 14 Years longitudinal evaluation of clinical information systems acceptance: The HEGP case. Int J Med Inform 2016 Feb;86:20-29. [doi: 10.1016/j.ijmedinf.2015.11.016] [Medline: 26725691]
39. Kim S, Lee KH, Hwang H, Yoo S. Analysis of the factors influencing healthcare professionals' adoption of mobile electronic medical record (EMR) using the unified theory of acceptance and use of technology (UTAUT) in a tertiary hospital. BMC Med Inform Decis Mak 2016 Jan 30;16:12 [FREE Full text] [doi: 10.1186/s12911-016-0249-8] [Medline: 26831123]
40. Ye T, Xue J, He M, Gu J, Lin H, Xu B, et al. Psychosocial factors affecting artificial intelligence adoption in health care in China: cross-sectional study. J Med Internet Res 2019 Oct 17;21(10):e14316 [FREE Full text] [doi: 10.2196/14316] [Medline: 31625950]
41. Alam MZ, Hu W, Kaium MA, Hoque MR, Alam MMD. Understanding the determinants of mHealth apps adoption in Bangladesh: A SEM-Neural network approach. Technology in Society 2020 May;61:101255. [doi: 10.1016/j.techsoc.2020.101255]
42. Hsieh PJ, Lai HM. Exploring people's intentions to use the health passbook in self-management: An extension of the technology acceptance and health behavior theoretical perspectives in health literacy. Technological Forecasting and Social Change 2020 Dec;161:120328. [doi: 10.1016/j.techfore.2020.120328]
43. Zainab B, Awais Bhatti M, Alshagawi M. Factors affecting e-training adoption: an examination of perceived cost, computer self-efficacy and the technology acceptance model. Behaviour & Information Technology 2017 Oct 04;36(12):1261-1273. [doi: 10.1080/0144929x.2017.1380703]
44. Basak E, Gumussoy CA, Calisir F. Examining the factors affecting PDA acceptance among physicians: An extended technology acceptance model. J Healthc Eng 2015;6(3):399-418 [FREE Full text] [doi: 10.1260/2040-2295.6.3.399] [Medline: 26753441]

45. Rahimi B, Nadri H, Lotfnezhad Afshar H, Timpka T. A systematic review of the technology acceptance model in health informatics. Appl Clin Inform 2018 Jul;9(3):604-634 [FREE Full text] [doi: 10.1055/s-0038-1668091] [Medline: 30112741]
46. Ibrahim R, Leng NS, Yusoff RCM, Samy GN, Masrom S, Rizman ZI. E-learning acceptance based on technology acceptance model (TAM). J. Fundam and Appl Sci 2018 Jan 23;9(4S):871. [doi: 10.4314/jfas.v9i4s.50]
47. Mackert M, Champlin SE, Holton A, Muñoz II, Damásio MJ. eHealth and health literacy: A research methodology review. J Comput-Mediat Comm 2014 Apr 12;19(3):516-528. [doi: 10.1111/jcc4.12044]
48. Chisolm DJ, Hardin DS, McCoy KS, Johnson LD, McAlearney AS, Gardner W. Health literacy and willingness to use online health information by teens with asthma and diabetes. Telemed J E Health 2011 Nov;17(9):676-682 [FREE Full text] [doi: 10.1089/tmj.2011.0037] [Medline: 21943161]
49. Mackert M, Mabry-Flynn A, Champlin S, Donovan EE, Pounders K. Health literacy and health information technology adoption: the potential for a new digital divide. J Med Internet Res 2016 Oct 04;18(10):e264 [FREE Full text] [doi: 10.2196/jmir.6349] [Medline: 27702738]
50. Rahman MS, Ko M, Warren J, Carpenter D. Healthcare Technology Self-Efficacy (HTSE) and its influence on individual attitude: An empirical study. Computers in Human Behavior 2016 May;58:12-24. [doi: 10.1016/j.chb.2015.12.016]
51. Zhang X, Han X, Dang Y, Meng F, Guo X, Lin J. User acceptance of mobile health services from users' perspectives: The role of self-efficacy and response-efficacy in technology acceptance. Inform Health Soc Care 2017 Mar;42(2):194-206. [doi: 10.1080/17538157.2016.1200053] [Medline: 27564428]
52. Compeau DR, Higgins CA. Computer self-efficacy: development of a measure and initial test. MIS Quarterly 1995 Jun;19(2):189 [FREE Full text] [doi: 10.2307/249688]
53. Tavares J, Oliveira T. Electronic Health Record Portal Adoption: a cross country analysis. BMC Med Inform Decis Mak 2017 Jul 05;17(1):97 [FREE Full text] [doi: 10.1186/s12911-017-0482-9] [Medline: 28679423]
54. Abd-Alrazaq A, Bewick BM, Farragher T, Gardner P. Factors affecting patients' use of electronic personal health records in England: cross-sectional study. J Med Internet Res 2019 Jul 31;21(7):e12373 [FREE Full text] [doi: 10.2196/12373] [Medline: 31368442]
55. Jian WS, Syed-Abdul S, Sood SP, Lee P, Hsu MH, Ho CH, et al. Factors influencing consumer adoption of USB-based personal health records in Taiwan. BMC Health Serv Res 2012 Aug 27;12:277 [FREE Full text] [doi: 10.1186/1472-6963-12-277] [Medline: 22925029]
56. Magnani R. Sampling Guide. Washington, DC: Food and Nutrition Technical Assistance (FANTA); Dec 1997.
57. Magnani R, Sabin K, Saidel T, Heckathorn D. Review of sampling hard-to-reach and hidden populations for HIV surveillance. AIDS 2005 May;19 Suppl 2:S67-S72. [doi: 10.1097/01.aids.0000172879.20628.e1] [Medline: 15930843]
58. Hair JF, Hult GTM, Ringle C, Sarstedt M. A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM). Thousand Oaks, CA: SAGE Publications; 2016.
59. Hair JF, Anderson RE, Tatham RL. Multivariate Data Analysis with Readings. Hoboken, NJ: Prentice Hall; 1995.
60. Werts CE, Joreskog KG, Linn RL. Identification and estimation in path analysis with unmeasured variables. American Journal of Sociology 1973 May;78(6):1469-1484. [doi: 10.1086/225474]
61. Wheaton B. Assessment of fit in overidentified models with latent variables. Sociological Methods & Research 2016 Jun 30;16(1):118-154. [doi: 10.1177/0049124187016001005]
62. Doll WJ, Xia W, Torkzadeh G. A confirmatory factor analysis of the end-user computing satisfaction instrument. MIS Quarterly 1994 Dec;18(4):453. [doi: 10.2307/249524]
63. MacCallum RC, Hong S. Power analysis in covariance structure modeling using GFI and AGFI. Multivariate Behav Res 1997 Apr 01;32(2):193-210. [doi: 10.1207/s15327906mbr3202_5] [Medline: 26788758]
64. Hair JF, Black WC, Babin BJ, Anderson RE. Multivariate data analysis : a global perspective. Upper Saddle River, NJ: Pearson; 2010.
65. Hu LT, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural Equation Modeling: A Multidisciplinary Journal 1999 Jan;6(1):1-55. [doi: 10.1080/10705519909540118]
66. Taiwan Can Help - National Health Insurance’s Contribution in Combating COVID-19. National Health Insurance Administration. URL: https://www.nhi.gov.tw/English/Content_List.aspx?n=F083F4F3605179D1 [accessed 2020-12-12]
67. Lee PC, Chen SC, Chiu TY, Chen CM, Chi C. What we can learn from Taiwan's response to the COVID-19 epidemic. The BMJ Opinion. 2020 Jul. URL: https://blogs.bmj.com/bmj/2020/07/21/what-we-can-learn-from-taiwans-response-to-the-covid-19-epidemic/#content [accessed 2020-12-04]
68. About Global Aging. Global Aging Institute. URL: https://www.globalaginginstitute.org/about-gai/about-global-aging.html [accessed 2020-11-08]
69. Simborg DW. Consumer empowerment versus consumer populism in healthcare IT. J Am Med Inform Assoc 2010;17(4):370-372 [FREE Full text] [doi: 10.1136/jamia.2010.003392] [Medline: 20595301]
70. Huang YM, Huang YM, Huang SH, Lin YT. A ubiquitous English vocabulary learning system: evidence of active/passive attitudes vs. usefulness/ease-of-use. Computers & Education 2012 Jan;58(1):273-282. [doi: 10.1016/j.compedu.2011.08.008]

71. Fleming ML, Hatfield MD, Wattana MK, Todd KH. Exploratory study of emergency physicians' use of a prescription monitoring program using a framework of technology acceptance. J Pain Palliat Care Pharmacother 2014 Mar;28(1):19-27. [doi: 10.3109/15360288.2013.879242] [Medline: 24520884]
72. Tung FC, Chang SC, Chou CM. An extension of trust and TAM model with IDT in the adoption of the electronic logistics information system in HIS in the medical industry. Int J Med Inform 2008 May;77(5):324-335. [doi: 10.1016/j.ijmedinf.2007.06.006] [Medline: 17644029]
73. Markle Survey on Health in a Networked Life. Markle Foundation. URL: http://www.markle.org/sites/default/files/20101007_bluebutton_summary_1.pdf [accessed 2020-11-20]
74. Nguyen A, Mosadeghi S, Almario CV. Persistent digital divide in access to and use of the Internet as a resource for health information: results from a California population-based study. Int J Med Inform 2017 Jul;103:49-54. [doi: 10.1016/j.ijmedinf.2017.04.008] [Medline: 28551001]
75. Health Literacy. World Health Organization. URL: https://www.who.int/healthpromotion/health-literacy/en/ [accessed 2020-11-15]
76. Berens EM, Vogt D, Messer M, Hurrelmann K, Schaeffer D. Health literacy among different age groups in Germany: results of a cross-sectional survey. BMC Public Health 2016 Nov 09;16(1):1151 [FREE Full text] [doi: 10.1186/s12889-016-3810-6] [Medline: 27829401]
77. Hadjimina E, Furnham A. Influence of age and gender on mental health literacy of anxiety disorders. Psychiatry Res 2017 May;251:8-13. [doi: 10.1016/j.psychres.2017.01.089] [Medline: 28189082]
78. Duong TV, Sørensen K, Pelikan JM, Van den Broucke S, Lin IF, Lin YC, et al. Health-related behaviors moderate the association between age and self-reported health literacy among Taiwanese women. Women Health 2018 Jul;58(6):632-646. [doi: 10.1080/03630242.2017.1333074] [Medline: 28537772]
79. Mohsen MO, Aziz HA. The blue button project: engaging patients in healthcare by a click of a button. Perspect Health Inf Manag 2015;12:1d [FREE Full text] [Medline: 26755898]
80. Do NV, Barnhill R, Heermann-Do KA, Salzman KL, Gimbel RW. The military health system's personal health record pilot with Microsoft HealthVault and Google Health. J Am Med Inform Assoc 2011;18(2):118-124 [FREE Full text] [doi: 10.1136/jamia.2010.004671] [Medline: 21292705]
81. Ho CH, Wene HC, Chu CM, Wu YS, Wang JL. Importance-satisfaction analysis for primary care physicians' perspective on EHRs in Taiwan. Int J Environ Res Public Health 2014 Jun 06;11(6):6037-6051 [FREE Full text] [doi: 10.3390/ijerph110606037] [Medline: 24914640]
82. Parrott R, Burgoon JK, Burgoon M, LePoire BA. Privacy between physicians and patients: more than a matter of confidentiality. Soc Sci Med 1989;29(12):1381-1385. [doi: 10.1016/0277-9536(89)90239-6] [Medline: 2629120]
83. Tao D, LeRouge C, Smith KJ, De Leo G. Defining information quality into health websites: a conceptual framework of health website information quality for educated young adults. JMIR Hum Factors 2017 Oct 06;4(4):e25 [FREE Full text] [doi: 10.2196/humanfactors.6455] [Medline: 28986336]
84. Dexheimer JW, Greiner MV, Beal SJ, Johnson D, Kachelmeyer A, Vaughn LM. Sharing personal health record data elements in protective custody: youth and stakeholder perspectives. J Am Med Inform Assoc 2019 Aug 01;26(8-9):714-721 [FREE Full text] [doi: 10.1093/jamia/ocz067] [Medline: 31365097]
85. Blumenthal D, Tavenner M. The "meaningful use" regulation for electronic health records. N Engl J Med 2010 Aug 05;363(6):501-504. [doi: 10.1056/NEJMp1006114] [Medline: 20647183]
86. Coronavirus Disease (COVID-19): Masks. World Health Organization. URL: https://www.who.int/news-room/q-a-detail/coronavirus-disease-covid-19-masks [accessed 2020-10-08]
87. Chen CM, Jyan HW, Chien SC, Jen HH, Hsu CY, Lee PC, et al. Containing COVID-19 among 627,386 persons in contact with the diamond princess cruise ship passengers who disembarked in Taiwan: big data analytics. J Med Internet Res 2020 May 05;22(5):e19540 [FREE Full text] [doi: 10.2196/19540] [Medline: 32353827]
88. Taiwan's Top Apps Rankings 2020. National Health Insurance Agency. URL: https://www.nhi.gov.tw/News_Content.aspx?n=FC05EB85BD57C709&sms=587F1A3D9A03E2AD&s=3588080DED1457A2 [accessed 2020-12-10]
89. Sedgwick P. Cross sectional studies: advantages and disadvantages. BMJ 2014 Mar 26;348:g2276-g2276. [doi: 10.1136/bmj.g2276]

Abbreviations

AGFI: adjusted goodness-of-fit index
EMRs: electronic medical record
GFI: goodness-of-fit index
ICTs: information and communication technologies
MHB: My Health Bank
mHealth: mobile health
MIT: medical information technology
PHRs: personal health records

RMSEA: root mean square error of approximation
TAM: technology acceptance model

Edited by C Basch; submitted 18.01.21; peer-reviewed by D Tao, A Chong, E Mastriani; comments to author 08.02.21; revised version received 10.03.21; accepted 01.04.21; published 21.04.21.

Please cite as:
Tsai WH, Wu YS, Cheng CS, Kuo MH, Chang YT, Hu FK, Sun CA, Chang CW, Chan TC, Chen CW, Lee CC, Chu CM
A Technology Acceptance Model for Deploying Masks to Combat the COVID-19 Pandemic in Taiwan (My Health Bank): Web-Based Cross-sectional Survey Study
J Med Internet Res 2021;23(4):e27069
URL: https://www.jmir.org/2021/4/e27069
doi:10.2196/27069
PMID:33819168

Original Paper

Differential Effects of Outpatient Portal User Status on Inpatient Portal Use: Observational Study

Naleef Fareed[1,2], MBA, PhD; Pallavi Jonnalagadda[1], MBBS, DrPH; Sarah R MacEwan[1], PhD; Gennaro Di Tosto[1], PhD; Seth Scarborough[1], MAS; Timothy R Huerta[1,2,3], MS, PhD; Ann Scheck McAlearney[1,2,3], MS, ScD

[1]The Center for the Advancement of Team Science, Analytics, and Systems Thinking, College of Medicine, The Ohio State University, Columbus, OH, United States

[2]Department of Biomedical Informatics, College of Medicine, The Ohio State University, Columbus, OH, United States

[3]Department of Family and Community Medicine, College of Medicine, The Ohio State University, Columbus, OH, United States

Corresponding Author:
Naleef Fareed, MBA, PhD
The Center for the Advancement of Team Science, Analytics, and Systems Thinking
College of Medicine
The Ohio State University
460 Medical Center Drive
Suite 512
Columbus, OH, 43210
United States
Phone: 1 614 366 0283
Email: naleef.fareed@osumc.edu

Abstract

Background: The decision to use patient portals can be influenced by multiple factors, including individuals' perceptions of the tool, which are based on both their personal skills and experiences. Prior experience with one type of portal may make individuals more comfortable with using newer portal technologies. Experienced outpatient portal users in particular may have confidence in their ability to use inpatient portals that have similar functionality. In practice, the use of both outpatient and inpatient portal technologies can provide patients with continuity of access to their health information across care settings, but the influence of one type of portal use on the use of other portals has not been studied.

Objective: This study aims to understand how patients' use of an inpatient portal is influenced by outpatient portal use.

Methods: This study included patients from an academic medical center who were provided access to an inpatient portal during their hospital stays between 2016 and 2018 (N=1571). We analyzed inpatient portal log files to investigate how inpatient portal use varied by using 3 categories of outpatient portal users: prior users, new users, and nonusers.

Results: Compared with prior users (695/1571, 44.24%) of an outpatient portal, new users (214/1571, 13.62%) had higher use of a select set of inpatient portal functions (messaging function: incidence rate ratio [IRR] 1.33, 95% CI 1.06-1.67; function that provides access to the outpatient portal through the inpatient portal: IRR 1.34, 95% CI 1.13-1.58). Nonusers (662/1571, 42.14%), compared with prior users, had lower overall inpatient portal use (all active functions: IRR 0.68, 95% CI 0.60-0.78) and lower use of specific functions, which included the function to review vitals and laboratory results (IRR 0.51, 95% CI 0.36-0.73) and the function to access the outpatient portal (IRR 0.53, 95% CI 0.45-0.62). In comparison with prior users, nonusers also had lower odds of being comprehensive users (defined as using 8 or more unique portal functions; odds ratio [OR] 0.57, 95% CI 0.45-0.73) or composite users (defined as comprehensive users who initiated a 75th or greater percentile of portal sessions) of the inpatient portal (OR 0.42, 95% CI 0.29-0.60).

Conclusions: Patients' use of an inpatient portal during their hospital stay appeared to be influenced by a combination of factors, including prior outpatient portal use. For new users, hospitalization itself, a major event that can motivate behavioral changes, may have influenced portal use. In contrast, nonusers might have lower self-efficacy in their ability to use technology to manage their health, contributing to their lower portal use. Understanding the relationship between the use of outpatient and inpatient portals can help direct targeted implementation strategies that encourage individuals to use these tools to better manage their health across care settings.

(J Med Internet Res 2021;23(4):e23866) doi:10.2196/23866

KEYWORDS

patient portals; technology experience; health information technology; technology acceptance; technology use; log files; mobile phone

Introduction

Background

Health information technologies (HITs) are increasingly being introduced and implemented to provide individuals with tools to help them manage their health. Patient portals are one such HIT tool that provides patients with the opportunity to learn about their health and participate in their health care [1]. These portals offer patients access to personal health information and educational materials, a means to manage their health and health care, and methods to communicate with their health care providers. There is evidence that links the use of patient portals to better patient-provider communication, clinical decision making, and patient satisfaction [2,3]. Furthermore, the use of patient portals has been associated with improved health outcomes such as controlled blood pressure and better glycemic and cholesterol control in patients with chronic diseases [4,5].

Despite the potential benefits of patient portals, there are still barriers to their adoption and use. First, individuals must have access to the internet to use outpatient portals [6]. In addition, individuals must have adequate eHealth literacy—defined as the ability to acquire, comprehend, and apply health information from electronic sources [7,8]—to understand the information that a patient portal can provide [9]. The decision to use patient portals also involves multiple factors related to an individual's perception of the tool in the context of their personal skills and experiences. For example, prior experience with technology may allow individuals to become more comfortable with new technologies compared with individuals who do not have technology experience [10]. Prior technology experience may be particularly valuable in supporting the adoption of technologies that are similar, as this experience may give individuals confidence in their abilities to use other technologies with similar functionalities [11].

Patient portals were first employed in outpatient settings but are now increasingly being implemented in inpatient settings, with functionality tailored to the hospital environment [12]. Inpatient portals have been suggested to have multiple benefits for patients, including the promotion of independence, reduction of anxiety, and increasing empowerment of patients during their hospitalization [13]. Their use has been shown to help uncover medical errors, improve medication adherence, and facilitate patient-provider communication [14]. Furthermore, there is evidence supporting the association between inpatient portal use and lower hospital readmission rates [15], medical errors, and adverse drug events [16]. In practice, the use of both outpatient and inpatient portals together can provide patients with continuity of access to their health information and their health care providers across care settings [17], but the influence of one type of portal use on the use of the other has not been studied.

Objectives

This study seeks to understand the relationship between the use of an outpatient portal and the use of an inpatient portal by examining the portal use of 3 groups of patients: patients who had used an outpatient portal before using an inpatient portal during hospitalization (*prior users*); patients who used an outpatient portal only after using an inpatient portal during a hospital stay (*new users*); and patients who did not use an outpatient portal, despite the fact that they used an inpatient portal during their hospitalization (*nonusers*). We hypothesized that prior users of the outpatient portal would have higher frequency of use and more comprehensive use of the inpatient portal compared with new users or nonusers of the outpatient portal and that these differences would persist after adjusting for demographic and clinical variables.

Methods

Study Setting and Design

A large-scale, pragmatic, randomized controlled trial (RCT) was conducted across 6 noncancer hospitals at a large academic medical center (AMC) between September 2016 and August 2019 [18]. The trial considered the use of 2 patient portals developed by Epic Systems (Epic Systems Corporation). The inpatient portal application, MyChart Bedside (MCB), available on tablet computers provisioned as part of the hospital admission, was introduced in 2013 in select units of the AMC and rolled out system wide in 2016. The outpatient portal, MyChart (MC), was a web-based patient portal that had been in use at the AMC since 2011 and was available as an app on a variety of personal devices, including computers and cell phones. The overall objective of the RCT was to examine how the introduction of MCB influenced patient experience. Briefly, the RCT was structured to examine the impact of 2 interventions, technology and training, using a factorial design. Individuals were provided with a tablet with either full-tech (full MCB) functionality or lite-tech (limited MCB) functionality and received either a high-touch (in-person training from a technology navigator to use the patient portal) or low-touch (instructional video) intervention. In the context of this larger study, we examined MCB use in relation to MC user status among participants in the arm of the RCT that was not impacted by study design considerations—that is, patients who were provisioned with the full-tech tablets and received the low-touch intervention.

MCB and MC Functions

The MCB app available to hospitalized patients included the following functions: Access MyChart (log in to or create an MC account), Dining on Demand (order food from a predefined menu), Happening Soon (view expected interactions with care team during the hospital stay), I Would Like (request ancillary services), Messages (communicate with the care team), My Health (review vitals and laboratory results), Notes (type notes), Taking Care of Me (view active members of the care team), To

Learn (access educational materials), and Tutorial (view a tutorial on how to use MCB). The Access MyChart function could be used by inpatients to create an MC account that was available in the outpatient environment. We refer to all these functions together as active functions in contrast to the administrative functions that were not part of our investigation. The methods for processing MCB log files have been previously described [19].

The MC functions included Messaging (links to messaging center, letters to the patient, and prescription refill option), Visits (list of past and upcoming visits), My Record (list of medications and allergies, medical history and immunizations, test results and health summary, preventive care, and a summary of plan of care), Medical Tools (share medical records with other services, participate in research studies, and connect tracking devices), Resources (terms and conditions, patient education, and frequently asked questions), Proxy (request or renew proxy), and Preferences (personal and security settings and notification preferences). Although MC is tailored for outpatient care, it also includes access to information from inpatient stays within 3 days of discharge. The methods for processing MC log files have also been described previously [20].

Study Data

We used audit log file data from September 2016 to August 2018, a period after MCB rollout that was considered stable with respect to MCB implementation, to test associations between MCB use by MC user status. Our analysis was restricted to a subsample who received access to the full suite of MCB functions and an instructional training video available on the tablet. This subsample provided an adequate sample size to test associations between MCB use by MC user status and avoided confounding effects that might have been related to the in-person technology navigation intervention. Our analytic sample comprised 1571 patients who met the study criteria for this focused analysis.

Our primary outcomes were based on the frequency of MCB function use. We defined MCB function use at 3 levels: the patient level, representing use aggregated over multiple admissions; the admission level, representing use during a single admission; and the sessions level, representing a single, defined instance of coherent MCB use. The definition of MCB use metrics has been previously described in detail [19]. In short, MCB use was defined as the number of times a function was used. We calculated the frequency of use of each of the 10 MCB functions as well as the total frequency of use of all 10 MCB functions.

First, we determined the number of MCB sessions for each patient. Next, we aggregated MCB function use from up to the first 3 admissions for all patients, where the first admission was when patients were enrolled into the RCT (ie, enrollment admission). We selected this cutoff for the number of admissions as most patients had admissions that fell within this range (number of admissions; median 1; 90th percentile=3). We excluded patients whose length of stay (LOS) during their enrollment admission was less than 3 days and those who had no recorded MCB use during their enrollment admission. We also calculated the comprehensiveness of MCB use, defined as the use of 8 or more unique MCB functions [19]. Finally, we defined composite use, a combination of comprehensive use and high-frequency use of MCB. High-frequency MCB use was defined as the total number of MCB sessions equal to or greater than the 75th percentile. We report the results at the patient level to present the aggregate perspective of patient use.

Our main predictor variable was MC user status at the time MCB use first occurred, which we defined as having occurred upon enrollment into the RCT. To define MC user status, we selected 90 days as the cutoff point based on the pattern of MC account activations among patients after their enrollment into the RCT. An MC account is activated when a user signs up to use the app. The number of MC account activations plateaued 3 months after enrollment, which was the basis for our decision that MC account activations occurring within 90 days of trial enrollment was a reasonable time frame during which portal activations could be considered. MC use was defined as any use of the MC functions. Using these definitions, patients were classified into 3 groups: (1) *prior users* (695/1571, 44.24%) were patients with any past recorded MC use before their enrollment admission, (2) *new users* (214/1571, 13.62%) were patients who first used MC during their enrollment admission or within 90 days of their enrollment into the RCT, and (3) *nonusers* (662/1571, 42.14%) were patients who did not use MC before, during, or within 90 days after enrollment. Other covariates included age at enrollment, sex, race, the length of time (days) the tablet with MCB was provisioned (ie, length of provisioning [LOP]) to the patient during their admission, and the Charlson Comorbidity Index (CCI). We did not include the LOS as a covariate because of variability in the length of time between admission and provisioning of the tablet to the patient; given this variability, the LOS does not reflect the duration of time that the patient had to use the inpatient tablet with MCB. Of note, this intersection between the LOS and LOP has been previously described in work pertaining to the processing of log files [19].

Data Analysis

For our analysis, we first examined the distributions of the outcome and predictor variables and assessed differences in the demographic characteristics of MC user groups via analysis of variance, Chi-square tests, and Kruskal-Wallis tests, as applicable. We then modeled the frequency of MCB use by using logistic and negative binomial regression and adjusting for demographic variables, the LOP, and CCIs.

We also performed several sensitivity analyses. First, we repeated the process of calculating MCB function use at the admission and sessions levels from up to 3 admissions. The purpose of examining MCB function use at different levels was to gain a better understanding of patient engagement with the app by considering both granular (sessions) and more aggregate (admission and patient) data. The 1571 patients in our sample had 2227 admissions and 53,823 sessions. Of the 2227 admissions, 1025 (46.02%) were prior users, 310 (13.92%) were new users, and 892 (40.05%) were nonusers. Of the 53,823 sessions, 25,810 (47.95%) were initiated by prior users, 9481 (17.62%) by new users, and 18,532 (34.43%) by nonusers. Our estimates for SEs were clustered by patient. We did not define

comprehensiveness and composite use at the sessions level, as the recording of function use may be influenced by system idiosyncrasies, such as automatically logging the patient out after 10 minutes of inactivity [19-21]. Second, we restricted the analysis to MCB function use to the first admission or the enrollment admission. Finally, we repeated the analyses by redefining MC user status as follows: (1) *prior users* (n=695) were patients with any past recorded MC use before their enrollment admission; (2) *new users* (n=240) were patients who first used MC during their enrollment admission or within 180 days of their enrollment into the trial; and (3) *nonusers* (n=636) were patients who did not use MC before, during, or within 180 days after their enrollment in to the RCT.

Results

Baseline Characteristics of the Participants

Table 1 summarizes the characteristics of the study participants for the analytic sample based on MC user status. The sample mostly consisted of female (899/1571, 57.22%) and White (1180/1571, 75.11%) participants, and participants had a mean age of 47 years (SD 15.09). The median CCI was 1 (range 0-15), and the median LOP was 7 (range 1-124) days. Across the different MC user groups, there were statistically significant differences in gender, race, age, and CCIs. Prior users were typically older, female, and White. New users were likely to have the portal provisioned to them for one more day than the other user groups.

The frequency of MCB use in relation to MC user status is summarized in Table 2. The prior users group had a median of 23 MCB sessions per patient, whereas new users and nonusers had 29 and 16 sessions per patient, respectively. Among prior users, 34.9% (243/695) were comprehensive users and 19.4% (135/695) were composite users. Among new users, 36.9% (79/214) were comprehensive users and 23.4% (50/214) were composite users. For nonusers, 23.1% (153/662) were classified as comprehensive and 9.8% (65/662) were classified as composite users.

Table 1. Baseline characteristics of the study population by MyChart user status.

Characteristics	Overall (N=1571)	Prior users (n=695)	New users (n=214)	Nonusers (n=662)	*P* value[a]
Gender (female), n (%)	899 (57.22)	433 (62.30)	127 (59.35)	339 (51.21)	<.001
Race, n (%)					.004
Black	326 (20.75)	119 (17.12)	42 (19.63)	165 (24.92)	
White	1180 (75.11)	541 (77.84)	162 (75.70)	477 (72.05)	
Other	65 (4.14)	35 (5.04)	10 (4.67)	20 (3.02)	
Age (years), mean (SD)	47 (15.09)	48 (14.95)	44 (14.75)	46 (15.15)	<.001
Charlson Comorbidity Index, median (range)	1 (0-15)	1 (0-15)	1 (0-11)	1 (0-13)	<.001
Length of provisioning (days), median (range)	7 (1-124)	7 (1-124)	8 (1-116)	7 (2-117)	.05

[a]Differences in gender and race were examined using the Chi-square test. Analysis of variance was used to assess differences in mean age. The Kruskal-Wallis test was used to assess differences in the Charlson Comorbidity Index and length of provisioning.

Table 2. Frequency of MyChart Bedside use by MyChart user status[a].

Characteristics	Overall (N=1571)	Prior users (n=695)	New users (n=214)	Nonusers (n=662)
Number of sessions, median (Q1[b], Q3[c])	20 (10, 41)	23 (12, 44)	29 (14, 51)	16 (8, 31)
Active functions, median (Q1, Q3)	93 (37, 198)	103 (44, 237)	136 (58, 259)	69 (26, 151)
Access MyChart, median (Q1, Q3)	2 (0, 4)	2 (1, 5)	4 (1, 7)	1 (0, 2)
Dining on Demand, median (Q1, Q3)	14 (4, 31)	15 (6, 33)	20 (6, 40)	9 (3, 26)
Happening Soon, median (Q1, Q3)	24 (6, 76)	27 (7, 82)	45 (14, 98)	17 (3, 61)
I Would Like, median (Q1, Q3)	0 (0, 0)	0 (0, 0)	0 (0, 0)	0 (0, 0)
Messages, median (Q1, Q3)	2 (0, 5)	2 (1, 6)	3 (1, 7)	1 (0, 3)
My Health, median (Q1, Q3)	1 (0, 16)	3 (0, 22)	5 (0, 27)	0 (0, 4)
Notes, median (Q1, Q3)	0 (0, 0)	0 (0, 0)	0 (0, 0)	0 (0, 0)
Taking Care of Me, median (Q1, Q3)	2 (1, 6)	3 (1, 8)	4 (1, 10)	2 (0, 5)
To Learn, median (Q1, Q3)	0 (0, 1)	0 (0, 1)	0 (0, 2)	0 (0, 1)
Tutorial, median (Q1, Q3)	6 (4, 11)	7 (4, 12)	7 (5, 13)	6 (3, 9)
Comprehensive user[d], n (%)	475 (30.24)	243 (34.96)	79 (36.92)	153 (23.11)
Composite user[e], n (%)	250 (15.91)	135 (19.42)	50 (23.36)	65 (9.81)

[a]MyChart (MC) user status was defined as follows: prior users (n=695) were patients with any past recorded MC use before their enrollment admission; new users (n=214) were patients who first used MC during their enrollment admission or within 90 days of their enrollment into the trial; and nonusers (n=662) were patients who did not use MC either before, during, or after their enrollment admission or those who first used MC 90 days after enrollment into the trial.

[b]Q1: 25th percentile.

[c]Q3: 75th percentile.

[d]A comprehensive user was defined as those who used 8 or more MyChart Bedside functions at both the patient and admission levels.

[e]A composite user was defined as a comprehensive user and high-frequency user of MyChart Bedside. High-frequency users were defined as users who had a total number of MyChart Bedside sessions greater than or equal to the 75th percentile (41 sessions).

Association Between MCB Use and MC User Status

The association between MCB use and MC user status is summarized in Table 3. In the unadjusted analysis, with the exception of the Notes and I Would Like functions, there were significant differences across the 3 MC user groups in the use of the remaining 8 MCB functions. Overall, compared with prior users, we found that new users had higher use of MCB, and nonusers had lower use of MCB. After adjustment, although the frequency of use of active MCB functions was not significantly different between new users and prior users, we found a significantly higher use of certain individual MCB functions among new users. For instance, new users had 33% higher use of the Messages function (incidence rate ratio [IRR] 1.33, 95% CI 1.06-1.67; *P*=.02) and 34% higher use of the Access MyChart function (IRR 1.34, 95% CI 1.13-1.58; *P*=.001), compared with prior users. Furthermore, new users had 12% higher use of the Tutorial function (IRR 1.12, 95% CI 0.99-1.26; *P*=.07), compared with prior users. In contrast, nonusers had a significantly lower frequency of MCB use compared with prior users for all but 2 functions (ie, I Would Like and Notes), even after adjusting for demographic factors, the LOP, and CCIs. The full model estimates are available in Multimedia Appendix 1.

Table 3. MyChart Bedside function use in relation to MyChart user status (in reference to prior users) at the patient level[a].

Functions	New users (unadjusted)	New users (adjusted[b])	Nonusers (unadjusted)	Nonusers (adjusted[b])
Number of sessions				
IRR[c] (95% CI)	1.20 (0.99-1.46)	1.07 (0.93-1.23)	0.75 (0.65-0.86)	0.74 (0.67-0.82)
P value	.07	.35	<.001	<.001
Active functions				
IRR (95% CI)	1.24 (0.98-1.56)	1.11 (0.93-1.32)	0.69 (0.58-0.82)	0.69 (0.60-0.79)
P value	.07	.27	<.001	<.001
Access MyChart				
IRR (95% CI)	1.38 (1.16-1.67)	1.34 (1.13-1.58)	0.54 (0.46-0.64)	0.53 (0.45-0.62)
P value	<.001	.001	<.001	<.001
Dining on Demand				
IRR (95% CI)	1.20 (1.01-1.44)	1.10 (0.96-1.27)	0.81 (0.69-0.95)	0.79 (0.70-0.88)
P value	.04	.16	.01	<.001
Happening Soon				
IRR (95% CI)	1.38 (1.00-1.92)	1.16 (0.90-1.50)	0.71 (0.56-0.91)	0.68 (0.56-0.82)
P value	.05	.26	.01	<.001
I Would Like				
IRR (95% CI)	1.59 (0.89-2.84)	1.47 (0.83-2.61)	0.96 (0.67-1.37)	0.97 (0.68-1.38)
P value	.12	.19	.84	.87
Messages				
IRR (95% CI)	1.45 (1.15-1.83)	1.33 (1.06-1.67)	0.67 (0.55-0.83)	0.65 (0.53-0.78)
P value	.002	.02	<.001	<.001
My Health				
IRR (95% CI)	0.89 (0.55-1.42)	0.93 (0.66-1.32)	0.44 (0.30-0.65)	0.51 (0.36-0.73)
P value	.61	.69	<.001	<.001
Notes				
IRR (95% CI)	0.81 (0.30-2.23)	0.63 (0.24-1.71)	0.58 (0.25-1.32)	0.54 (0.24-1.24)
P value	.69	.36	.19	.15
Taking Care of Me				
IRR (95% CI)	1.22 (0.99-1.51)	1.09 (0.90-1.33)	0.65 (0.54-0.79)	0.64 (0.54-0.75)
P value	.07	.35	<.001	<.001
To Learn				
IRR (95% CI)	1.18 (0.84-1.64)	1.14 (0.81-1.59)	0.74 (0.58-0.95)	0.75 (0.58-0.95)
P value	.34	.45	.02	.02
Tutorial				
IRR (95% CI)	1.11 (0.99-1.25)	1.12 (0.99-1.26)	0.89 (0.81-0.98)	0.92 (0.84-1.00)
P value	.10	.07	.02	.05
Comprehensive user[d]				
OR[e] (95% CI)	1.09 (0.79-1.50)	1.02 (0.73-1.42)	0.56 (0.44-0.71)	0.57 (0.45-0.73)
P value	.60	.90	<.001	<.001
Composite user[f]				
OR (95% CI)	1.26 (0.88-1.83)	1.15 (0.77-1.72)	0.45 (0.33-0.62)	0.42 (0.29-0.60)

Functions	New users (unadjusted)	New users (adjusted[b])	Nonusers (unadjusted)	Nonusers (adjusted[b])
P value	.21	.50	<.001	<.001

[a]MyChart (MC) user status was defined as follows: prior users (n=695) were patients with any past recorded MC use before their enrollment admission; new users (n=214) were patients who first used MC during their enrollment admission or within 90 days of their enrollment into the trial; and nonusers (n=662) were patients who did not use MC either before, during, or after their enrollment admission or those who first used MC 90 days after enrollment into the trial.

[b]Adjusted for age at enrollment, female gender, race, Charlson Comorbidity Index, and the length of provisioning time of the inpatient tablet.

[c]IRR: incidence rate ratio.

[d]A comprehensive user was defined as those who used 8 or more MyChart Bedside functions at both the patient and admission levels.

[e]OR: odds ratio.

[f]A composite user was defined as a comprehensive user and high-frequency user of MyChart Bedside (MCB). High-frequency users were defined as users who had a total number of MCB sessions greater than or equal to the 75th percentile (41 sessions).

Association Between Comprehensive and Composite Use by MC User Status

Table 3 also summarizes the association between MCB comprehensive use and composite use by MC user status. Prior users and new users were similar in terms of comprehensive and composite MCB use. However, compared with prior users, nonusers had lower odds of being comprehensive users (odds ratio [OR] 0.57, 95% CI 0.45-0.73; *P*<.001) or composite users (OR 0.42, 95% CI 0.29-0.60; *P*<.001).

Sensitivity Analyses of the Association Between MCB Use and MC User Status

Multimedia Appendices 2 and 3 summarize the findings from our sensitivity analyses at the admission and sessions levels. These results were largely consistent with those from analyses conducted at the patient level. There were notable exceptions at the admission level: the Dining on Demand function was used 12% more among new users (patient level: IRR 1.10, 95% CI 0.96-1.27; *P*=.16; admission level: IRR 1.12, 95% CI 0.98-1.28, *P*=.10) compared with prior users and the use of the Tutorial function was not significantly different between nonusers and prior users (patient level: IRR 0.92, 95% CI 0.84-1.00; *P*=.05; admission level: IRR 0.99, 95% CI 0.91-1.08, *P*=.85). At the sessions level, the use of the Dining on Demand (IRR 1.02, 95% CI 0.98-1.06; *P*=.28) and Tutorial functions (IRR 1.04, 95% CI 0.97-1.12; *P*=.27) were similar between prior users and new users. For the Access MyChart (patient: IRR 1.34, 95% CI 1.13-1.58; *P*=.001; admission: IRR 1.37, 95% CI 1.16-1.61; *P*<.001; session: IRR 1.25, 95% CI 1.15-1.36; *P*<.001) and Messages (patient: IRR 1.33, 95% CI 1.06-1.67; *P*=.02; admission: IRR 1.37, 95% CI 1.09-1.71; *P*=.01; session: IRR 1.24, 95% CI 1.14-1.34; *P*<.001) functions, we found smaller differences between new and prior users than at the patient and admission levels.

Although nonusers had lower use of MCB compared with prior users, the difference between the 2 groups was smaller than that found at the patient and admission levels. At the sessions level, the frequency of use of the Happening Soon, Messages, and To Learn functions was not statistically different between nonusers and prior users. In contrast to the patient and admission levels, where nonusers had significantly lower use of most MCB functions compared with prior users, at the sessions level, the use of the Tutorial and Dining on Demand functions was with notable exceptions. Compared with prior users, nonusers had 27% higher use of the Tutorial function (IRR 1.27, 95% CI 1.20-1.32; *P*<.001) and 9% higher use of the Dining on Demand function (IRR 1.09, 95% CI 1.06-1.12; *P*<.001). Although the I Would Like function did not differ by MC user status at the other 2 levels, at the sessions level, nonusers had 38% higher use (IRR 1.38, 95% CI 0.99-1.91; *P*=.06) compared with prior users.

Multimedia Appendix 4 summarizes the results from our sensitivity analyses using information solely from the enrollment admission into the RCT (model 2) and those using information from up to 3 admissions but also using 180 days after enrollment as the cutoff to define MC user status (model 3). For comparison, we provide the results from the analysis of MCB use from up to 3 admissions and 90 days after enrollment (the cutoff to define MC user status; model 1). On restricting the analysis to the enrollment admission, the results were generally consistent with our main findings, with some exceptions. New users had 19% higher use (IRR 1.19, 95% CI 1.00-1.42; *P*=.048) of all active MCB functions compared with prior users. Furthermore, compared with prior users, new users had 15% (IRR 1.15, 95% CI 1.00-1.32; *P*=.05) and 30% (IRR 1.30, 95% CI 1.02-1.67; *P*=.04) higher use of the Dining on Demand and Happening Soon functions, respectively. Finally, compared with prior users, new users had 44% (OR 1.44, 95% CI 1.12-1.85; *P*=.01) greater odds of being comprehensive users of MCB. Using 180 days after enrollment in the trial as the cutoff to define MC user status did not affect the number of prior users (n=695), but there were more new users (n=240) and fewer nonusers (n=636). Our findings were consistent with the results of analyses using 90 days after enrollment in the trial as the cutoff to define MC user status.

Discussion

Principal Findings

In our study of inpatient portal use in relation to outpatient portal user status, we found that the use of specific portal functions as well as the frequency and comprehensiveness of inpatient portal use differed across user groups. Our analyses showed that Dining on Demand and Happening Soon were the most frequently used inpatient portal functions, whereas I Would Like and Notes were the least frequently used functions. These results align with previous findings from a cluster analysis, which also highlighted the variability in function use based on

distinct behaviors of subgroups [21]. Among new users in this study, we found more use of the Tutorial, Messages, and Access MyChart functions than those among prior users. Nonusers were found to use most MCB functions less than prior users. Nonusers were also less likely to be comprehensive or composite users of MCB than prior users.

Prior research has found that an individual's experience with technology may increase their adoption of new technologies by increasing their familiarity with technology and reducing their dependence on external resources for help [22]. Experience may thereby impact an individual's perceptions of the usefulness and ease of use of a new technology [23,24]. This may, in part, explain our finding of higher inpatient portal use by prior users compared with nonusers. At the same time, individuals with less technology experience may use technologies differently (ie, use fewer functions or use less frequently) than those with more technology experience [25], and this prior finding may explain the differences we observed in portal function use between prior users and nonusers. As previous studies have shown that individuals with less prior technology experience may benefit from supplemental education or ongoing support to help them use new technologies [10], it is possible that training patients without prior experience of portal functionality or the benefits of using the portal may increase portal use for nonusers.

Prior users in our sample were mostly older, female, and White. This is similar to the results of another recent study on portal usage from 3 iterations of the Health Information National Trends Survey, which showed that portal users were more likely to be female, White, and have higher education or income. Furthermore, this prior study also showed a widening gender gap and narrowing age gap in portal use among Americans over time [26].

Beyond experience with the outpatient portal, there are likely additional factors that influence both patients' use of the inpatient portal while hospitalized and their use of the outpatient portal post discharge. For instance, barriers to portal use such as low levels of patient activation or low health literacy may contribute to a patient's choice not to use patient portals; such factors may have been at play among the nonusers in our study. Low patient activation may diminish the patient's interest in using portals, as it has been shown that individuals with lower levels of activation have a lower likelihood of using the internet to access health information [27]. In addition, low health literacy has been recognized as a barrier to the use of patient portals, as this may prevent patients from effectively comprehending information found via the portal [28].

The abovementioned barriers to portal adoption and use are more likely to be experienced by vulnerable and underserved patients [29,30], and it is concerning that these patients may also be left behind in the use of portals and in leveraging the potential benefits these tools provide [2-5,14]. Notably, the digital divide may also explain some differences in MCB use among nonusers. Although the inpatient portal is offered on hospital-provided tablets with wireless internet connectivity, the use of the outpatient portal requires a patient to have access to both technology and the internet. Not all individuals have this access, contributing to the digital divide that has been found [31]. For instance, households with low incomes (<US $30,000 per year) are less likely to have home broadband internet access, computers, or smartphones [32]. In addition, enrollment to use an outpatient portal and use of its secure messaging function have been shown to be associated with whether individuals have internet access in their home neighborhoods [6]. Individuals may be less likely to adopt and use an inpatient portal when they do not have internet access at home to enable the use of an outpatient portal post discharge. It is also likely that nonusers include patients visiting the health system for a specific procedure but whose usual care is provided through a separate health system that has a different patient portal platform. These individuals may be less motivated to use either MCB or MC if they perceive their interaction with the AMC to be only temporary.

Interestingly, we found greater use of the inpatient portal for a specific set of MCB functions, such as the Messages function, among new users compared with prior users. As hospitalization may represent a major life event that may spark behavioral changes and increase an individual's engagement in their health care [33], this experience may have contributed to new users' high levels of MCB use during their hospital stay and their creation of MC accounts to manage their health after discharge. New users may also have used the Messages function more because of its novelty and the high expectations placed on it by this highly motivated group of users. In contrast, prior users may have viewed the Messages function as potentially less useful, and more research is needed to understand what factors may have contributed to this behavior.

However, we did not find differences between the new user and prior user groups in the odds of being a composite or comprehensive user. Although these findings contradict our study hypothesis, these results suggest that it is possible that prior users may have had a priori assumptions that they already completely understand portal functionality and perhaps overlooked unique functions that were available in the inpatient portal. Promoting awareness of patient portals and their varied functions may be important across settings, as it can support seamless care transitions and maintain patient engagement. Ultimately, knowing more about the relationship between the use of outpatient and inpatient portals can help inform implementation practices that encourage individuals to use these tools across care settings to better manage their health and participate in their health care.

The results at the patient level were generally consistent with those at the admission and sessions levels. However, there were some notable exceptions. New users had higher numbers of sessions compared with prior users when analyzed at the sessions level (Multimedia Appendices 2 and 3 provide detailed results at the admission and sessions levels). Nonusers had significantly higher use of the Dining on Demand, I Would Like, and Tutorial functions. Furthermore, at the sessions level, there were no significant differences in the use of the Messages and To Learn functions between nonusers and prior users. The results at the other levels, especially at the sessions level, could be highly sensitive to additional system artifacts in spite of our processing of the raw log file data, and more research is required

to explore how different data processing techniques and analytical models may account for these potential idiosyncrasies.

Finally, our findings were generally consistent when examined using information on MCB function use solely from the enrollment admission. It is notable that new users had significantly higher use of all active functions in the enrollment admission, but this association was no longer significant when examining MCB use from up to 3 admissions. This pattern was observed for the Happening Soon and Dining on Demand functions, suggesting a novelty effect. Similarly, the odds of being a comprehensive user were high when examining MCB use from enrollment admission alone.

Limitations

Limitations should be considered when interpreting our results. First, our analysis focused on a single health system and a single patient portal platform, potentially restricting the generalizability of our findings. However, our inclusion of multiple hospitals across this health system and the similarities in patient portal functionality across this and other vendor platforms does support the potential applicability of our findings to other settings. Second, we restricted our focus to the associations between outpatient portal use and inpatient portal use, without considering the alternative influence of inpatient portal use on outpatient portal use. Although we decided to examine this relationship because outpatient portals are more established than inpatient portals, giving patients more opportunity to have been introduced to outpatient portal technology, future studies can be designed to study the alternate relationship. Third, although experience with technology can be measured in multiple ways [34], we defined experience as having an active outpatient portal account. Additional work examining more granular measures such as length of experience with patient portals might provide greater insight into the impact of experience on patient portal use. Fourth, MC users can access the same set of MC functions using the inpatient tablet or personal devices. Therefore, the difference in MCB use (specifically, number of sessions, active functions, and Access MyChart function) by MC user status may be influenced by a patient's use of their mobile device to access their MC app during an inpatient stay. Finally, our study was not designed to explore the implications of portal use on health outcomes. Although functions such as My Health and To Learn have been previously noted to influence patients' health outcomes [21], studying such relationships within and among outpatient user groups was beyond the scope of this study.

Conclusions

Together, outpatient and inpatient portals can provide patients access to their health information and a means to communicate with their health care providers, including along the continuum of care. We found that in comparison with patients who had previously used the outpatient portal, new users of the outpatient portal had higher inpatient portal use for a select set of functions, whereas nonusers of the outpatient portal had lower use of the inpatient portal during their hospitalization. Understanding how the use of one type of patient portal affects the use of the other can help us better understand how these tools can be both implemented and promoted to increase patients' involvement in their health care across settings.

Acknowledgments

This research was supported by grants from the Agency for Healthcare Research and Quality (grant R01HS024091 and grant P30HS024379). This research was reviewed and approved by the Institutional Review Board of the Ohio State University. The authors wish to thank Alice Gaughan, Shonda Vink, Lindsey Sova, and Lakshmi Gupta (all affiliated with the authors' organization) for their assistance with this project as well as all patients and providers who participated in this study.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Fully adjusted models showing MyChart Bedside function use in relation to MyChart user status (in reference to prior users and White patients) at the patient level. These models used information from up to 3 admissions, including the study enrollment admission.
[DOCX File , 37 KB - jmir_v23i4e23866_app1.docx]

Multimedia Appendix 2
Frequency of MyChart Bedside function use in relation to MyChart user status at the patient, admission, and sessions levels from up to 3 admissions, including the study enrollment admission.
[DOCX File , 27 KB - jmir_v23i4e23866_app2.docx]

Multimedia Appendix 3
MyChart Bedside function use in relation to MyChart user status at the patient, admission, and sessions levels.
[DOCX File , 35 KB - jmir_v23i4e23866_app3.docx]

Multimedia Appendix 4
MyChart Bedside function use in relation to MyChart user status at the patient level among different samples (in reference to prior users).
[DOCX File , 40 KB - jmir_v23i4e23866_app4.docx]

References

1. Kruse CS, Bolton K, Freriks G. The effect of patient portals on quality outcomes and its implications to meaningful use: a systematic review. J Med Internet Res 2015 Feb 10;17(2):e44 [FREE Full text] [doi: 10.2196/jmir.3171] [Medline: 25669240]
2. Wade-Vuturo AE, Mayberry LS, Osborn CY. Secure messaging and diabetes management: experiences and perspectives of patient portal users. J Am Med Inform Assoc 2013 May 01;20(3):519-525 [FREE Full text] [doi: 10.1136/amiajnl-2012-001253] [Medline: 23242764]
3. Vink S, Fareed N, MacEwan SR, McAlearney AS. An exploration of the association between inpatient access to tablets and patient satisfaction with hospital care. Perspect Health Inf Manag 2019;16 [FREE Full text] [Medline: 31908632]
4. Zhou YY, Kanter MH, Wang JJ, Garrido T. Improved quality at Kaiser Permanente through e-mail between physicians and patients. Health Aff (Millwood) 2010 Jul 06;29(7):1370-1375. [doi: 10.1377/hlthaff.2010.0048] [Medline: 20606190]
5. Graetz I, Huang J, Muelly ER, Fireman B, Hsu J, Reed ME. Association of mobile patient portal access with diabetes medication adherence and glycemic levels among adults with diabetes. JAMA Netw Open 2020 Feb 05;3(2) [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.21429] [Medline: 32074289]
6. Perzynski AT, Roach MJ, Shick S, Callahan B, Gunzler D, Cebul R, et al. Patient portals and broadband internet inequality. J Am Med Inform Assoc 2017 Sep 01;24(5):927-932 [FREE Full text] [doi: 10.1093/jamia/ocx020] [Medline: 28371853]
7. Hoogenbosch B, Postma J, de Man-van Ginkel JM, Tiemessen NA, van Delden JJ, van Os-Medendorp H. Use and the users of a patient portal: cross-sectional study. J Med Internet Res 2018 Sep 17;20(9):e262 [FREE Full text] [doi: 10.2196/jmir.9418] [Medline: 30224334]
8. Holt KA, Karnoe A, Overgaard D, Nielsen SE, Kayser L, Røder ME, et al. Differences in the level of electronic health literacy between users and nonusers of digital health services: an exploratory survey of a group of medical outpatients. Interact J Med Res 2019 Apr 05;8(2) [FREE Full text] [doi: 10.2196/ijmr.8423] [Medline: 30950809]
9. Norman CD, Skinner HA. eHealth literacy: essential skills for consumer health in a networked world. J Med Internet Res 2006 Jun 16;8(2):e9 [FREE Full text] [doi: 10.2196/jmir.8.2.e9] [Medline: 16867972]
10. Carey JC, Wade SL, Wolfe CR. Lessons learned: the effect of prior technology use on web-based interventions. Cyberpsychol Behav 2008 Apr;11(2):188-195 [FREE Full text] [doi: 10.1089/cpb.2007.0025] [Medline: 18422412]
11. Blackler A, Popovic V, Mahar D. The nature of intuitive use of products: an experimental approach. Des Stud 2003 Nov;24(6):491-506 [FREE Full text] [doi: 10.1016/s0142-694x(03)00038-3]
12. Kelly MM, Coller RJ, Hoonakker PLT. Inpatient portals for hospitalized patients and caregivers: a systematic review. J Hosp Med 2018 Jun 01;13(6):405-412 [FREE Full text] [doi: 10.12788/jhm.2894] [Medline: 29261819]
13. McAlearney AS, Fareed N, Gaughan A, MacEwan SR, Volney J, Sieck CJ. Empowering patients during hospitalization: perspectives on inpatient portal use. Appl Clin Inform 2019 Jan;10(1):103-112 [FREE Full text] [doi: 10.1055/s-0039-1677722] [Medline: 30759491]
14. Dendere R, Slade C, Burton-Jones A, Sullivan C, Staib A, Janda M. Patient portals facilitating engagement with inpatient electronic medical records: a systematic review. J Med Internet Res 2019 Apr 11;21(4) [FREE Full text] [doi: 10.2196/12779] [Medline: 30973347]
15. Creber RM, Grossman L, Ryan B, Qian M, Polubriaginof F, Restaino S, et al. Engaging hospitalized patients with personalized health information: a randomized trial of an inpatient portal. J Am Med Inform Assoc 2019 Feb 01;26(2):115-123 [FREE Full text] [doi: 10.1093/jamia/ocy146] [Medline: 30534990]
16. Kelly MM, Hoonakker PL, Dean SM. Using an inpatient portal to engage families in pediatric hospital care. J Am Med Inform Assoc 2017 Jan;24(1):153-161 [FREE Full text] [doi: 10.1093/jamia/ocw070] [Medline: 27301746]
17. McAlearney AS, Sieck CJ, Gaughan A, Fareed N, Volney J, Huerta TR. Patients' perceptions of portal use across care settings: qualitative study. J Med Internet Res 2019 Jun 06;21(6) [FREE Full text] [doi: 10.2196/13126] [Medline: 31172960]
18. McAlearney AS, Sieck CJ, Hefner JL, Aldrich AM, Walker DM, Rizer MK, et al. High Touch and High Tech (HT2) Proposal: transforming patient engagement throughout the continuum of care by engaging patients with portal technology at the bedside. JMIR Res Protoc 2016 Nov 29;5(4):e221 [FREE Full text] [doi: 10.2196/resprot.6355] [Medline: 27899338]
19. Huerta T, Fareed N, Hefner JL, Sieck CJ, Swoboda C, Taylor R, et al. Patient engagement as measured by inpatient portal use: methodology for log file analysis. J Med Internet Res 2019 Mar 25;21(3) [FREE Full text] [doi: 10.2196/10957] [Medline: 30907733]
20. Tosto GD, McAlearney AS, Fareed N, Huerta TR. Metrics for outpatient portal use based on log file analysis: algorithm development. J Med Internet Res 2020 Jun 12;22(6) [FREE Full text] [doi: 10.2196/16849] [Medline: 32530435]
21. Fareed N, Walker D, Sieck C, Taylor R, Scarborough S, Huerta T, et al. Inpatient portal clusters: identifying user groups based on portal features. J Am Med Inform Assoc 2019 Jan 01;26(1):28-36 [FREE Full text] [doi: 10.1093/jamia/ocy147] [Medline: 30476122]

22. Venkatesh V, Thong JY, Xu X. Consumer acceptance and use of information technology: extending the unified theory of acceptance and use of technology. MIS Q 2012;36(1):157. [doi: 10.2307/41410412]
23. Venkatesh V, Davis FD. A theoretical extension of the technology acceptance model: four longitudinal field studies. Manage Sci 2000 Feb;46(2):186-204. [doi: 10.1287/mnsc.46.2.186.11926]
24. Taylor S, Todd P. Assessing IT usage: the role of prior experience. MIS Q 1995 Dec;19(4):561. [doi: 10.2307/249633]
25. O'brien MA, Rogers WA, Fisk AD. Understanding age and technology experience differences in use of prior knowledge for everyday technology interactions. ACM Trans Access Comput 2012 Mar;4(2):1-27. [doi: 10.1145/2141943.2141947]
26. Hong YA, Jiang S, Liu PL. Use of patient portals of electronic health records remains low from 2014 to 2018: results from a national survey and policy implications. Am J Health Promot 2020 Jul;34(6):677-680. [doi: 10.1177/0890117119900591] [Medline: 32030989]
27. Smith SG, Pandit A, Rush SR, Wolf MS, Simon C. The association between patient activation and accessing online health information: results from a national survey of US adults. Health Expect 2015 Dec 05;18(6):3262-3273 [FREE Full text] [doi: 10.1111/hex.12316] [Medline: 25475371]
28. Sarkar U, Karter AJ, Liu JY, Adler NE, Nguyen R, Lopez A, et al. The literacy divide: health literacy and the use of an internet-based patient portal in an integrated health system-results from the diabetes study of Northern California (DISTANCE). J Health Commun 2010;15 Suppl 2:183-196 [FREE Full text] [doi: 10.1080/10810730.2010.499988] [Medline: 20845203]
29. Hibbard JH, Greene J. What the evidence shows about patient activation: better health outcomes and care experiences; fewer data on costs. Health Aff (Millwood) 2013 Feb;32(2):207-214. [doi: 10.1377/hlthaff.2012.1061] [Medline: 23381511]
30. Lyles CR, Sarkar U. Health literacy, vulnerable patients, and health information technology use: where do we go from here? J Gen Intern Med 2015 Mar;30(3):271-272 [FREE Full text] [doi: 10.1007/s11606-014-3166-5] [Medline: 25588688]
31. Brodie M, Flournoy RE, Altman DE, Blendon RJ, Benson JM, Rosenbaum MD. Health information, the internet, and the digital divide. Health Aff (Millwood) 2000 Nov 01;19(6):255-265. [doi: 10.1377/hlthaff.19.6.255] [Medline: 11192412]
32. Anderson M, Kumar M. Digital divide persists even as lower-income Americans make gains in tech adoption. Pew Research Center. 2019. URL: https://www.pewresearch.org/fact-tank/2019/05/07/digital-divide-persists-even-as-lower-income-americans-make-gains-in-tech-adoption/ [accessed 2021-04-05]
33. Lawson PJ, Flocke SA. Teachable moments for health behavior change: a concept analysis. Patient Educ Couns 2009 Jul;76(1):25-30 [FREE Full text] [doi: 10.1016/j.pec.2008.11.002] [Medline: 19110395]
34. Sun H, Zhang P. The role of moderating factors in user technology acceptance. Int J Hum Comput Stud 2006 Feb;64(2):53-78 [FREE Full text] [doi: 10.1016/j.ijhcs.2005.04.013]

Abbreviations

AMC: academic medical center
CCI: Charlson Comorbidity Index
HIT: health information technology
IRR: incidence rate ratio
LOP: length of provisioning
LOS: length of stay
MC: MyChart
MCB: MyChart Bedside
OR: odds ratio
RCT: randomized controlled trial

Edited by R Kukafka; submitted 26.08.20; peer-reviewed by C Yousef, D Heisey-Grove; comments to author 28.09.20; revised version received 23.11.20; accepted 16.03.21; published 30.04.21.

Please cite as:
Fareed N, Jonnalagadda P, MacEwan SR, Di Tosto G, Scarborough S, Huerta TR, McAlearney AS
Differential Effects of Outpatient Portal User Status on Inpatient Portal Use: Observational Study
J Med Internet Res 2021;23(4):e23866
URL: https://www.jmir.org/2021/4/e23866
doi:10.2196/23866
PMID:33929328

Original Paper

Attitudes of Health Care Professionals Toward Older Adults' Abilities to Use Digital Technology: Questionnaire Study

Ittay Mannheim[1,2], BA, MA; Eveline J M Wouters[1,2], MD, PhD; Leonieke C van Boekel[2], PhD; Yvonne van Zaalen[1], PhD

[1]School for Allied Health Professions, Fontys University of Applied Science, Eindhoven, Netherlands

[2]Tranzo, Tilburg School of Social and Behavioral Sciences, Tilburg University, Tilburg, Netherlands

Corresponding Author:
Ittay Mannheim, BA, MA
School for Allied Health Professions
Fontys University of Applied Science
2 Dominee Theodor Fliednerstraat
Eindhoven, 5631 BN
Netherlands
Phone: 31 62 8754948
Email: ittay.mannheim@gmail.com

Abstract

Background: Digital technologies (DTs) for older adults focus mainly on health care and are considered to have the potential to improve the well-being of older adults. However, adoption rates of these DTs are considered low. Although previous research has investigated possible reasons for adoption and acceptance of DT, age-based stereotypes (eg, those held by health care professionals) toward the abilities of older adults to use DTs have yet to be considered as possible barriers to adoption.

Objective: The aim of this study was to investigate the influencing role of ageism in the context of health care professionals attitudes toward older adults' abilities to use health care DT. A further goal was to examine if social comparison and stereotype activation affect and moderate this association.

Methods: A new measurement to assess health care professionals' attitudes toward older adults using technology (ATOAUT-10) was developed and used in 2 studies. Study 1 involved the development of the ATOAUT-10 scale using a principal component analysis and further examined health care professionals' attitudes toward the use of health care DTs and correlations with ageism. Study 2 further explored the correlation between ageism and ATOAUT in an experimental design with health care professionals.

Results: In study 1, physiotherapists (N=97) rated older adults as young as 50 years as less able to use health care DT compared to younger adults (P<.001). A multiple regression analysis revealed that higher levels of ageism, beyond other predictors, were predictive of more negative ATOAUT, (β=.36; t=3.73; P<.001). In study 2, the salience of age was manipulated. Health care professionals (N=93) were randomly assigned to rate the abilities of a young or old person to use health care DT. Old age salience moderated the correlation between ageism and ATOAUT (R^2=0.19; $F_{6,85}$=3.35; P=.005), such that higher levels of ageism correlated with more negative ATOAUT in the old age salient condition, but not the young condition. Stereotype activation accounted for health care professionals' attitudes more than did the experience of working with older patients or the professionals' age.

Conclusions: Negative and ageist attitudes of health care professionals can potentially affect how older adults are viewed in relation to DT and consequently might influence actual use and adoption of technology-based treatment. Future studies should broaden the validation of the ATOAUT-10 scale on more diverse samples and focus on the discriminatory aspect of ageism and self-ageism of older adults. This study calls for a focus on ageism as a determinant of adoption of DT.

(J Med Internet Res 2021;23(4):e26232) doi:10.2196/26232

KEYWORDS

ageism; attitudes; stereotype activation; digital technology

Introduction

Digital technology (DT), hereby defined as technological devices, services or platforms that use, collect, and often process data and are connected to the internet, other devices, or apps [1], are thought by many to have the potential to improve quality of life and promote independent and active aging of older adults [2-4]. However, older adults are often discoursed as a homogenous group of "nonusers"[5], associated with illness, frailty, cognitive decline, and dependency [6]. This might be one of the reasons why DTs developed for the use of older adults primarily focus on health care [2], which positions health care professionals in the forefront of using DT with older adults. Substantive research has attempted to explain the factors for adoption of DT by older adults in general [7-9] and specifically for health care [10]. Yet, the specific influence of age-based stereotypes and ageism on the use and adoption of DT in care and health care, and the effect of social comparison and stereotype activation on health care professionals' attitudes toward older adults in relation to DT, have not been investigated.

Ageism comprises stereotypes, prejudice, and discrimination toward a person based on their age [11]. This definition reflects a cognitive component (eg, the belief that older adults are less able to use DTs), an emotional component (eg, the feeling that instructing older adults how to use DT is annoying), and a behavioral component (eg, not offering older adults treatments based on DT). The pervasiveness and social acceptability [12,13] of ageism is to some extent explained by its indirect and often implicit nature. Social behavior is often implicitly shaped by environmental cues and activation of stereotypical traits. Ageism is thus internalized throughout the life course and can operate implicitly [14] often without awareness as to how it influences our judgments [15]. Activation of age stereotypes and ageism, therefore, do not necessarily encompass explicit intention to do harm, as it is often expressed subtly in forms of benevolence [16], and older adults being disrespected, ignored, or patronized [17]. Therefore, it is important to measure both implicit and explicit attitudes [18]. Ageism can indeed be harmful and affect the opportunities of older adults for active aging, equal participation, and access to services such as health care [19] or DTs [20]. Furthermore, implicit attitudes and stereotypes are often embodied and self-directed [21] and may eventually lead to decreased physical and mental health [22].

Expressions of ageism manifest in different contexts and life domains, such as health care, leisure and employment, and different domains have different age thresholds as to when a person is considered "old" [23]. Ageism can also be context specific [24], meaning that negative attitudes toward older adults, (eg, use of DT) can vary in the context of family, work, or health care. A central stereotype about older adults that is very much apparent in relation to DT is that they are less competent [12], and simply presenting a question about "old" or "young" can lead to stronger associations with negative age traits [25].

Contrary to stereotypes that portray older adults as "laggards" [26], nonusers [5], and technophobic [27], accumulating evidence suggests that older adults find DT to be fascinating and empowering [28], and hold more positive than negative attitudes toward DT [29]. According to a recent report by the American Association of Retired Persons (AARP), use of various forms of DTs (eg, smartphones, tablets, smart home technologies) by adults aged 50 years and above has consistently increased since 2014, and for many devices, adoption rates are nearly similar to those of younger adults [30]. More importantly, reasons for which older adults use or do not use DT are complex and include social context, emotions, experience, support, and individual preferences [8,31], and perhaps also relate to social influence and attitudes of others [7], such as family members or health care professionals. Noticeably, there is often a mismatch between what is designed for the use of older adults (mainly in the context of health care) and what they actually want and need [32], which may be DT that is both enjoyable and empowering [33]. Older adults express high willingness to use certain care and health care DTs (eg, monitoring sensors), however, only if they perceive that their health status might severely decline [34]. Subsequently, this mismatch might lead to low adoption rates and abandonment of health care–related DT [10].

There is considerable evidence for ageism among health care providers, both self-reported and patient reported [35]. Negative age stereotypes, usually operating in indirect or implicit manners, were found to influence diagnosis, prognosis, and treatments provided. Ambady et al [36] found that implicit measures such as distancing and nonverbal communication of physiotherapists during treatment (eg, looking away from the person) were associated with short- and long-term physical decline of older patients. Patients with similar symptoms or complaints are often diagnosed differently or misdiagnosed because of their age. In a study by Linden and Kurtz [37], the same case description (differing only by age) led to a diagnosis of depression for younger patients and a diagnosis of dementia for older patients. Meanwhile, in a study by Gewirtz-Meydan and Ayalon [38], manipulating the age salient in a case description led to a different prognosis and treatment trajectory of a sexual function complaint. Unfortunately, use of chronological age in triage strategies (eg, in the Covid-19 pandemic) also raises ethical discussions about the influence of ageism on medical decisions [39]. As many aspects of health care are digitalizing, it remains unclear if these ageist manifestations might affect the use of DT with older adults.

There is contradicting evidence regarding characteristics of health care professionals that are associated with ageist outcomes [35]. Some studies indicate that negative attitudes are often associated with younger age of health care professionals [40], whereas positive attitudes are associated with older age, female professionals, and positive experience of working with older patients [41]. Other findings suggest that knowledge about aging and choosing to work with older adults might determine professionals' attitudes and reduce stereotypes and prejudice [42]. The latter might fit in with the idea of social contact theory [43].

In their daily work, health care professionals need to categorize patients in order to reach practical medical decisions. This type of categorization might be considered functional [44]. However, with a view of stereotypes as a process of internalizing and

learning [14], possible biases of health care professionals might be seen as consequence of increased exposure to older adults in situations of illness and dependency. This "clinician bias" might shape the general image health care professionals have of older patients, leading to "diagnostic overshadowing" [45] and pushing them to differentiate themselves from older patients, for example, because of existential fear of their own death [46]. It is therefore plausible that implicitly activated age stereotypes, created in situations of social comparison or categorization [47], determine health care professionals' attitudes toward older adults, more than characteristics of age, gender, or experience. This might lead to the disregarding of individuating information, perception of older adults as a homogeneous outgroup, and discriminatory behaviors.

The aim of this study was to investigate whether ageism plays a role in the perception of health care professionals toward older adults' abilities to use health care DT, or, in other words, whether people who reach a certain age are considered "too old" to use health care DT. Therefore, we first were interested in determining whether higher levels of ageism in health care professionals would be associated with negative attitudes regarding older adults' abilities to use DT. Second, we looked into whether social comparison and stereotype activation would affect this association. Generally, we hypothesized that health care professionals' attitudes of older adults' abilities to use DT would be negative and that higher levels of ageism would be associated with more negative attitudes. Furthermore, we hypothesized that social comparison and stereotype activation would moderate this association, leading to attitudes that are more negative. In study 1 we developed and tested new measurement tools and subsequently assessed the association between ageism and attitudes toward older adults' abilities to use DT. In study 2 we further tested how manipulating stereotype activation might moderate this association. Both studies received ethical approval from the Fontys University of Applied Science ethics research committee (approval file no. Mannheim22022019).

Methods

Study 1

The goal of study 1 was to initially assess the explicit and implicit attitudes of health care professionals toward older adults' abilities to use DT and the association between these attitudes and ageism. As specific measures are currently unavailable, an additional goal was to develop and test the use of new direct and indirect measurements of DT-related attitudes and ageism. We hypothesized that health care professionals would express negative attitudes toward the abilities of older adults (compared to younger adults) to use DT. We further hypothesized that higher levels of ageism would be correlated with more negative attitudes.

Participants

We recruited physiotherapists working in the Netherlands and fourth year physiotherapy students who had already gained professional experience during their internships using available mailing lists between April 2019 and May 2019. Out of the 155 who were contacted, 97 participants voluntarily completed the questionnaire. A statistical power analysis was performed for sample size estimation, using G*Power 3.1.9.4 (Heinrich Heine University Düsseldorf). We used the assumption of a minimal effect size of 0.2 (as no prior knowledge of this scale is available), with an α of .05 and a power of 0.9. The projected sample size needed for a multiple regression with 4 predictors was 82, 75% (73/97) of participants were female, M_{age} = 32.39 (SD 11.24), 23% (22/97) were fourth year students, 34% (33/97) had 1-5 years of work experience, 28% (27/97) had 6-20 years of experience, and the remaining 15% (15/97) exceeded 20 years of experience. Finally, 47% (46/97) indicated that most of their patients were 65 years or older.

Measures

Ageism

The Fraboni Scale of Ageism (FSA) was used as a direct measure of ageism [48]. The FSA assesses all 3 dimensions of ageism: stereotypes, prejudice, and discrimination [11]. As no available translation in Dutch was available, the questionnaire was forward translated from English to Dutch by 2 assessors (YVZ and another assessor), and back translated to English by 2 different assessors (EJMW and another assessor). Differences in interpretation and culturally sensitive aspects were then discussed with all 4 assessors and the corresponding author (IM). The scale consists of 29 items ranked on a 4-point ordinal scale (1, strongly disagree; 2, disagree; 3, agree; 4, strongly agree). Reversed items were recoded and a sum score of the scale was calculated (scale range 29-116). Higher scores represent higher levels of ageism. Missing values in 4 items (7 missing values in total) were replaced by the mean of the item as previously suggested in the use of the FSA by Helmes and Pachana [49]. The Cronbach α coefficient of the scale was .85, similar to the reliability levels found by Fraboni et al [48].

Attitudes Toward Older Adults Using Technology

As a direct measure, we developed the attitudes toward older adults using technology (ATOAUT) scale. Items were developed in accordance with known literature about stereotypes on older adults and technology, such as ease of use and perceived benefit [7,29], fear, anxiety, and self-efficacy [50,51]; our experience from interviewing technology designers and focus groups with older adults [52]; and feedback from experts. We eventually reached a group consensus regarding 15 items that potentially assess stereotypes (eg, "Using digital technology is harder for most older adults") and prejudice (eg, "One needs a lot of patience to explain to an older adult how to use digital technologies") toward older adults and DT (for the full list of items see Table 1). Participants rated their agreement with statements about older adults and DT on a Likert-type scale from 1 (totally disagree) to 6 (totally agree). Five reversed items were recoded, and a sum score of the scale was calculated (scale range 15-90). Higher scores represent attitudes that are more negative.

As an indirect measure, we modified a vignette technique previously used to assess health care–related ageism [37,38]. Participants were presented with 3 descriptions of health care–related DTs, namely a health care app, smartwatches, and rehabilitation videogames (see Multimedia Appendix 1).

Participants were then asked to rate (yes or no) if they believed different age groups (18-30 years, 31-50 years, 51-64 years, 65-79 years, and 80+ years) could use this DT. Positive answers were coded as 1 and negative answers as 0. Answers for each age category in all 3 vignettes were summed, creating a measure between 0-3 for each age group. The Cronbach α coefficient of all items was .82.

Additionally, we used a direct question to assess the belief that age might be a barrier to using DT. Participants rated (yes or no) if they believed that gender, age, culture, or financial situation can limit a person's ability to use technology.

Procedure

Participants received an invitation through email to participate in a study about how older adults use technology in health care and everyday life. Participants were directed to an online questionnaire on Qualtrics, where they gave consent to participate. Afterward, they answered questions about demographic information and to which age group most of their patients belonged. Following this, they answered the DT-related ageism measures (vignettes and ATOAUT scale) and the FSA.

Additionally, we inserted an unrelated validity item ("For this question only, mark the number 2") in the middle of the FSA and ATOAUT scales to check actual reading of the question.

Analysis

SPSS Statistics 26 (IBM Corp) was used to perform the analysis. In order to examine the ATOAUT scale, a principal component analysis was performed, and modifications were made to the scale. In order to examine our hypothesis on the attitudes of physiotherapists on the indirect measure of the vignettes, a repeated measures analysis of variance (ANOVA) was performed, with the sum of vignettes as the dependent variable, the age group assessed as the within-subject independent variable, primary age group of patients the physiotherapists work with as the between-subject independent variable, and the age of the physiotherapists as a covariate. Finally, to examine the correlation between ATOAUT and ageism (FSA scale) and other variables, we used a correlation matrix and a multiple regression.

Study 2

The goal of study 2 was to test how age salience and stereotype activation might moderate the correlation between ageism and ATOAUT. Specifically, we wanted to address the limitations of study 1: that by merely asking all participants to rank the abilities of every age group to use DT, we actually primed social categorization [47] and age-based stereotypes. Therefore, we sought to control the age group salient in the vignettes, so that participants would need to rate the ability of a young or contrastively an old person to use the DTs described in the vignettes. Additionally, we sought to broaden our findings to a more diverse group of health care professionals. We hypothesized that older adults would be assessed as less able than younger adults in using health care DTs. Furthermore, we hypothesized that the "old" salience condition would prime age stereotypes and the need to categorize and differentiate oneself from the older group, leading to attitudes that are more negative. In contrast, rating the "young" condition would allow participants to affirm their self-concept without categorizing and comparing themselves to the older group. Therefore, we hypothesized that the age salient manipulation would act as a moderator in the correlation between ageism and ATOAUT.

Participants

We recruited 93 health care professionals and fourth year health care students in the Netherlands between December 2019 and February 2020. A statistical power analysis was performed for sample size estimation, using G*Power 3.1.9.4 (Heinrich Heine University Düsseldorf). We used the assumption of a minimal effect size of 0.2, with an α of .05 and a power of 0.9. The projected sample size needed for a multiple regression with 6 predictors was 94. Participants were recruited among the students and staff at a university of applied science, and the questionnaire was further distributed within their networks of health care professionals. Of the respondents, 67% (62/93) of the participants were female, M_{age} = 37.01 (SD 11.89), 38% (35/93) were physiotherapists, 25% (23/93) were speech therapists, 17% (16/93) were medical doctors, and 20% (19/93) belonged to other health professions. Moreover, 18% (17/93) were students, 33% (31/93) had 1-5 years of experience, 29% (27/93) had 6-20 years, and the remaining 20% (18/93) had more than 20 years of experience. For patient age, 41% (38/92) of the professionals indicated that most of their patients are 65 or older.

Measures

Ageism

Due to the limitations of the FSA in study 1, we used the Expectations Regarding Aging (ERA-12) scale [53]. Consisting of 12 items, the ERA-12 is a shorter and more updated scale compared to other available ageism scales and is considered to have the most adequate psychometric properties [11]. The items reflect stereotypes about aging in general and toward one's own aging. As a Dutch translation to the ERA-12 was not available, we used the forward–backward translation method described in study 1. Items were ranked on a 4-point ordinal scale (1, definitely false; 2, somewhat false; 3, somewhat true; 4, definitely true). A summed score was calculated (scale range 12-48), with higher scores representing more negative expectations regarding aging. The Cronbach α coefficient of the scale was .78, which was slightly lower than that reported in Sarkisian et al [53].

Attitudes Toward Older Adults Using Technology

As a direct measure, we used the ATOAUT-10 scale developed in study 1 (for the factor analysis and description of how we reduced the scale from 15 to 10 items see the Results section). The Cronbach α coefficient was .73, which slightly lower than that in study 1.

For the indirect measure we used the vignettes developed in study 1 with certain modifications. We used 2 of the 3 vignettes from study 1 (health care app and smartwatch), and replaced the third (videogames in rehabilitation) as it was specific to the context of physiotherapists. Instead, we added a new vignette of using a voice-activated personal assistance, such as Siri,

google personal assistant, or Alexa (see Multimedia Appendix 1). Participants rated the probability that a person (25 or 75 years old) would be able to use the described DT on a Likert-type scale between 1 (not at all) to 6 (very much so). The scores of the 3 vignettes were summed to create the total measure (scale range 3-18). The Cronbach α coefficient was .86.

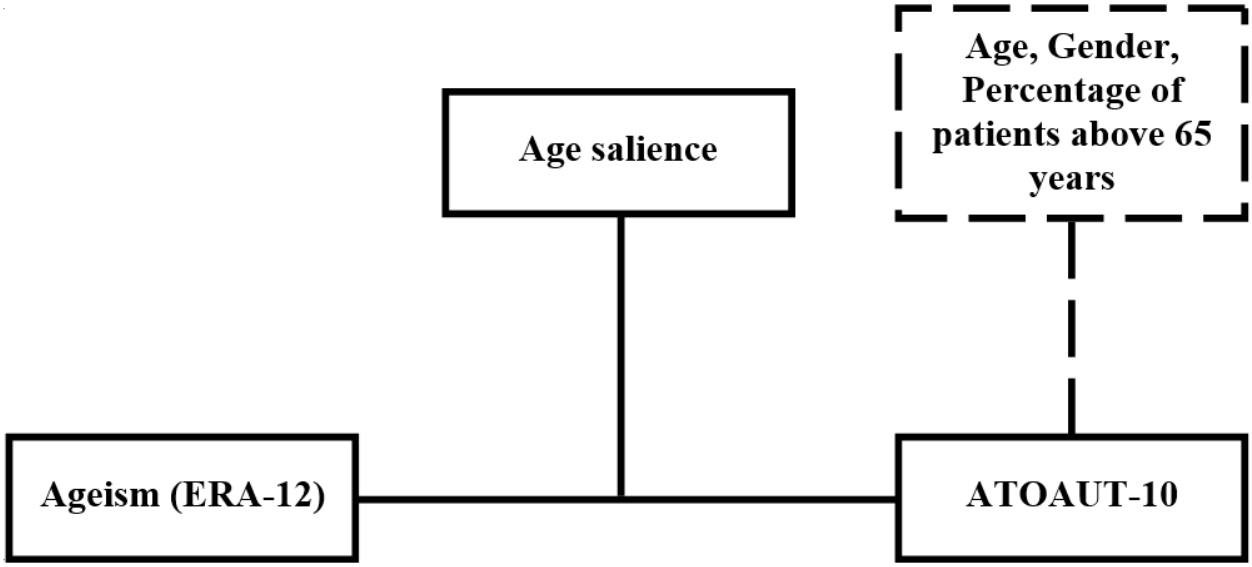
Procedure

Participants received an invitation through email to participate in a study about health care professionals' perspectives about using DT in everyday life and in health care. Participants were directed to an online questionnaire on Qualtrics, where they gave consent to participate and answered demographic questions. Participants were then randomly assigned to rate one of the contrastive age groups (young or old) in the 3 vignettes, and then completed the ATOAUT-10 and ERA-12 scales, and the question used in study 1 about the belief of age being a limiting factor. Finally, participants indicated the percentage of their patients who are above the age of 65 years. One validity check item was inserted in the middle of the ATOAUT-10 scale.

Analysis

SPSS 26 (IBM Corp.**)** was used to perform the analysis. In order to examine our hypothesis that older adults would be assessed as less able than younger adults to use health care DTs on the indirect measure, we used a correlation matrix and multiple regression. In order to examine our hypothesis that age salience would moderate the correlation between ATOAUT and ageism (ERA-12), we employed a regression procedure using the PROCESS bootstrapping macro [54] in SPSS (model 1: 5000 iterations). Figure 1 presents the assumed moderation model.

Figure 1. Assumed moderation model of age salience (young and old) on the correlation between ageism (measured by ERA-12) and attitudes towards older adults using technology (ATOAUT-10). Age, gender, and percentage of patients above 65 years were added to the model as covariates. ATOAUT: attitudes toward older adults using technology; ERA-12: Expectations Regarding Aging scale.



Results

Study 1

We found that 27 participants did not answer the validity items correctly. Consequently, concern regarding their attention in answering was raised. Analysis including and excluding these participants, revealed that the results would not be different. We therefore included all 97 participants in this study.

ATOAUT Scale

The initial Cronbach α coefficient of the 15-item scale was .77. We then used a principal component analysis in order to examine the latent components of the scale. Bartlett test confirmed no violation of sphericity (approximate $\chi^2 = 379.1$; $P<.001$). Five factors with eigenvalues larger than 1 were identified and accounted for 62.07% of the variance (see Table 1). Eight items loaded mainly on the first factor (all loadings above 0.43), representing stereotypes and prejudice toward older adults' abilities to use DT (eg, "Using digital technology is harder for most older adults"), with the exception of item 2, were also strongly loaded on the fourth factor. Two additional items (5 and 12), loaded on the second factor, represented attitudes toward older adults' access to DT and online digital services. Examining the remaining 5 items that did not load on the first and second factor revealed that the phrases used might have been ambiguous and interpreted variably. Consequently, answering them might not necessarily reflect stereotypes or prejudice toward older adults, but rather general attitudes toward the role of DT in improving well-being (item 8), matters of privacy (item 9), accessibility of the design (items 4 and 11), and how playful people of different ages are (item 3).

Item 10, which loaded on the first factor, seemed to have a weaker loading compared to the other items. This might have been related to confusion in the use of negation in this item. We therefore concluded to change the phrasing for future use to "Most older adults *can* give useful feedback about new digital technologies." Finally we formed the new ATOAUT-10 scale, comprising items 1, 2, 5, 6, 7, 10, 12, 13, 14, and 15. The Cronbach α coefficient of the new 10-item scale was higher

(.82), compared to the 15-item scale coefficient (.77), and explained 91.2% of the variance of the 15-item scale.

Table 1. Initial eigenvalues and explained variance and loadings after Varimax rotation of ATOAUT items as sorted by loading size (N=97).[a]

Items and factors	Rotated component matrix				
	1	2	3	4	5
Initial eigenvalues	4.28	1.55	1.28	1.19	1.01
Variance explained (%)	28.51	10.36	8.52	7.95	6.74
ATOAUT[b] item (loadings)					
15. One needs a lot of patience to explain to an older adult how to use digital technologies	.81	.01	–.12	–.02	–.20
7. Using digital technology is harder for most older adults	.76	.19	.04	.22	.08
1. It's hard to explain to older adults how to use digital technology	.71	.06	.01	.21	.08
6. Most older people do not see the benefits of using digital technology	.63	.38	.13	–.07	–.12
14. Most older adults are not interested in learning about using new digital technology	.58	.51	.10	.02	.02
13. Most older adults fear using digital technology because they believe they will break or ruin something	.47	.15	–.35	.11	.19
10. Most older adults cannot give useful feedback about new digital technologies	.43	.19	.21	–.12	.07
5. Most older people have less access to digital technology	.27	.80	–.03	–.02	.10
12. Online services can be used by adults of any age (for example online banking or government services)[c]	.08	.60	–.18	.53	–.05
4. When designing new digital technologies for older adults, older adults should take part in the design process[c]	.19	–.20	.83	–.03	.07
8. Using digital technology can improve older adults well–being and health[c]	.00	.34	.62	.30	–.05
9. Using digital technology can cause more harm to older adults' personal safety and privacy compared to younger adults	.01	.02	.07	.76	–.04
2. Most older adults can use digital technology just as well as younger adults[c]	.53	–.01	.07	.55	.07
11. Digital technology for older adults should be designed in a way that is accessible and easy to use[c]	–.30	.02	.34	–.19	.74
3. Video game devices are mainly for younger adults	.34	.05	–.34	.12	.69

[a]The final ATOAUT-10 scale comprised items 1, 2, 5, 6, 7, 10, 12, 13, 14, and 15.

[b]ATOAUT: attitudes toward older adults using technology.

[c]Reversed item.

Indirect Measure of Attitudes: Vignettes

Table 2 presents the means, SDs, and correlations of the sum of the 3 vignettes for each age category, ATOAUT-10, FSA, and other variables. Physiotherapists assessed older age groups as less able to use the health care DTs described in the vignettes (Figure 2). We further examined these differences using a repeated measures ANOVA, with the sum of vignettes as the dependent variable, age group assessed as the within-subject independent variable, primary age group of patients the physiotherapists work with as the between-subject independent variable, and age of the physiotherapists as a covariate. By interpreting the Greenhouse-Geisser test (due to violation of sphericity), we found a significant main effect of the age group assessed ($F_{2.33,\ 211.74}$=34.66; P<.001; ηp^2=0.28). Post hoc comparisons using the Bonferroni adjustment revealed that each age group above 31-50 years was assessed as significantly less able to use health care DT than the younger age group before it (all P values <.001). The interaction of age group assessed and age of the physiotherapists reached significance ($F_{2.33,211.74}$=5.79; P=.002; ηp^2=0.06), such that younger physiotherapists assessed the ability of older adults to use the DTs as lower than that of older physiotherapists. The interaction between age group assessed and the primary age group of patients the physiotherapists work with was not significant.

Table 2. Means, SDs, and correlations of the sum of the 3 vignettes for each age category, ATOAUT-10, FSA, and other variables (N= 95-97)[a].

Variable	Mean (SD)	1, *r*	2, *r*	3, *r*	4, *r*	5, *r*	6, *r*	7, *r*	8, *r*	9, *r*	10, *r*
1. ATOAUT-10[b]	35.45 (8.45)	—[c]									
2. FSA[d]	50.62 (7.98)	.38**	—								
3. Gender[e]	0.75 (0.43)	.03	–.02	—							
4. Age	32.39 (11.25)	–.18	–.17	.12	—						
5. Primary age group physiotherapists work with[f]	0.47 (0.50)	.10	.05	.02	–.04	—					
6. Belief that age is a barrier to use of DT[g,h]	0.68 (0.47)	.48**	.13	.07	–.25*	.08	—				
7. Sum of 3 vignettes for 18-30 years age group	2.88 (0.36)	.17	.09	.13	.07	.09	.13	—			
8. Sum of 3 vignettes for 31-50 years age group	2.83 (0.45)	–.11	.08	.11	.02	.02	.04	.52**	—		
9. Sum of 3 vignettes for the 51-64 years age group	2.54 (0.78)	–.18	.06	.14	.12	.08	–.06	.28**	.64**	—	
10. Sum of 3 vignettes for the 65-79 years age group	1.77 (1.05)	–.35**	–.05	.16	.32**	.15	–.17	.06	.33**	.64**	—
11. Sum of 3 vignettes for the 80+ years age group	1.22 (1.02)	–.34**	–.22*	.22*	.21*	.14	–.10	.04	.27**	.51**	.76**

[a]2-tailed significant levels presented.

[b]ATOAUT: attitudes toward older adults using technology. Higher score represents more negative attitudes.

[c]Not applicable.

[d]FSA: Fraboni Scale of Ageism. Higher score represents higher levels of ageism.

[e]This variable was dummy coded (0 = male, 1 = female).

[f]This variable was dummy coded (0 = 0-64, 1 = 65+).

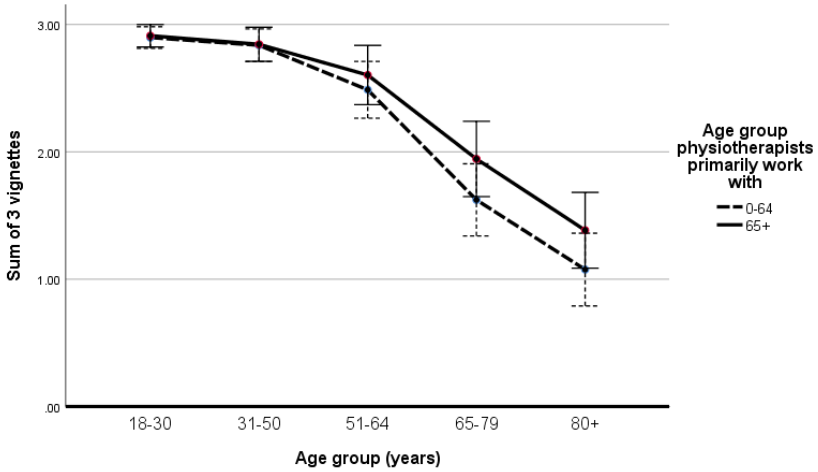
[g]DT: digital technology.

[h]This variable was dummy coded (0 = no, 1 = yes).

*P<.05.

**P<.01.

Figure 2. Assessed ability to use health care digital technologies by age group assessed and the primary age the physiotherapists work with (N=94). Error bars: 95% CI.



Correlation Between Ageism and ATOAUT

As seen in Table 2, a more negative ATOAUT score was significantly correlated with higher levels of ageism (FSA), a belief that age can be a barrier to using DT, and lower perceptions of older adults' abilities to use health care DTs (sum of vignettes for the 65-79 years and 80+ years age groups). Additionally, higher scores of ageism as measured by the FSA were correlated with lower perceptions of older adults' abilities to use health care DTs (80+ years age group). To further examine the correlation between ATOAUT and ageism, we performed a multiple regression with ATOAUT-10 as the dependent variable, and FSA, age, gender, and the primary age group physiotherapists work with as independent variables. A significant model was found (adjusted R^2=0.14; $F_{4,92}$=4.75; P=.002), indicating that beyond all predictors, FSA was the only significant predictor (β=.36; t=3.73; P<.001).

Study 2

One participant did not answer the validity item correctly. Analysis that included this participant revealed that the results would not be altered, and the participant was hence included in the analysis. Table 3 presents the means and SDs of the main variables by condition and correlations. No significant differences were found between participants in the different age salient conditions (young or old) regarding background variables of age, gender, health care profession, and percentage of patients over the age of 65 years that the professional works with. We then analyzed the indirect measure of assessing the ability in using actual health care DTs, using a multiple regression with the sum of the vignettes as the dependent variable and age salience condition and background variables (age, gender, and percentage of patients over the age of 65 years) as independent variables. As observed in Table 3, younger adults were perceived as more likely to use the described DTs compared to older adults. This was qualified by a significant regression model (adjusted R^2=0.59; $F_{4,87}$= 34.15; P<.001), with the age salience condition being the only significant predictor (β=–.78; t =–11.61; P<.001).

Table 3. Means and SDs by age salience (young and old) and correlations (N=93)[a].

Variable	Total (N=93), mean (SD)	Young (n=49), mean (SD)	Old (n=44), mean (SD)	1, *r*	2, *r*	3, *r*	4, *r*	5, *r*	6, *r*	7, *r*
1.Age salience[b]	—[c]	—	—	—		—				
2. ATOAUT-10[d]	36.16 (6.50)	35.73 (6.96)	36.64 (6.00)	.07	—					
3. ERA-12[e]	28.27 (5.10)	27.61 (5.18)	29.00 (4.96)	.14	.06	—				
4. Sum of vignettes	12.72 (3.66)	15.41 (2.28)	9.73 (2.35)	–.78**	–.15	–.14	—			
5. Age	37.01 (11.89)	36.31 (12.15)	37.80 (11.69)	.06	–.31**	.11	–.05	—		
6. Gender[f]	0.67 (.47)	0.61 (.49)	0.73 (.45)	.12	–.03	–.06	–.07	–.06	—	
7. Percentage of patients over the age of 65 years that the professional works with[g]	39.02 (27.66)	38.16 (29.63)	40.00 (25.54)	.03	.20	.26*	–.01	–.11	.01	—
8. Belief that age is a barrier to use of DT[h]	0.73 (.45)	0.69 (.47)	0.77 (.42)	.09	.42**	.03	–.09	–.36**	–.07	.08

[a]2-tailed significant levels presented.

[b]This variable was dummy coded (0 = young, 1 = old).

[c]Not applicable.

[d]ATOAUT: attitudes toward older adults using technology. A higher score represents a more negative attitude.

[e]ERA: Expectations Regarding Aging. A higher score represents a higher level of ageism.

[f]This variable was dummy coded (0 = male, 1 = female).

[g]n=92.

[h]This variable was dummy coded (0 = no, 1 = yes).

*P<.05.

**P<.01.

Examining the direct measure of the ATOAUT-10 scale revealed that negative ATOAUT scores correlated with the younger age of the health care professionals and belief that age is a barrier to using DT. Yet, the simple correlation with ageism (ERA-12) was insignificant. We then tested our hypothesis that age salience would moderate the correlation between ATOAUT and ageism. Table 4 and Figure 3 present the regression coefficients and interaction of age salience X ERA-12. A significant moderation model was found (R^2=0.19, $F_{6,85}$=3.35; P=.005) and the age salience X ERA-12 interaction qualified as a significant moderator, adding significant explanatory variance to the model (R^2 change=0.05; $F_{1,85}$=4.90; P=.03). For the conditional effect of the old age salience condition, negative ERA-12 (higher ageism) was (marginally) associated with a

more negative ATOAUT score (β=.34; *t*=0.82 *P*=.07; CI –0.03 to 0.72), whereas for the young age salience condition, it was not (β =–.22; *t*=–1.23; *P*=.22; CI –0.57 to 0.13). Moreover, the age of the professional was identified as a significant predictor of ATOAUT score, with the younger age of the professional being associated with a negative ATOAUT score.

Table 4. Regression coefficients of the moderation model between ERA-12 and attitudes toward older adults using technology (ATOAUT-10) by age salience, with controlling for age, gender, and professionals' percentage of patients above 65 (N=92).

Variable	Coefficient	SE	*t*	*P* value	LLCI[a]	ULCI[b]
Constant	46.82	5.12	9.14	<.001	36.63	57.00
ERA-12[c]	–0.22	0.18	–1.23	.22	–0.57	0.13
Age salience[d]	–14.97	7.31	–2.05	.04	–29.51	–0.43
Age salience X ERA-12	0.56	0.25	2.21	.03	0.06	1.07
Age	–0.17	0.06	–3.14	.002	–0.28	–0.06
Percentage of patients above 65 years	0.04	0.02	1.74	.09	–0.01	0.09
Gender[e]	–0.74	1.36	–0.54	.59	–3.43	1.96

[a]LLCI: lower level confidence interval.

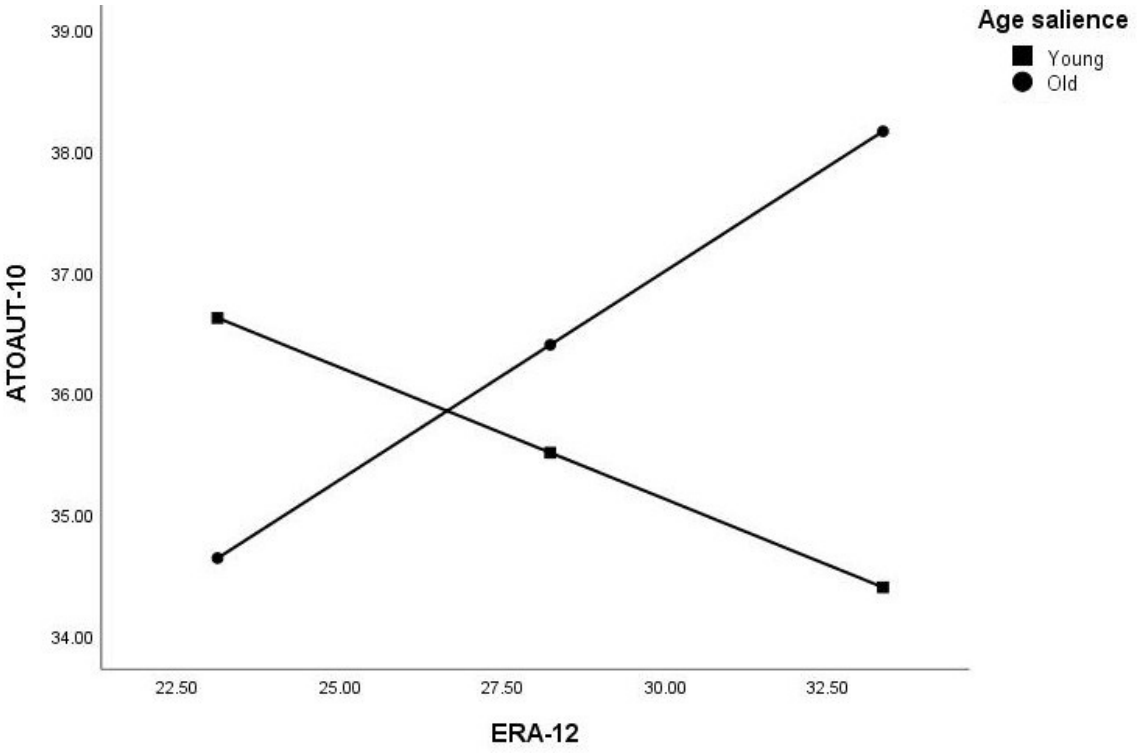
[b]ULCI: upper level confidence interval.

[c]ERA-12: Expectations Regarding Aging. A higher score represents a higher level of ageism.

[d]This variable was dummy coded (0 = young, 1 = old).

[e]This variable was dummy coded (0 = male, 1 = female).

Figure 3. Moderation of the correlation between ATOAUT and ageism (ERA-12) by age salience (N=92). ATOAUT: attitudes toward older adults using technology; ERA-12: Expectations Regarding Aging.



Discussion

Principal Findings

The aim of these studies was to explore attitudes of health care professionals toward older adults' abilities to use DT, as a specific domain of ageism. In study 1 we developed the ATOAUT-10 scale, a direct measurement of stereotypes and prejudice toward use of DT by older adults, and an indirect measurement (vignettes), which were thereafter used in study 2. In the process of developing the ATOAUT-10, we identified ten items that represent stereotypes and prejudice toward older adults in the context of using DT. As hypothesized, significant correlations were found between negative ATOAUT and higher levels of ageism, measured by 2 different ageism scales (FSA in study 1 and ERA-12 in study 2). More so, negative ATOAUT correlated with lower perception of older adults' abilities to use actual health care DTs, measured by the vignettes; belief that

age is a barrier to using DT; and younger age of the health care professionals (in study 2). Thus, enhancing its construct validity. The correlation with ageism measured by the FSA in study 1 was significant and accounted for additional predictors in a multiple regression. Yet, this correlation was small to medium, suggesting that ageism alone does not fully explain physiotherapists' attitudes toward older adults' abilities to use DT.

By using an indirect measure of the vignettes, we were able to assess health care professionals' perceptions on how they believe older adults use health care DT. Our main finding in study 1, that individuals from older age categories are perceived as less able to use DTs is not surprising [10]. Yet still, the difference found between age groups was quite dramatic, with a significant difference between each age group above 31-50 years with the previous age group. Hence, a DT-specific age threshold [23] for negative assumptions about the ability to use health care DTs, might be as early as 50 years, which includes people in their working age who are considered to use DT on a daily basis, and not only people above 80 years. This finding alone might confirm an assessment that is based on stereotypes and not facts, as the majority of older adults in some countries use smartphones and the internet [30]. In study 1, rating all age groups by the participants might have activated age stereotypes by means of categorization and social comparison [47], but in study 2, the age group salient in the vignettes was manipulated (young or old). Once more, a significant difference was found in how the technological abilities of older adults were perceived by health care professionals. More importantly, as hypothesized, this subtle manipulation of age salience moderated the correlation between ageism and ATOAUT, such that higher levels of ageism were associated with a more negative ATOAUT score but only when old age was made salient beforehand. Although the moderator of age salience was significant, it should be noted that the effect size of the moderator's addition to the explained variance was small.

These results reveal perspectives that are quite ageist, considering the accumulating evidence on the increasing prevalence of using DT by older adults. Furthermore, they demonstrate how DT-specific ageism can operate implicitly by merely inducing social comparison or making the concept of age salient. Consequently, this raises the concern that older adults might be discriminated in how they receive (or do not receive) technology-based treatments, as found in other studies with nontechnological treatments [37,38]. This is worrying, considering the discussed benefits of DT in facilitating health care and reducing costs [2]. Notably, we did not explicitly ask health care professionals about their intentions to offer DT-based treatments; therefore, the behavioral and discriminatory aspect of ageism was not addressed in this study. However, it can be assumed that attitudes and beliefs might influence intentions to use and actual use, as is often emphasized by technology acceptance models [7,9,10]. Interestingly, the correlation between ageism and ATOAUT was not significant for participants in the young age salient condition. This could be explained by the matching of the age category (young vs old) and the specific attributes of the context (eg, competency of older adults in using DT) that occurs only in the old age salient condition and may activate specific age-based stereotypes [55]. Although people might be unaware of the underlying processes that influence their attitudes, there is still a matter of controllability of induced behavior and actual use of stereotypes [15]. Explicit and implicit stereotypical evaluations are both, to some extent, prone to belief-based learning processes [18]. Therefore, control over behavioral expressions (namely discrimination) is also a matter of social norms and social acceptability. As ageism is still relatively socially acceptable, expressions of discrimination, especially in domains where older adults are highly stereotyped, such as DT, might be prevalent.

Although previous studies reported gender and more experience of treating older patients [41] as possible predictors of attitudes of health care professionals, these effects on ATOAUT or ageism were not found in this study. Younger age of the professional was found to predict negative attitudes only to a limited extent. Unfortunately, these findings do not shed new light into the inconclusive findings in the literature [35] and do not strongly support the idea that increased social contact with older adults [43] can reduce negative attitudes in relation to health care DTs. This might be due to the higher exposure of health care professionals to older adults who are ill or suffer from chronic conditions [45]. Subsequently, these findings suggest that stereotype activation might be a stronger predictor of negative ATOAUT score. However, the effect sizes found were relatively low. Perhaps a stronger manipulation of stereotype activation, such as priming negative age stereotypes, could lead to stronger effects. Alternatively, other characteristics should be taken into account, such as professionals' desire to work with older adults, the valence of their professional and social contact with older adults, or previous experience of using DT in health care.

Knowledge and training are important aspects when trying to control or combat automatic stereotype activation. Gawronski et al [56] found that training people to acknowledge information that contradicts stereotypes may reduce automatic stereotype activation. Training might offer positive outcomes in reducing ageism in health care and enhance positive contact [57]. Nevertheless, formal training of health care professionals regarding ageing (more so stereotype activation) is still lacking in curricula, and the esteem of working with older patients is still low [58]. We therefore suggest that training should include modules to raise awareness on biases and how easily our behavior is affected by them.

This study focused mainly on the perspective of health care professionals. It is however important to consider that indirect expressions of ageism, including patronizing speech [17], "Elderspeak" [59] or other nonverbal communications [36], might in turn lead to stereotype activation within older patients. These stereotypes are often embodied [21] and directed toward oneself, further affecting participation [19] and actual use of DTs. Hence, attitudes of older adults, as well the reciprocal nature of the interaction between them and others, should be a focus of future research on DT-related ageism.

Limitations

The findings of this study provide a preliminary basis for future research and validation of the ATOAUT-10 scale. However, there were 3 main limitations regarding our measurements. First, our sample size in study 1, while sufficient for the regression analysis, can be considered to be on the lower boundary of minimal sample sizes for factor analysis [60]. Furthermore, the sample in study 1 consisted only of physiotherapists. The sample in study 2, therefore, was broadened to a wider range of health care professionals. However, in order to expand the scale's validity, future studies focusing solely on the validation of the scale with an appropriate sample size and including a diversity of health care professionals and other stakeholders, is needed. Second, it can be claimed that our sample was biased in gender and experience. Although early career and experienced professionals were represented in both studies, a better balanced and planned sampling might have enhanced the validation of the scale. Third, in study 1, a large number of participants missed the validity check item inserted in the FSA, which raised concerns they did not answer the FSA seriously enough. This could be due to the length of the FSA, and some outdated items which might be less suited to Dutch society. A recent review by Ayalon et al [11] found that the psychometric properties of the FSA are considered low compared to other ageism scales. This was the reason we used a different ageism scale in study 2 (ERA-12). Therefore, continued use of the ERA-12 in future studies aimed at replicating these findings is warranted. Finally, as mentioned, our study did not focus on the actual intentions of professionals to use health care DT, or, in other words, the behavioral aspect of discrimination and adoption of DT. This would also be a recommendation for future research with this scale, which could also be used to broaden theoretical models of technology adoption [7,9,10].

Conclusions

The highly negative attitudes of health care professionals toward older adults' abilities to use DTs raise the question of how DTs are used in treatment. Activation of age-based stereotypes seems to play a pivotal role in these attitudes, possibly suggesting that nonadoption of DT is not entirely attributable to chronological age or individual characteristics. Using health care DTs indeed presents an opportunity to improve treatment of chronic diseases and well-being. However, when categorizing the lower access of older adults to use DTs and discussing the digital divide as a purportedly well-established fact [28], we are often preserving a negative view that might hamper the successful implementation of DTs in health care. Instead, it is essential to acknowledge how the field of DT is constantly evolving and includes individuals from all ages with different wants and needs. It is suggested that in order to increase adoption of DT, the focus ought to shift toward how we can change stereotypes and their activation on the individual and societal level.

Acknowledgments

This work was supported by the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie actions (grant agreement no. 764632; "Euroageism").

We would like to thank Anne van Tintelen and Snædís Guðmundsdóttir for assisting in the data collection as well as the colleagues from the Fontys University of Applied Science, Madelon Pijnenburg, Thom Rutten, and Jaron Schnitzer, for participating in the translation process of the FSA and ERA-12. Lastly, we would like to thank Maria Varlamova from the "Euroageism" Innovative Training Network for providing useful statistical and methodological advice. The corresponding author may be contacted for additional information regarding the methods, data analysis, and study material. This study was not preregistered.

Authors' Contributions

All authors were involved in the conceptualization of the study, development of the measurements and scales, data collection, and analysis. IM prepared the first draft of the paper. All authors contributed by reviewing and editing the final draft.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Vignettes used in study 1 and 2.
[DOCX File , 14 KB - jmir_v23i4e26232_app1.docx]

References

1. Mannheim I, Schwartz E, Xi W, Buttigieg SC, McDonnell-Naughton M, Wouters EJM, et al. Inclusion of older adults in the research and design of digital technology. Int J Environ Res Public Health 2019 Oct 02;16(19):1-17 [FREE Full text] [doi: 10.3390/ijerph16193718] [Medline: 31581632]
2. Schulz R, Wahl H, Matthews JT, De Vito Dabbs A, Beach SR, Czaja SJ. Advancing the aging and technology agenda in gerontology. Gerontologist 2015 Oct;55(5):724-734 [FREE Full text] [doi: 10.1093/geront/gnu071] [Medline: 25165042]
3. Czaja S. The potential role of technology in supporting older adults. Public Policy & Aging Report 2017;27(2):44-48 [FREE Full text] [doi: 10.1093/ppar/prx006]

4. Sixsmith A. Technology and the Challenge of Aging. In: Sixsmith A, Gutman G, editors. Technologies for active aging. New York: Springer; 2013:7-25.
5. Quan-Haase A, Williams C, Kicevski M, Elueze I, Wellman B. Dividing the grey divide: deconstructing myths about older adults' online activities, skills, and attitudes. American Behavioral Scientist 2018 May 29;62(9):1207-1228 [FREE Full text] [doi: 10.1177/0002764218777572]
6. Neven L. 'But obviously not for me': robots, laboratories and the defiant identity of elder test users. Sociol Health Illn 2010 Feb 01;32(2):335-347 [FREE Full text] [doi: 10.1111/j.1467-9566.2009.01218.x] [Medline: 20149151]
7. Heerink M, Kröse B, Evers V, Wielinga B. Assessing acceptance of assistive social agent technology by older adults: the Almere model. Int J of Soc Robotics 2010 Sep 4;2(4):361-375. [doi: 10.1007/s12369-010-0068-5]
8. Peek STM, Wouters EJM, van HJ, Luijkx KG, Boeije HR, Vrijhoef HJM. Factors influencing acceptance of technology for aging in place: a systematic review. Int J Med Inform 2014 Apr;83(4):235-248 [FREE Full text] [doi: 10.1016/j.ijmedinf.2014.01.004] [Medline: 24529817]
9. Venkatesh V, Morris M, Davis G, Davis F. User acceptance of information technology: toward a unified view. MIS Quarterly 2003;27(3):425-478 [FREE Full text] [doi: 10.2307/30036540]
10. Greenhalgh T, Wherton J, Papoutsi C, Lynch J, Hughes G, A'Court C, et al. Beyond adoption: a new framework for theorizing and evaluating nonadoption, abandonment, and challenges to the scale-up, spread, and sustainability of health and care technologies. J Med Internet Res 2017 Nov 01;19(11):e367 [FREE Full text] [doi: 10.2196/jmir.8775] [Medline: 29092808]
11. Ayalon L, Dolberg P, Mikulionienė S, Perek-Białas J, Rapolienė G, Stypinska J, et al. A systematic review of existing ageism scales. Ageing Res Rev 2019 Sep;54:100919 [FREE Full text] [doi: 10.1016/j.arr.2019.100919] [Medline: 31228647]
12. Cuddy A, Norton M, Fiske S. This old stereotype: the pervasiveness and persistence of the elderly stereotype. J Social Issues 2005 Jun;61(2):267-285 [FREE Full text] [doi: 10.1111/j.1540-4560.2005.00405.x]
13. Officer A, de la Fuente-Núñez V. A global campaign to combat ageism. Bull World Health Organ 2018 Apr 01;96(4):295-296 [FREE Full text] [doi: 10.2471/BLT.17.202424] [Medline: 29695887]
14. Levy B, Banaji M. Implicit ageism. In: Nelson TD. ed. Ageism: Stereotyping and Prejudice against Older Persons. Cambrige, MA: The MIT Press; 2002:49-75.
15. Bargh J. The four horsemen of automaticity: awareness, intention, efficiency, control in social cognition. In: Wyer RS, Srull TK. eds. Handbook of Social cognition. Vol 1. New York: Psychology Press; 1994:1-40.
16. Cary LA, Chasteen AL, Remedios J. The ambivalent ageism scale: developing and validating a scale to measure benevolent and hostile ageism. Gerontologist 2017 Apr 01;57(2):e27-e36. [doi: 10.1093/geront/gnw118] [Medline: 27520730]
17. Abrams D, Russell P, Vauclair M, Swift H. Ageism in Europe: findings from the European social survey. University of Kent. 2011. URL: https://kar.kent.ac.uk/29733/ [accessed 2021-07-04]
18. Peters KR, Gawronski B. Are we puppets on a string? Comparing the impact of contingency and validity on implicit and explicit evaluations. Pers Soc Psychol Bull 2011 Apr;37(4):557-569. [doi: 10.1177/0146167211400423] [Medline: 21441218]
19. Swift H, Abrams D, Lamont R, Drury L. The risks of ageism model: how ageism and negative attitudes toward age can be a barrier to active aging. Social Issues and Policy Review 2017 Jan 13;11(1):195-231 [FREE Full text] [doi: 10.1111/sipr.12031]
20. McDonough CC. The effect of ageism on the digital divide among older adults. GGM 2016 Jun 16;2(1):1-7. [doi: 10.24966/ggm-8662/100008]
21. Levy B. Stereotype embodiment: a psychosocial approach to aging. Curr Dir Psychol Sci 2009 Dec 01;18(6):332-336 [FREE Full text] [doi: 10.1111/j.1467-8721.2009.01662.x] [Medline: 20802838]
22. Wurm S, Benyamini Y. Optimism buffers the detrimental effect of negative self-perceptions of ageing on physical and mental health. Psychol Health 2014;29(7):832-848. [doi: 10.1080/08870446.2014.891737] [Medline: 24527737]
23. Kornadt AE, Rothermund K. Contexts of aging: assessing evaluative age stereotypes in different life domains. J Gerontol B Psychol Sci Soc Sci 2011 Sep;66(5):547-556. [doi: 10.1093/geronb/gbr036] [Medline: 21571702]
24. Diekman AB, Hirnisey L. The effect of context on the silver ceiling: a role congruity perspective on prejudiced responses. Pers Soc Psychol Bull 2007 Oct;33(10):1353-1366. [doi: 10.1177/0146167207303019] [Medline: 17933733]
25. Perdue C, Gurtman M. Evidence for the automaticity of ageism. Journal of Experimental Social Psychology 1990 May;26(3):199-216 [FREE Full text] [doi: 10.1016/0022-1031(90)90035-k]
26. Essén A, Östlund B. Laggards as innovators? Old users as designers of new services & service systems. Int J Des 2011;5(3):89-98.
27. Nimrod G. Technophobia among older internet users. Educational Gerontology 2018 Jan 24;44(2-3):148-162 [FREE Full text] [doi: 10.1080/03601277.2018.1428145]
28. Wu Y, Damnée S, Kerhervé H, Ware C, Rigaud A. Bridging the digital divide in older adults: a study from an initiative to inform older adults about new technologies. Clin Interv Aging 2015;10:193-200 [FREE Full text] [doi: 10.2147/CIA.S72399] [Medline: 25624752]
29. Mitzner TL, Boron JB, Fausset CB, Adams AE, Charness N, Czaja SJ, et al. Older adults talk technology: technology usage and attitudes. Comput Human Behav 2010 Nov 1;26(6):1710-1721 [FREE Full text] [doi: 10.1016/j.chb.2010.06.020] [Medline: 20967133]

30. Nelson-Kakulla B. Older adults keep pace on tech usage. AARP. URL: https://doi.org/10.26419/res.00329.001 [accessed 2021-07-04]
31. van Boekel LC, Peek ST, Luijkx KG. Diversity in older adults' use of the internet: identifying subgroups through latent class analysis. J Med Internet Res 2017 May 24;19(5):e180 [FREE Full text] [doi: 10.2196/jmir.6853] [Medline: 28539302]
32. Peek S, Aarts S, Wouters E. Can smart home technology deliver on the promise of independent living? A critical re-flection based on the perspectives of older adults. Understanding technology acceptance by older adults who are aging in place: a dynamic perspective 2015:a. [doi: 10.1007/978-3-319-01904-8_41-1]
33. Astell A. Technologyfun for a happy old age. In: Sixsmith A, Gutman G, editors. Technologies for Active Aging. New York: Springer; 2013:169-187.
34. Claes V, Devriendt E, Tournoy J, Milisen K. Attitudes and perceptions of adults of 60 years and older towards in-home monitoring of the activities of daily living with contactless sensors: an explorative study. Int J Nurs Stud 2015 Jan;52(1):134-148. [doi: 10.1016/j.ijnurstu.2014.05.010] [Medline: 24951084]
35. Wyman M, Shiovitz-Ezra S, Bengel J. Ageism in the health care system: Providers, patients and systems. In: Ayalon L, Tesch-Römer C. eds. Contemporary perspectives on ageism. New York: Springer Publishing; 2018:193-212.
36. Ambady N, Koo J, Rosenthal R, Winograd CH. Physical therapists' nonverbal communication predicts geriatric patients' health outcomes. Psychol Aging 2002 Sep;17(3):443-452. [doi: 10.1037/0882-7974.17.3.443] [Medline: 12243386]
37. Linden M, Kurtz G. A randomised controlled experimental study on the influence of patient age on medical decisions in respect to the diagnosis and treatment of depression in the elderly. Curr Gerontol Geriatr Res 2009:475958 [FREE Full text] [doi: 10.1155/2009/475958] [Medline: 20182532]
38. Gewirtz-Meydan A, Ayalon L. Physicians' response to sexual dysfunction presented by a younger vs. an older adult. Int J Geriatr Psychiatry 2017 Dec;32(12):1476-1483. [doi: 10.1002/gps.4638] [Medline: 27981728]
39. Previtali F, Allen LD, Varlamova M. Not only virus spread: the diffusion of ageism during the outbreak of COVID-19. J Aging Soc Policy 2020;32(4-5):506-514. [doi: 10.1080/08959420.2020.1772002] [Medline: 32507060]
40. Lookinland S, Anson K. Perpetuation of ageist attitudes among present and future health care personnel: implications for elder care. J Adv Nurs 1995 Jan;21(1):47-56. [doi: 10.1046/j.1365-2648.1995.21010047.x] [Medline: 7897077]
41. Leung S, Logiudice D, Schwarz J, Brand C. Hospital doctors' attitudes towards older people. Intern Med J 2011 Apr;41(4):308-314. [doi: 10.1111/j.1445-5994.2009.02140.x] [Medline: 20002850]
42. Liu Y, Norman IJ, While AE. Nurses' attitudes towards older people: a systematic review. Int J Nurs Stud 2013 Sep;50(9):1271-1282. [doi: 10.1016/j.ijnurstu.2012.11.021] [Medline: 23265870]
43. Allport G. The Nature of Prejudice. Cambridge, MA: Perseus Books; 1954.
44. Snyder M, Miene PK. Stereotyping of the elderly: a functional approach. Br J Soc Psychol 1994 Mar;33 (Pt 1):63-82. [doi: 10.1111/j.2044-8309.1994.tb01011.x] [Medline: 8173867]
45. Thornicroft G, Rose D, Kassam A. Discrimination in health care against people with mental illness. Int Rev Psychiatry 2007 Apr;19(2):113-122. [doi: 10.1080/09540260701278937] [Medline: 17464789]
46. Greenberg J, Schimel J, Martens A. Ageism: Denying the face of the future. In: TD N. ed. Ageism: Stereotyping and Prejudice Against Older Persons. Cambridge, MA: The MIT Press; 2002:27-48.
47. Tajfel H, Turner J. An integrative theory of intergroup conflict. In: Austin WG, Worchel S, eds. The Social Psychology of Intergroup Relations. Pacific Grove, CA: Brooks/Cole; 1979:33-48.
48. Fraboni M, Saltstone R, Hughes S. The Fraboni Scale of Ageism (FSA): an attempt at a more precise measure of ageism. Can. J. Aging 2010 Nov 29;9(1):56-66. [doi: 10.1017/s0714980800016093]
49. Helmes E, Pachana NA. Dimensions of stereotypical attitudes among older adults: analysis of two countries. Geriatr Gerontol Int 2016 Nov;16(11):1226-1230. [doi: 10.1111/ggi.12613] [Medline: 26460207]
50. Niemelä‐Nyrhinen J. Baby boom consumers and technology: shooting down stereotypes. Journal of Consumer Marketing 2007 Aug 07;24(5):305-312. [doi: 10.1108/07363760710773120]
51. Czaja SJ, Charness N, Fisk AD, Hertzog C, Nair SN, Rogers WA, et al. Factors predicting the use of technology: findings from the Center for Research and Education on Aging and Technology Enhancement (CREATE). Psychol Aging 2006 Jun;21(2):333-352 [FREE Full text] [doi: 10.1037/0882-7974.21.2.333] [Medline: 16768579]
52. Mannheim I, Weiss D, van Zaalen Y, van Boekel L, Wouters E. Why is that robot following me?? Older participants' perspectives of co-designing digital technology. In: Gerontechnology 2020;19(0). 2020 Presented at: 12th World Conference of Gerontechnology; 7-9 October 2020; Trondheim, Norway p. 1 URL: https://doi.org/10.4017/gt.2020.19.s.69734.2 [doi: 10.4017/gt.2020.19.s.69734.2]
53. Sarkisian CA, Steers WN, Hays RD, Mangione CM. Development of the 12-item Expectations Regarding Aging Survey. Gerontologist 2005 Apr;45(2):240-248. [doi: 10.1093/geront/45.2.240] [Medline: 15799989]
54. Hayes A. Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach. New York: Guilford Publications; 2017.
55. Casper C, Rothermund K, Wentura D. The activation of specific facets of age stereotypes depends on individuating information. Social Cognition 2011 Aug;29(4):393-414 [FREE Full text] [doi: 10.1521/soco.2011.29.4.393]

56. Gawronski B, Deutsch R, Mbirkou S, Seibt B, Strack F. When "just say no" is not enough: affirmation versus negation training and the reduction of automatic stereotype activation. Journal of Experimental Social Psychology 2008 Mar;44(2):370-377 [FREE Full text] [doi: 10.1016/j.jesp.2006.12.004]
57. Chonody J. Addressing ageism in students: a systematic review of the pedagogical intervention literature. Educational Gerontology 2015 Jun 17;41(12):859-887 [FREE Full text] [doi: 10.1080/03601277.2015.1059139]
58. Kydd A, Touhy T, Newman D, Fagerberg I, Engstrom G. Attitudes towards caring for older people in Scotland, Sweden and the United States. Nurs Older People 2014 Mar;26(2):33-40. [doi: 10.7748/nop2014.02.26.2.33.e547] [Medline: 24576249]
59. Nussbaum J, Pitts M, Huber F, Krieger J, Ohs J. Ageism and ageist language across the life span: intimate relationships and non-intimate interactions. J Social Issues 2005 Jun;61(2):287-305 [FREE Full text] [doi: 10.1111/j.1540-4560.2005.00406.x]
60. Mundfrom D, Shaw D, Ke T. Minimum sample size recommendations for conducting factor analyses. International Journal of Testing 2005 Jun;5(2):159-168 [FREE Full text] [doi: 10.1207/s15327574ijt0502_4]

Abbreviations

AARP: American Association of Retired Persons
ANOVA: analysis of variance
ATOAUT: attitudes toward older adults using technology
DT: digital technology
ERA-12: Expectations Regarding Aging scale
FSA: Fraboni Scale of Ageism

Edited by G Eysenbach; submitted 05.12.20; peer-reviewed by B Breil, J D'Arcey, L Kremer, P Heyn; comments to author 30.12.20; revised version received 23.02.21; accepted 15.03.21; published 21.04.21.

Please cite as:
Mannheim I, Wouters EJM, van Boekel LC, van Zaalen Y
Attitudes of Health Care Professionals Toward Older Adults' Abilities to Use Digital Technology: Questionnaire Study
J Med Internet Res 2021;23(4):e26232
URL: https://www.jmir.org/2021/4/e26232
doi:10.2196/26232
PMID:33881408

Original Paper

Evaluation of a Novel e-Learning Program for Physiotherapists to Manage Knee Osteoarthritis via Telehealth: Qualitative Study Nested in the PEAK (Physiotherapy Exercise and Physical Activity for Knee Osteoarthritis) Randomized Controlled Trial

Sarah E Jones[1], PhD; Penny K Campbell[1], BAppSc; Alexander J Kimp[1], DPT; Kim Bennell[1], PhD; Nadine E Foster[2,3], PhD; Trevor Russell[4], PhD; Rana S Hinman[1], PhD

[1]Department of Physiotherapy, School of Health Sciences, Centre for Health, Exercise and Sports Medicine, The University of Melbourne, Melbourne, Australia

[2]Primary Care Centre Versus Arthritis, School of Medicine, Keele University, Keele, United Kingdom

[3]STARS Education and Research Alliance, School of Health and Behavioural Sciences, The University of Queensland, Brisbane, Australia

[4]RECOVER Injury Research Centre, The University of Queensland, Brisbane, Australia

Corresponding Author:
Rana S Hinman, PhD
Department of Physiotherapy, School of Health Sciences
Centre for Health, Exercise and Sports Medicine
The University of Melbourne
Alan Gilbert Building, 161 Barry St
The University of Melbourne
Melbourne, 3010
Australia
Phone: 61 03 8344 3223
Email: ranash@unimelb.edu.au

Abstract

Background: The delivery of physiotherapy via telehealth could provide more equitable access to services for patients. Videoconference-based telehealth has been shown to be an effective and acceptable mode of service delivery for exercise-based interventions for chronic knee pain; however, specific training in telehealth is required for physiotherapists to effectively and consistently deliver care using telehealth. The development and evaluation of training programs to upskill health care professionals in the management of osteoarthritis (OA) has also been identified as an important priority to improve OA care delivery.

Objective: This study aims to explore physiotherapists' experiences with and perceptions of an e-learning program about best practice knee OA management (focused on a structured program of education, exercise, and physical activity) that includes telehealth delivery via videoconferencing.

Methods: We conducted a qualitative study using individual semistructured telephone interviews, nested within the *Physiotherapy Exercise and Physical Activity for Knee Osteoarthritis* randomized controlled trial, referred to as the *PEAK* trial. A total of 15 Australian physiotherapists from metropolitan and regional private practices were interviewed following the completion of an e-learning program. The *PEAK* trial e-learning program involved self-directed learning modules, a mock video consultation with a researcher (simulated patient), and 4 audited practice video consultations with pilot patients with chronic knee pain. Interviews were audio recorded and transcribed verbatim. Data were thematically analyzed.

Results: A total of five themes (with associated subthemes) were identified: the experience of self-directed e-learning (physiotherapists were more familiar with in-person learning; however, they valued the comprehensive, self-paced web-based modules. Unwieldy technological features could be frustrating); practice makes perfect (physiotherapists benefited from the mock consultation with the researcher and practice sessions with pilot patients alongside individualized performance feedback, resulting in confidence and preparedness to implement new skills); the telehealth journey (although inexperienced with telehealth before training, physiotherapists were confident and able to deliver remote care following training; however, they still experienced some technological challenges); the *whole package* (the combination of self-directed learning modules, mock consultation, and practice consultations with pilot patients was felt to be an effective learning approach, and patient information booklets supported the

training package); and impact on broader clinical practice (training consolidated and refined existing OA management skills and enabled a switch to telehealth when the COVID-19 pandemic affected in-person clinical care).

Conclusions: Findings provide evidence for the perceived effectiveness and acceptability of an e-learning program to train physiotherapists (in the context of a clinical trial) on best practice knee OA management, including telehealth delivery via videoconferencing. The implementation of e-learning programs to upskill physiotherapists in telehealth appears to be warranted, given the increasing adoption of telehealth service models for the delivery of clinical care.

(J Med Internet Res 2021;23(4):e25872) doi:10.2196/25872

KEYWORDS

osteoarthritis; knee; physiotherapy; exercise; e-learning; qualitative; telehealth; pain; education

Introduction

Background

Physiotherapy care is traditionally delivered via in-person consultations. However, for many people, access to physiotherapy is limited by geographical isolation, or limited local services, or both [1]. To this end, experts in Australia [2] and the United Kingdom National Health Service [3] have identified the need to make full use of digital technologies to facilitate patient convenience and access to care. Such services have become even more critical during the COVID-19 pandemic. Social distancing requirements and *lockdown* restrictions have affected the delivery of in-person health care worldwide, including for noncommunicable diseases, where rehabilitation services have been among the hardest hit [4]. The COVID-19 pandemic has accelerated the drive toward telehealth service delivery as a safe and viable model for physiotherapy services [5,6].

Telehealth is defined by the World Health Organization as the "delivery of health care services, where patients and providers are separated by distance. Telehealth uses information communication technology for the exchange of information for the diagnosis and treatment of diseases and injuries, research and evaluation, and for the continuing education of health professionals" [7]. Data suggest that only a minority of physiotherapists were providing telehealth services prepandemic [8,9], highlighting the lack of telehealth experience within the profession. Delivering physiotherapy via telehealth requires new technical skills and new clinical skills to adapt clinical practice to treating a patient located remotely from the clinician [10,11]. Although generally computer literate, clinicians delivering rehabilitation services for chronic pain typically have limited confidence and knowledge in the use of telehealth [12]. Furthermore, clinician acceptance and confidence in telehealth increases with both training and repeated exposure to telehealth practice [10,12,13]. Thus, specific training in telehealth is required for physiotherapists to effectively and consistently deliver care via this medium [14].

Advancing technologies allow education and training for medical and allied health students and professionals to overcome learning barriers such as distance and the limited availability of specialist training staff while also providing a standardized experience for learners [15]. e-Learning broadly relates to the delivery of educational material through information and communication technology (ICT) [16,17], using the internet to wholly or partially replace the need for a human instructor [18,19]. e-Learning can be synchronous (mediated in real time, eg, videoconference), asynchronous (self-directed, self-paced learning), or a combination of both. A systematic review of studies in practicing or trainee doctors and other health professionals compared e-learning interventions with noninternet learning or no learning interventions. Findings revealed that e-learning was associated with large positive effects on education outcomes (knowledge and skills) compared with no learning, whereas only small, inconsistent effects were seen between e-learning and noninternet learning [18], suggesting that e-learning may be similarly effective to more traditional teaching methods.

Osteoarthritis (OA) is a common and often debilitating chronic joint disease and is one of the leading causes of pain and disability in Australia [20] and worldwide [21]. Symptoms can become increasingly debilitating over time and can greatly affect quality of life, contributing to feelings of dependence and loss of autonomy in older people [22]. As a result of an aging population, combined with increasing obesity rates, the disease burden associated with knee OA is forecast to increase substantially over the coming decade [23]. There is no cure for OA. However, improvements in pain, physical function, and quality of life have been demonstrated with exercise-based interventions [24,25]. Thus, clinical guidelines consistently emphasize education, exercise, weight loss (if required), and support for self-management to alleviate knee OA symptoms before using surgical or pharmacological interventions [26-30].

Given the central role of exercise in disease management, physiotherapists in primary care settings play an important role in providing care to people with knee OA. However, physiotherapists often feel underprepared to manage OA, lacking knowledge about evidence-based practice and confidence in implementing recommendations into routine care [31-34]. The development and evaluation of training programs to upskill health care professionals in OA management has thus been identified as an important priority for improving OA care [35-37]. Telephone- and videoconference-based telehealth interventions have been shown to be effective and acceptable [14,38-40] modes of service delivery for exercise-based interventions aimed at relieving chronic knee pain and improving physical dysfunction and are as effective as in-person care for adults with musculoskeletal pain [41]. Australia's National Osteoarthritis Strategy [20] has also called for the increased implementation of remotely delivered evidence-based OA services.

Objectives

In this study, we explore the experiences of physiotherapists with and their perceptions of an e-learning program aimed at educating physiotherapists about best practice knee OA management, including the implementation of a protocolized management program focused on education, exercise, and physical activity, and how to deliver such care remotely using a videoconferencing platform.

Methods

Design

A qualitative study nested within an ongoing randomized controlled trial (RCT; Australian New Zealand Clinical Trials Registry: ACTRN12619001240134) [42] was conducted. The Physiotherapy Exercise and Physical Activity for Knee OA RCT (known and referred to as the *PEAK* trial) is a noninferiority trial comparing physiotherapist-delivered in-person consultations with physiotherapist-delivered video consultations for people with knee OA. For the RCT, 15 physiotherapists were provided with a structured program of e-learning in best practice knee OA management, including the implementation of a protocolized management program focused on education, exercise, and physical activity as well as how to deliver such care remotely using a videoconferencing platform. In this study, a qualitative approach was chosen to explore the physiotherapist's experience of e-learning as well as its impact and perceived effectiveness. The research design was centered around a constructivist paradigm, which asserts that people generate their own understanding and knowledge subjectively through experience and reflection [43]. The Standards for Reporting Qualitative Research checklist guided the reporting [44].

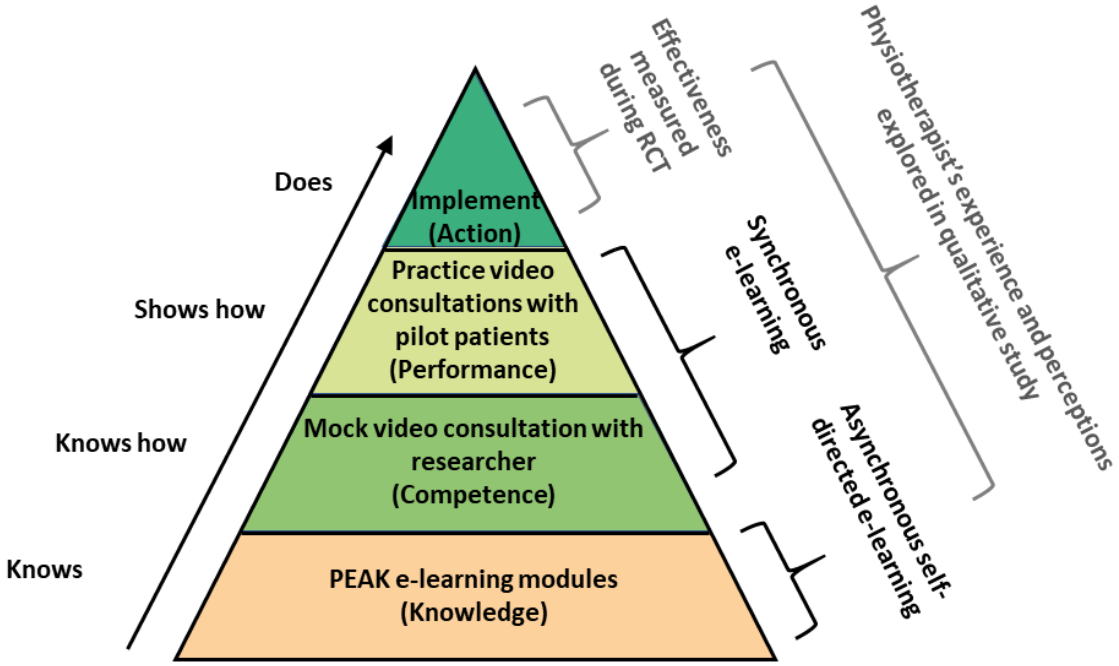
Participants

All 15 physiotherapists recruited to deliver trial interventions for the PEAK RCT participated in this qualitative study. Eligibility criteria for physiotherapist participation in the RCT included current registration to practice as a physiotherapist, private practice located in metropolitan or regional Victoria or Queensland (Australia), access to a computer with internet connection, suitable workspace for confidential video consultations, and some previous experience with videoconferencing software, for example, Skype, Zoom, or Facetime (not necessarily for delivering clinical care). Physiotherapists who met the eligibility criteria were considered for the study, and participating physiotherapists were selected based on their availability and location to ensure geographical spread across metropolitan and regional areas of Victoria and Queensland. All participants provided written informed consent, and the qualitative study was approved by the Institutional Human Research Ethics Committee, separate from the RCT.

Training Program

The PEAK trial training program was developed by the research team, using an e-learning approach featuring both asynchronous and synchronous learning, developed in line with the Miller Pyramid (Figure 1), a learning model designed for use in health education [45]. The model proposes that the foundation of the pyramid is knowledge (*knows*), with learners understanding the relevant facts and theories. The next level is competence (*knows how*), with learners developing skills necessary to apply knowledge, demonstrated via a mock video consultation with a researcher (simulated patient). Next is performance (*shows how*), with learners demonstrating, via practice video consultations with pilot patients, that they can apply knowledge and skills in a realistic clinical context. The final stage represents action (*does*)—the stage at which the learner has integrated their knowledge, skills, and attitudes and is prepared to integrate the competency into their clinical practice. Results from the linked RCT will provide data on the effectiveness of the implementation.

Figure 1. Schematic representation of the PEAK (Physiotherapy Exercise and Physical Activity for Knee Osteoarthritis) training components mapped to the Miller Pyramid learning model. PEAK: Physiotherapy Exercise and Physical Activity for Knee Osteoarthritis. RCT: randomized controlled trial.



Physiotherapists first completed approximately 5 hours of self-directed e-learning modules delivered at the University of Melbourne Learning Management System (LMS; Canvas LMS by Instructure, 2019) covering OA management best practices (including a structured physiotherapy treatment protocol), telehealth (the delivery of care via Zoom videoconferencing), and trial procedures. Each module included a quiz to help reinforce user's knowledge and learning. e-Learning was completed at the user's own self-selected pace (ideally within 4 weeks) and delivered knowledge as the foundation of the physiotherapist's learning. The PEAK training program e-learning modules have since been released for use by clinicians more widely [46].

The e-learning then moved to *practical* synchronous components, whereby physiotherapists participated in a mock initial consultation via videoconferencing with a physiotherapist researcher (AJK; simulated patient), who provided immediate verbal feedback on consultation quality and performance and evaluated the physiotherapist's competency according to a standardized competency checklist (Multimedia Appendix 1). The next stage of *practical* synchronous e-learning required physiotherapists to complete 4 practice video consultations, with 2 pilot patients with chronic (>3 months) knee pain (recruited by research staff) to practice their video consultation skills and apply the education, exercise, and physical activity management program taught in the e-learning modules in a realistic scenario. The research staff conducted spot-checks of consultation recordings to assess the fidelity of the practice consultations to the protocolized treatment plan. Written feedback on common mistakes was provided to participants, and personalized feedback was provided verbally to individual physiotherapists after all pilot patient consultations had been completed. Physiotherapists were financially compensated for the time away from private practice that was dedicated to all elements of participating in the training program and the time taken to complete the interviews.

Semistructured Interviews

Semistructured interviews were conducted after physiotherapists completed all components of the training program and commenced treating participants in the PEAK RCT. The interview schedule (Table 1) was developed based on a constructivist schema whereby physiotherapist expectations, training experience, and perceptions of effectiveness of the e-learning program in changing confidence, imparting knowledge, and addressing learning needs were explored. Questions were loosely aligned with the theoretical framework of acceptability [47], focusing on the framework's five most relevant constructs to the e-learning program (affective attitude, burden, perceived effectiveness, intervention coherence, and self-efficacy). All interviews were conducted over the telephone by SEJ (a researcher who is not a physiotherapist but is trained in qualitative research) who was not involved in the RCT or in developing the training program and was otherwise unknown to the participants. Interviews lasted approximately 30 minutes and were audio recorded and then transcribed verbatim by a third party. Audio recordings and transcripts were deidentified, with transcripts assigned gender-matched pseudonyms to ensure participant confidentiality. All data were stored in digital format on a password-protected university server.

Table 1. Semistructured interview schedule (loosely aligned with the theoretical framework of acceptability).

Topic and construct	Questions (prompts)
Expectations	
Affective attitude; self-efficacy	
	1. Can you tell me how confident you were in your ability to deliver an OA[a] management program before the PEAK[b] trial training? Previous OA management training? *Usual* management of knee OA before training?
	2. Can you tell me about how confident you were in your ability to deliver physiotherapy to OA patients via videoconferencing, before you started the PEAK training? Training or professional development or experience in telehealth?
	3. What were your overall impressions and experience of the training program? Expectations? Comparison with previous professional development?
Training experience	
Burden	
	4. What did you like and dislike about the training program? Consider both the web-based modules and the *practical* aspects of training (ie, the *competency* video consultation and 4 practice patient video consultations)? Time taken for web-based modules? Volume of material?
Intervention coherence	
	5. To what extent did you feel your learning needs were met (or not) through the PEAK training program? Was there anything else you felt need to be covered or a required a greater focus in the training? Was there anything included that you felt was unnecessary?
	6. What were your thoughts on the content and presentation of the content in the PEAK web-based modules? How did you feel about the volume of text presented? How would you prefer to see the content delivered? If you were to change the content or its presentation, what would you suggest?
	7. What were your impressions about the usability of the web-based training platform? Like/dislikes/changes?
	8. How did you find the *competency video consultation* with the research staff member? Was it useful?
	9. Can you tell me a little bit about what you thought about the practice video consultation sessions with the pilot patients with knee pain? Were these consultations useful?
Perceived effectiveness	
	10. Can you tell me overall how useful you found the PEAK trial training program? Intent to implement changes to usual clinical practice?
	11. How confident do you feel in delivering OA management to patients now, after completing the PEAK trial training? What elements do you feel most/least confident with?
	12. How confident do you feel in delivering physiotherapy to your OA patients over videoconference, now that you have completed the PEAK training? What elements do you feel most/least confident with?
Concluding remarks	
	13. Given the restrictions on clinical practice imposed by the current COVID-19 pandemic, has the PEAK training program enabled you to make any changes to your own current clinical practice?
	14. Thank you very much for all your time today and all your time and efforts with this study. Is there anything else you would like to add about your experiences with the PEAK training program?

[a]OA: osteoarthritis.

[b]PEAK: Physiotherapy Exercise and Physical Activity for Knee Osteoarthritis.

Data Analysis

Data analysis was conducted using a thematic approach [48]. Shortly after transcription, interview transcripts were read by SEJ to familiarize with the data. Transcripts were then reread and coded, with text indexed into topics, each identified with a short descriptor. To demonstrate the credibility and confirmability of the emergent topics and patterns, coding was also performed by a second researcher PKC (not a physiotherapist but assisted with developing the training program and recruited physiotherapists into the RCT, and trained in qualitative research). Topics identified by both SEJ and PKC were reviewed, and in collaboration, closely related topics were collated to generate emergent themes within the data. All transcripts were also read by RSH (a physiotherapist who led the development of the training program and an experienced qualitative researcher) who confirmed the relevance of emergent themes across transcripts. Emergent themes were then further refined, ensuring clear and encompassing definitions were generated when naming final themes and subthemes. Themes and subthemes were presented with exemplary quotes from interviews to demonstrate the transferability of the results [49].

Results

Physiotherapist Characteristics

All 15 physiotherapists recruited to deliver care as part of the PEAK RCT were invited to participate in the qualitative interviews. The sample comprised physiotherapists who worked in private practices across 2 Australian states. Physiotherapist characteristics are summarized in Table 2, including the number of PEAK RCT participant consultations conducted by each physiotherapist at the time of the interview. Participating physiotherapists were more often male (11/15, 73%). There was an even divide between those who worked in major cities (8/15, 53%) and those who worked in regional areas (7/15, 47%). Of the 14 physiotherapists, 8 of them (60%) had no previous experience with telehealth. The mean (SD) of years of clinical experience was 11 years (SD 4), and the mean numbers of PEAK RCT participants who physiotherapists had already treated at the time of the interview were 2 (SD 1) via videoconferencing and 2 (SD 2) in-person.

Table 2. Physiotherapists' characteristics (n=15).

Pseudonym	Sex	Geographical location[a]	Clinical experience (years)	Previous experience with telehealth	PEAK[b] RCT[c] videoconferencing participants at the time of the interview, n	PEAK RCT in-person participants at the time of the interview, n
Daniel	Male	Outer regional, Victoria	19	Yes	1	0
Edward	Male	Major city, Queensland	10	No	1	0
Gregory	Male	Major city, Victoria	4	No	0	0
Jason	Male	Inner regional, Queensland	12	Yes	5	2
Robert	Male	Outer regional, Victoria	12	No	2	2
Steven	Male	Inner regional, Victoria	9	No	2	3
Nicole	Female	Inner regional, Victoria	15	No	3	5
Anthony	Male	Major city, Queensland	6	Yes	2	3
Mark	Male	Major city, Queensland	9	Yes	2	3
Douglas	Male	Major city, Victoria	18	No	4	3
Caroline	Female	Major city, Victoria	7	No	1	4
Leslie	Female	Major city, Queensland	9	No	0	1
Brian	Male	Inner regional, Queensland	14	Yes	3	0
William	Male	Major city, Victoria	10	No	1	1
Vicki	Female	Outer regional, Queensland	5	Yes	1	4

[a]Level of remoteness, based on residential postcode, in accordance with the Australian Statistical Geographical Classification-Remoteness Area.

[b]PEAK: Physiotherapy Exercise and Physical Activity for Knee Osteoarthritis.

[c]RCT: randomized controlled trial.

Emergent Themes

A total of five themes, with associated subthemes, emerged and are described in Multimedia Appendix 2.

Theme 1: The Experience of Self-Directed e-Learning

Physiotherapists found the self-directed e-learning modules to be of high quality, comprehensive, and user-friendly and that they would "benefit the least experienced physiotherapist and the most experienced physiotherapist". Although physiotherapists were more familiar with "hands-on" in-person

professional development training for furthering their knowledge and clinical skills, they highly valued the web-based structure. In particular, the self-paced nature of the program was highly regarded and enabled the physiotherapists to fit the training into their busy daily lives and complete it as time allowed. However, the physiotherapists spoke of feeling annoyed and frustrated by some unwieldy features of the delivery platform (including log-in and navigation). Although most physiotherapists spoke highly of the web-based modules, 3 physiotherapists with divergent views spoke of the e-learning modules being "slow going", "tedious", or "fatiguing". The telehealth learning module content, in particular, was felt to be "dry" and "overtly obvious".

Theme 2: Practice Makes Perfect

Physiotherapists frequently referenced the benefit gained from individualized performance feedback from the research team during or after the practical components of the training. They spoke of "learning from applying" and that feedback was helpful ahead of implementing their new skills in a true clinical scenario. Physiotherapists found the mock consultation with the researcher (simulated patient) facilitated the transition from theory to implementation, commenting it "gave us a feel of what to expect for the upcoming pilots" and that it prevented them having to "worry about looking silly with the patients." Similarly, the practice consultations with the pilot patients with chronic knee pain were valued, with physiotherapists reporting that this stage of practical e-learning was "hugely beneficial" and "consolidated everything" ahead of consulting with actual patients. Upon completion of the practical components of the e-learning program, physiotherapists expressed a high level of confidence and readiness to go into their first patient consultations using their new skills and knowledge. Two physiotherapists with divergent views felt that the mock consultation with the researcher was "daunting" or "confronting", with one feeling that it was not necessary. The same physiotherapist also felt that the practice consultations with pilot patients were "surplus to needs".

Theme 3: The Telehealth Journey

For the most part, physiotherapists were inexperienced in telehealth before training. Experience was largely constrained to videoconferencing for social purposes rather than for health care delivery. Physiotherapists in general were nervous or uncertain about providing care to patients via telehealth before undertaking the training program. However, 3 physiotherapists felt that they were moderately confident in telehealth before training, with one stating that "it didn't seem like too much of a leap to deliver these services [exercise prescription via telehealth] to people." After completing the training, most physiotherapists were much more confident and reported that they felt ready to deliver OA care via telehealth using videoconferencing. Despite improved confidence and feelings of preparedness, physiotherapists still felt that telehealth posed some challenges, particularly in supporting patients with OA to be able to navigate videoconferencing technology effectively.

Theme 4: The Whole Package

Physiotherapists highlighted the benefits of the structured approach of the e-learning program. They found that the combination of self-directed e-learning modules, followed by practical components (mock consultation and pilot patient consultations), was very effective as a "whole package" to develop the knowledge and skills required for best practice OA care via telehealth. Physiotherapists valued the resources and patient information booklets that accompanied the training program, stating that the resources "made it really easy" to navigate the consultations and supported the implementation of their knowledge and skills into practice. One physiotherapist with a divergent view appreciated the package but felt that they would get more out of in-person training because of their learning style.

Theme 5: Implementation in Broader Clinical Practice

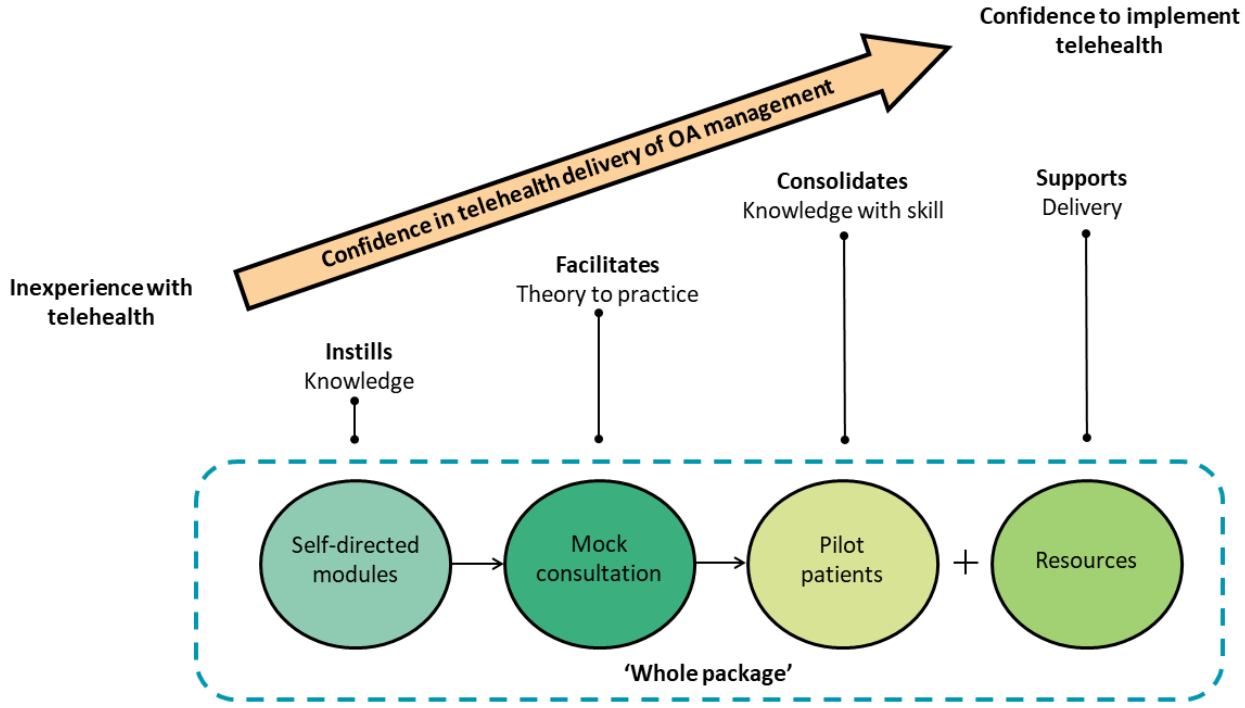
Although the greatest knowledge gains from the training program were regarding the practical implementation of telehealth for the PEAK randomized trial, physiotherapists also described their intent to apply new knowledge and skills in OA management more broadly to their own private practice. They spoke of how the training program filled "knowledge gaps", particularly regarding patient education, and provided them with a more structured approach to OA management that could be used when treating their own clients. Given that the training was completed before the COVID-19 pandemic, physiotherapists felt they were "well equipped for this COVID situation" and expressed feeling "ahead of the curve" when it came to implementing telehealth physiotherapy services more broadly, given social distancing and lockdown restrictions. They felt that the skills they had learned allowed them to confidently switch their private practice service delivery to telehealth as well as teach their colleagues about telehealth. Overall, 4 physiotherapists commented that despite the COVID-19 pandemic, they had not ramped up telehealth services or experienced demand for telehealth in their clinical practice.

Discussion

Principal Findings

Our study shows that physiotherapists accepted e-learning despite their lack of familiarity with professional development delivered entirely via the internet. Physiotherapists valued both the theoretical and practical components of training, which together formed the *whole package*. Most physiotherapists described increasing confidence in providing OA management via telehealth as they progressed through the elements of the e-learning program (Figure 2). The e-learning approach first built a foundation based on new knowledge, before facilitating a transition from theory to practice, then consolidating knowledge in a clinical scenario, and ultimately resulting in the implementation of telehealth into their broader clinical practice.

Figure 2. Schematic representation of emergent themes and their contribution to the development of confidence in delivering osteoarthritis care via telehealth. OA: osteoarthritis.



Physiotherapists perceived the asynchronous aspect of the self-directed e-learning modules to be generally user-friendly; however, the unwieldy technological features of the delivery platform could be frustrating. A previous systematic review of enablers and barriers affecting e-learning in health science education found that 33% (8/24) of the studies identified the lack of user-friendly ICT as a barrier to successful e-learning [50]. This highlights the importance of a delivery platform that is simple and user-friendly when developing e-learning resources to maximize user engagement. Similarly, a previous qualitative study identified user-friendly training and technology as key factors in the successful implementation of telehealth in diabetic foot care [51]. Physiotherapists valued the comprehensive and self-paced nature of the web-based modules, which is consistent with other research. A previous study evaluated the perceptions of physiotherapy students on asynchronous e-learning for the management of chronic health conditions [52]. Students consistently highlighted the flexibility to work at their own pace and time and access to comprehensive information as advantages of e-learning. Collectively, these findings suggest that the ability to self-pace a person's learning should be prioritized when creating adult e-learning materials.

Telehealth practical components of our e-learning program were valued by physiotherapists, which supports previous research highlighting the importance of practical components in health care education, given the need to develop *clinical or hands-on skills* alongside knowledge [53-56]. In a study evaluating e-learning for physiotherapy students on intensive care work placements [57], students perceived the e-learning modules to be helpful in preparing them for their clinical rotation and reducing their anxiety. However, e-learning modules could only sufficiently prepare students when integrated with an in-person clinical placement, where practical experience was considered necessary to build clinical reasoning skills. Similarly, other studies have noted that supervised practice and clinical simulations are valuable for building confidence and preparedness for physiotherapy students to provide care in clinical situations [58,59]. A qualitative study showed that physiotherapy students felt that clinically oriented learning, focusing on knowledge, skills, and *learning through doing*, was an important feature to include in an e-learning program about the management of chronic health conditions [52]. It is likely that the benefit of and amount of practical training required might differ depending on the level of the learner's previous relevant clinical experience as well as the nature of the clinical skill being taught, and therefore, all of these factors should be considered when developing e-learning programs. From a pedagogical perspective, learning approaches incorporating both knowledge and skills practice are consistent with adult learning theories centered around the concept of competency being developed in sequential stages of learning, grounded in theory with proficiency built through experience [45,60,61]. Our ongoing PEAK RCT will provide additional insights into the fidelity and clinical effectiveness of the structured physiotherapy treatment plan delivered using telehealth compared with the same treatment plan delivered in person.

Although most physiotherapists described increased confidence in the delivery of telehealth after training, we do not have quantitative measures in this study documenting changes in telehealth proficiency over time. Thus, we cannot draw strong conclusions about how clinical skills in telehealth changed as a result of the training. A previous study of e-learning in physiotherapists showed that asynchronous e-learning was effective in increasing quantitatively measured confidence and

knowledge in delivering self-management interventions for OA and lower back pain to patients and that the intervention was delivered with high fidelity despite a lack of supervised practical training [62]. Future research incorporating quantitative measures of intervention confidence, knowledge, and fidelity may be warranted to compare outcomes between e-learning approaches with and without synchronous practical elements of training. Future research may also aim to explore innovative alternatives to synchronous practice consultations. Such examples might include encouraging the learner to practice skills with friends or family, including completion of self-audit and self-reflection exercises.

Our sample of Australian physiotherapists was inexperienced with telehealth before training and lacked confidence in delivering care remotely. This is unsurprising given that the uptake of telehealth across Australian health services before 2020 was minimal [63,64], despite recognition of its potential to facilitate efficient health service delivery [65-67]. The COVID-19 pandemic has driven a step change in health care delivery, with rapid adoption of telehealth services by health care providers [4], coupled with increased funding for telehealth services (including physiotherapy) in several countries, including Australia [6,68] and the United Kingdom [69]. Our findings showed that the e-learning program facilitated most physiotherapists to make a rapid switch to telehealth with the sudden onset of the COVID-19 pandemic. This suggests that e-learning courses may be an effective means to train physiotherapists in best practice telehealth delivery. Education providers may also wish to incorporate e-learning telehealth training into the entry-to-practice curriculum to prepare emerging practitioners for the evolving digital health landscape.

Strengths and Limitations

Strengths of our study include the evaluation of an e-learning program for which the e-learning modules are now freely accessible to clinicians globally. Our qualitative evaluation allowed for a thorough understanding of participants' experiences and perceptions of participating in the e-learning program. Our interviews were conducted and the analysis was led by a person who was unknown to the participants and who was not a physiotherapist, minimizing the chances of personal or professional bias influencing findings. Limitations include that our sample was of limited size and comprised solely of physiotherapists who had applied and been selected to deliver the intervention as part of the PEAK RCT. Participants were bound by their clinical trial agreement to complete the training, and they were financially compensated for their time. Therefore, we cannot generalize our findings to the general population of physiotherapists who may be unwilling or unmotivated to complete training in their own time. As the training program is also only available in English, we cannot generalize the findings to non-English speakers nor to countries that may have a different scope of physiotherapy practice to that of Australia.

Conclusions

In conclusion, this study provides evidence for the perceived effectiveness and acceptability of an e-learning program to train physiotherapists (in the context of a clinical trial) on best practice knee OA management, including telehealth delivery via videoconferencing. The implementation of e-learning programs to upskill physiotherapists in telehealth appears to be warranted, given the increasing adoption of telehealth service models for the delivery of clinical care.

Acknowledgments

This work was supported by the National Health and Medical Research Council (grant 1157977). RSH was supported by the National Health and Medical Research Council Senior Research Fellowship (grant 1154217). KB was supported by the National Health and Medical Research Council (grant 1174431). NEF is a senior investigator at the UK National Institute for Health Research.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Standardized competency checklist used to provide personalized verbal feedback to physiotherapists during a mock consultation with a researcher (simulated patient).
[DOCX File , 17 KB - jmir_v23i4e25872_app1.docx]

Multimedia Appendix 2
Themes, subthemes, and exemplary quotes.
[DOCX File , 23 KB - jmir_v23i4e25872_app2.docx]

References

1. Papandony MC, Chou L, Seneviwickrama M, Cicuttini F, Lasserre K, Teichtahl A, et al. Patients' perceived health service needs for osteoarthritis (OA) care: a scoping systematic review. Osteoarthr Cartil 2017 Jul;25(7):1010-1025 [FREE Full text] [doi: 10.1016/j.joca.2017.02.799] [Medline: 28232144]

2. Bradford NK, Caffery LJ, Smith AC. Telehealth services in rural and remote Australia: a systematic review of models of care and factors influencing success and sustainability. Rural Remote Health 2016;16(4):4268 [FREE Full text] [Medline: 27817199]
3. NHS five year forward view. National Health Service (NHS) UK. 2014. URL: https://www.england.nhs.uk/publication/nhs-five-year-forward-view/ [accessed 2020-11-03]
4. Rapid assessment of service delivery for NCDs during the COVID-19 pandemic. World Health Organization. 2020. URL: https://www.who.int/publications/m/item/rapid-assessment-of-service-delivery-for-ncds-during-the-covid-19-pandemic [accessed 2020-10-14]
5. Duckett S. What should primary care look like after the COVID-19 pandemic? Aust J Prim Health 2020;26(3):207. [doi: 10.1071/py20095]
6. Stanhope J, Weinstein P. Learning from COVID-19 to improve access to physiotherapy. Aust J Prim Health 2020;26(4):271. [doi: 10.1071/py20141]
7. Telehealth. World Health Organisation. 2016. URL: https://www.who.int/gho/goe/telehealth/en/ [accessed 2020-02-11]
8. Lee AC, Davenport TE, Randall K. Telehealth physical therapy in musculoskeletal practice. J Orthop Sports Phys Ther 2018 Oct;48(10):736-739. [doi: 10.2519/jospt.2018.0613] [Medline: 30270782]
9. Victorian Allied Health Workforce Research Project; Allied Health Assistance Workforce Report. State of Victoria, Department of Health and Human Services. 2016. URL: https://www2.health.vic.gov.au/Api/downloadmedia/%7BB0FE1403-FB25-4F2A-8FE5-4FECE377B257%7D [accessed 2021-03-31]
10. Cottrell MA, Russell TG. Telehealth for musculoskeletal physiotherapy. Musculoskelet Sci Pract 2020 Aug;48 [FREE Full text] [doi: 10.1016/j.msksp.2020.102193] [Medline: 32560876]
11. Bishop A, Gamlin J, Hall J, Hopper C, Foster NE. PhysioDirect: supporting physiotherapists to deliver telephone assessment and advice services within the context of a randomised trial. Physiotherapy 2013 Jun;99(2):113-118 [FREE Full text] [doi: 10.1016/j.physio.2012.08.002] [Medline: 23219629]
12. Cottrell MA, Hill AJ, O'Leary SP, Raymer ME, Russell TG. Clinicians' perspectives of a novel home-based multidisciplinary telehealth service for patients with chronic spinal pain. Int J Telerehabil 2018 Dec 11;10(2):81-88 [FREE Full text] [doi: 10.5195/ijt.2018.6249] [Medline: 30588279]
13. Lawford BJ, Delany C, Bennell KL, Hinman RS. "I Was Really Pleasantly Surprised": firsthand experience and shifts in physical therapist perceptions of telephone-delivered exercise therapy for knee osteoarthritis-a qualitative study. Arthritis Care Res (Hoboken) 2019 Apr 28;71(4):545-557. [doi: 10.1002/acr.23618] [Medline: 29885026]
14. Hinman RS, Lawford BJ, Bennell KL. Harnessing technology to deliver care by physical therapists for people with persistent joint pain: telephone and video‐conferencing service models. J Appl Behav Res 2018 Oct 30;24(2). [doi: 10.1111/jabr.12150]
15. Munro V, Morello A, Oster C, Redmond C, Vnuk A, Lennon S, et al. E-learning for self-management support: introducing blended learning for graduate students - a cohort study. BMC Med Educ 2018 Sep 24;18(1):219 [FREE Full text] [doi: 10.1186/s12909-018-1328-6] [Medline: 30249238]
16. Lawn S, Zhi X, Morello A. An integrative review of e-learning in the delivery of self-management support training for health professionals. BMC Med Educ 2017 Oct 10;17(1):183 [FREE Full text] [doi: 10.1186/s12909-017-1022-0] [Medline: 29017521]
17. Sinclair PM, Kable A, Levett-Jones T, Booth D. The effectiveness of Internet-based e-learning on clinician behaviour and patient outcomes: a systematic review. Int J Nurs Stud 2016 May;57:70-81. [doi: 10.1016/j.ijnurstu.2016.01.011] [Medline: 27045566]
18. Cook DA, Levinson AJ, Garside S, Dupras DM, Erwin PJ, Montori VM. Internet-based learning in the health professions: a meta-analysis. J Am Med Assoc 2008 Sep 10;300(10):1181-1196. [doi: 10.1001/jama.300.10.1181] [Medline: 18780847]
19. Gagnon K, Young B, Bachman T, Longbottom T, Severin R, Walker MJ. Doctor of physical therapy education in a hybrid learning environment: reimagining the possibilities and navigating a "New Normal". Phys Ther 2020 Aug 12;100(8):1268-1277 [FREE Full text] [doi: 10.1093/ptj/pzaa096] [Medline: 32424417]
20. Feng Y. National osteoarthritis strategy. National Osteoarthritis Strategy Project Group. 2018. URL: https://ibjr.sydney.edu.au/wp-content/uploads/2019/05/National-Osteoarthritis-Strategy.pdf [accessed 2021-03-31]
21. Vos T, Flaxman AD, Naghavi M, Lozano R, Michaud C, Ezzati M, et al. Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet 2012 Dec 15;380(9859):2163-2196 [FREE Full text] [doi: 10.1016/S0140-6736(12)61729-2] [Medline: 23245607]
22. Wallis JA, Taylor NF, Bunzli S, Shields N. Experience of living with knee osteoarthritis: a systematic review of qualitative studies. BMJ Open 2019 Sep 24;9(9) [FREE Full text] [doi: 10.1136/bmjopen-2019-030060] [Medline: 31551381]
23. Ackerman IN, Pratt C, Gorelik A, Liew D. Projected burden of osteoarthritis and rheumatoid arthritis in Australia: a population-level analysis. Arthritis Care Res (Hoboken) 2018 Jun 12;70(6):877-883. [doi: 10.1002/acr.23414] [Medline: 28898565]
24. Fransen M, McConnell S, Harmer AR, Van der Esch M, Simic M, Bennell KL. Exercise for osteoarthritis of the knee: a Cochrane systematic review. Br J Sports Med 2015 Dec 24;49(24):1554-1557. [doi: 10.1136/bjsports-2015-095424] [Medline: 26405113]

25. Uthman OA, van der Windt DA, Jordan JL, Dziedzic KS, Healey EL, Peat GM, et al. Exercise for lower limb osteoarthritis: systematic review incorporating trial sequential analysis and network meta-analysis. Br Med J 2013 Sep 20;347(sep20 1):5555 [FREE Full text] [doi: 10.1136/bmj.f5555] [Medline: 24055922]
26. Bannuru RR, Osani MC, Vaysbrot EE, Arden NK, Bennell K, Bierma-Zeinstra SM, et al. OARSI guidelines for the non-surgical management of knee, hip, and polyarticular osteoarthritis. Osteoarthr Cartil 2019 Nov;27(11):1578-1589 [FREE Full text] [doi: 10.1016/j.joca.2019.06.011] [Medline: 31278997]
27. Fernandes L, Hagen KB, Bijlsma JW, Andreassen O, Christensen P, Conaghan PG, European League Against Rheumatism (EULAR). EULAR recommendations for the non-pharmacological core management of hip and knee osteoarthritis. Ann Rheum Dis 2013 Jul;72(7):1125-1135 [FREE Full text] [doi: 10.1136/annrheumdis-2012-202745] [Medline: 23595142]
28. Kolasinski SL, Neogi T, Hochberg MC, Oatis C, Guyatt G, Block J, et al. 2019 American College of Rheumatology/Arthritis Foundation guideline for the management of osteoarthritis of the hand, hip, and knee. Arthritis Rheumatol 2020 Feb 06;72(2):220-233 [FREE Full text] [doi: 10.1002/art.41142] [Medline: 31908163]
29. National Clinical Guideline Centre (UK). Osteoarthritis: Care and Management in Adults. London, United Kingdom: London: National Institute for Health and Care Excellence (UK); 2014 Feb. National Institute for Health and Clinical Excellence: Guidance; 2014.
30. Guideline for the management of knee and hip osteoarthritis, 2nd edn. The Royal Australian College of General Practitioners. 2018. URL: https://www.racgp.org.au/clinical-resources/clinical-guidelines/key-racgp-guidelines/view-all-racgp-guidelines/knee-and-hip-osteoarthritis [accessed 2021-03-31]
31. Briggs AM, Hinman RS, Darlow B, Bennell KL, Leech M, Pizzari T, et al. Confidence and attitudes toward osteoarthritis care among the current and emerging health workforce: a multinational interprofessional study. ACR Open Rheumatol 2019 Jun 22;1(4):219-235 [FREE Full text] [doi: 10.1002/acr2.1032] [Medline: 31777798]
32. Briggs AM, Houlding E, Hinman R, Desmond L, Bennell K, Darlow B, et al. Health professionals and students encounter multi-level barriers to implementing high-value osteoarthritis care: a multi-national study. Osteoarthr Cartil 2019 May;27(5):788-804. [doi: 10.1016/j.joca.2018.12.024] [Medline: 30668988]
33. Egerton T, Diamond L, Buchbinder R, Bennell K, Slade S. A systematic review and evidence synthesis of qualitative studies to identify primary care clinicians' barriers and enablers to the management of osteoarthritis. Osteoarthr Cartil 2017 May;25(5):625-638 [FREE Full text] [doi: 10.1016/j.joca.2016.12.002] [Medline: 27939622]
34. Selten EM, Vriezekolk JE, Nijhof MW, Schers HJ, van der Meulen-Dilling RG, van der Laan WH, et al. Barriers impeding the use of non-pharmacological, non-surgical care in hip and knee osteoarthritis. J Clin Rheumatol 2017;23(8):405-410. [doi: 10.1097/rhu.0000000000000562]
35. Eyles JP, Hunter DJ, Bennell KL, Dziedzic KS, Hinman RS, van der Esch M, Joint Effort Initiative Members. Priorities for the effective implementation of osteoarthritis management programs: an OARSI international consensus exercise. Osteoarthr Cartil 2019 Sep;27(9):1270-1279 [FREE Full text] [doi: 10.1016/j.joca.2019.05.015] [Medline: 31163271]
36. Eyles JP, Hunter DJ, Briggs AM, Hinman RS, Fitzpatrick J, March L, et al. National osteoarthritis strategy brief report: living well with osteoarthritis. Aust J Gen Pract 2020 Jul 01;49(7):438-442. [doi: 10.31128/ajgp-08-19-5051-02]
37. National strategic action plan for arthritis. Australian Government, Department of Health. 2019. URL: https://www.health.gov.au/resources/publications/national-strategic-action-plan-for-arthritis [accessed 2021-03-31]
38. Bennell KL, Nelligan R, Dobson F, Rini C, Keefe F, Kasza J, et al. Effectiveness of an internet-delivered exercise and pain-coping skills training intervention for persons with chronic knee pain. Ann Intern Med 2017 Feb 21;166(7):453. [doi: 10.7326/m16-1714]
39. Hinman RS, Campbell PK, Lawford BJ, Briggs AM, Gale J, Bills C, et al. Does telephone-delivered exercise advice and support by physiotherapists improve pain and/or function in people with knee osteoarthritis? Telecare randomised controlled trial. Br J Sports Med 2020 Jul 20;54(13):790-797. [doi: 10.1136/bjsports-2019-101183] [Medline: 31748198]
40. Hinman RS, Nelligan RK, Bennell KL, Delany C. "Sounds a Bit Crazy, But It Was Almost More Personal:" A qualitative study of patient and clinician experiences of physical therapist-prescribed exercise for knee osteoarthritis via Skype. Arthritis Care Res (Hoboken) 2017 Dec 02;69(12):1834-1844. [doi: 10.1002/acr.23218] [Medline: 28217864]
41. Salisbury C, Montgomery AA, Hollinghurst S, Hopper C, Bishop A, Franchini A, et al. Effectiveness of PhysioDirect telephone assessment and advice services for patients with musculoskeletal problems: pragmatic randomised controlled trial. Br Med J 2013 Jan 29;346(jan29 3):43 [FREE Full text] [doi: 10.1136/bmj.f43] [Medline: 23360891]
42. Hinman RS, Kimp AJ, Campbell PK, Russell T, Foster NE, Kasza J, et al. Technology versus tradition: a non-inferiority trial comparing video to face-to-face consultations with a physiotherapist for people with knee osteoarthritis. Protocol for the PEAK randomised controlled trial. BMC Musculoskelet Disord 2020 Aug 07;21(1):522 [FREE Full text] [doi: 10.1186/s12891-020-03523-8] [Medline: 32767989]
43. Baxter P, Jack S. Qualitative case study methodology: study design and implementation for novice researchers. Qual Rep 2015 Jan 14. [doi: 10.46743/2160-3715/2008.1573]
44. O'Brien BC, Harris IB, Beckman TJ, Reed DA, Cook DA. Standards for reporting qualitative research. Acad Med 2014;89(9):1245-1251. [doi: 10.1097/acm.0000000000000388]
45. Miller GE. The assessment of clinical skills/competence/performance. Acad Med 1990 Sep;65(9 Suppl):S63-S67. [doi: 10.1097/00001888-199009000-00045] [Medline: 2400509]

46. PEAK training. The University of Melbourne, Melbourne School of Health Sciences. URL: https://healthsciences.unimelb.edu.au/departments/physiotherapy/chesm/clinician-resources/peak-training [accessed 2021-04-09]
47. Sekhon M, Cartwright M, Francis JJ. Acceptability of healthcare interventions: an overview of reviews and development of a theoretical framework. BMC Health Serv Res 2017 Jan 26;17(1):88 [FREE Full text] [doi: 10.1186/s12913-017-2031-8] [Medline: 28126032]
48. Morse JM, Field PA. Qualitative Research Methods for Health Professionals. Thousand Oaks, United States: SAGE Publications Inc; 1995:1-272.
49. Mays N, Pope C. Rigour and qualitative research. Br Med J 1995 Jul 08;311(6997):109-112 [FREE Full text] [doi: 10.1136/bmj.311.6997.109] [Medline: 7613363]
50. Regmi K, Jones L. A systematic review of the factors - enablers and barriers - affecting e-learning in health sciences education. BMC Med Educ 2020 Mar 30;20(1):91 [FREE Full text] [doi: 10.1186/s12909-020-02007-6] [Medline: 32228560]
51. Kolltveit BH, Gjengedal E, Graue M, Iversen MM, Thorne S, Kirkevold M. Conditions for success in introducing telemedicine in diabetes foot care: a qualitative inquiry. BMC Nurs 2017 Jan 13;16(1):2 [FREE Full text] [doi: 10.1186/s12912-017-0201-y] [Medline: 28100957]
52. Gardner P, Slater H, Jordan JE, Fary RE, Chua J, Briggs AM. Physiotherapy students' perspectives of online e-learning for interdisciplinary management of chronic health conditions: a qualitative study. BMC Med Educ 2016 Mar 16;16(1):62 [FREE Full text] [doi: 10.1186/s12909-016-0593-5] [Medline: 26879982]
53. Duque G, Demontiero O, Whereat S, Gunawardene P, Leung O, Webster P, et al. Evaluation of a blended learning model in geriatric medicine: a successful learning experience for medical students. Australas J Ageing 2013 Jun 17;32(2):103-109. [doi: 10.1111/j.1741-6612.2012.00620.x] [Medline: 23773249]
54. Nartker AJ, Stevens L, Shumays A, Kalowela M, Kisimbo D, Potter K. Increasing health worker capacity through distance learning: a comprehensive review of programmes in Tanzania. Hum Resour Health 2010 Dec 31;8(1):30 [FREE Full text] [doi: 10.1186/1478-4491-8-30] [Medline: 21194417]
55. Rowe M, Frantz J, Bozalek V. The role of blended learning in the clinical education of healthcare students: a systematic review. Med Teach 2012 Mar 28;34(4):216-221. [doi: 10.3109/0142159x.2012.642831]
56. Zolfaghari M, Negarandeh R, Eybpoosh S. Developing a blended learning program for nursing and midwifery students in Iran: process and preliminary outcomes. Iran J Nurs Midwifery Res 2013 Jan;18(1):20-26 [FREE Full text] [Medline: 23983723]
57. Major ME, Ramaekers SP, Engelbert RH, Van der Schaaf M. Preparing undergraduate students for clinical work in a complex environment: evaluation of an e-learning module on physiotherapy in the intensive care unit. BMC Med Educ 2020 Apr 28;20(1):130 [FREE Full text] [doi: 10.1186/s12909-020-02035-2] [Medline: 32345330]
58. Ohtake PJ, Lazarus M, Schillo R, Rosen M. Simulation experience enhances physical therapist student confidence in managing a patient in the critical care environment. Phys Ther 2013 Mar;93(2):216-228. [doi: 10.2522/ptj.20110463] [Medline: 23329555]
59. Thomas EM, Rybski MF, Apke TL, Kegelmeyer DA, Kloos AD. An acute interprofessional simulation experience for occupational and physical therapy students: key findings from a survey study. J Interprof Care 2017 May 28;31(3):317-324. [doi: 10.1080/13561820.2017.1280006] [Medline: 28276844]
60. Dreyfus SE, Dreyfus HL. A five stage model of the mental activities involved in directed skill acquisition. Department of Industrial Engineering and Operations Research University of California, Berkeley. 1980. URL: https://apps.dtic.mil/dtic/tr/fulltext/u2/a084551.pdf [accessed 2021-03-31]
61. Dreyfus SE. The five-stage model of adult skill acquisition. Bull Sci Technol Soc 2016 Jul 24;24(3):177-181. [doi: 10.1177/0270467604264992]
62. Hurley DA, Keogh A, Mc Ardle D, Hall AM, Richmond H, Guerin S, et al. Evaluation of an e-learning training program to support implementation of a group-based, theory-driven, self-management intervention for osteoarthritis and low-back pain: pre-post study. J Med Internet Res 2019 Mar 07;21(3) [FREE Full text] [doi: 10.2196/11123] [Medline: 30843863]
63. Smith AC, Armfield NR, Croll J, Gray LC. A review of Medicare expenditure in Australia for psychiatric consultations delivered in person and via videoconference. J Telemed Telecare 2012 Feb 23;18(3):169-171. [doi: 10.1258/jtt.2012.sft111]
64. Wade V, Soar J, Gray L. Uptake of telehealth services funded by Medicare in Australia. Aust Health Review 2014;38(5):528. [doi: 10.1071/ah14090]
65. Nelson R. Telemedicine and telehealth: the potential to improve rural access to care. Am J Nurs 2017 Jul;117(6):17-18. [doi: 10.1097/01.NAJ.0000520244.60138.1c] [Medline: 28541980]
66. Polinski JM, Barker T, Gagliano N, Sussman A, Brennan TA, Shrank WH. Patients' satisfaction with and preference for telehealth visits. J Gen Intern Med 2016 Mar 13;31(3):269-275 [FREE Full text] [doi: 10.1007/s11606-015-3489-x] [Medline: 26269131]
67. Shigekawa E, Fix M, Corbett G, Roby DH, Coffman J. The current state of telehealth evidence: a rapid review. Health Aff (Millwood) 2018 Dec;37(12):1975-1982. [doi: 10.1377/hlthaff.2018.05132] [Medline: 30633674]
68. COVID-19 National Health Plan – Primary Care Package – MBS telehealth services and increased practice incentive payments. Australian Government, Department of Health. 2020. URL: https://www.health.gov.au/resources/publications/

covid-19-national-health-plan-primary-care-package-mbs-telehealth-services-and-increased-practice-incentive-payments [accessed 2020-10-14]

69. Tack CT, Grodon J, Shorthouse F, Spahr N. "Physio anywhere": digitally-enhanced outpatient care as a legacy of coronavirus 2020. Physiotherapy 2021 Mar;110:26-28 [FREE Full text] [doi: 10.1016/j.physio.2020.07.004] [Medline: 33308826]

Abbreviations

ICT: information and communication technology
LMS: Learning Management System
OA: osteoarthritis
PEAK: Physiotherapy Exercise and Physical Activity for Knee Osteoarthritis
RCT: randomized controlled trial

Edited by R Kukafka; submitted 19.11.20; peer-reviewed by S Kardes, A Samuel; comments to author 29.12.20; revised version received 14.01.21; accepted 18.03.21; published 30.04.21.

Please cite as:
Jones SE, Campbell PK, Kimp AJ, Bennell K, Foster NE, Russell T, Hinman RS
Evaluation of a Novel e-Learning Program for Physiotherapists to Manage Knee Osteoarthritis via Telehealth: Qualitative Study Nested in the PEAK (Physiotherapy Exercise and Physical Activity for Knee Osteoarthritis) Randomized Controlled Trial
J Med Internet Res 2021;23(4):e25872
URL: https://www.jmir.org/2021/4/e25872
doi:10.2196/25872
PMID:33929326

Original Paper

Combining Web-Based Gamification and Physical Nudges With an App (MoveMore) to Promote Walking Breaks and Reduce Sedentary Behavior of Office Workers: Field Study

André Mamede[1], BSc, MSc; Gera Noordzij[2], BSc, MSc, PhD; Joran Jongerling[1], BSc, MSc, PhD; Merlijn Snijders[3], BSc, MSc; Astrid Schop-Etman[2], BSc, MSc, PhD; Semiha Denktas[1], BSc, MSc, PhD

[1]Department of Psychology, Education and Child Studies, Erasmus School of Social and Behavioural Sciences, Erasmus University Rotterdam, Rotterdam, Netherlands

[2]Erasmus University College, Erasmus School of Social and Behavioural Sciences, Erasmus University Rotterdam, Rotterdam, Netherlands

[3]Department of Public Health, Welfare and Care, Municipality of Rotterdam, Rotterdam, Netherlands

Corresponding Author:
André Mamede, BSc, MSc
Department of Psychology, Education and Child Studies
Erasmus School of Social and Behavioural Sciences
Erasmus University Rotterdam
Burgemeester Oudlaan 50
Rotterdam, 3062 PA
Netherlands
Phone: 31 10 408 8789
Email: mamedesoaresbraga@essb.eur.nl

Abstract

Background: Sedentary behavior (SB) and lack of physical activity (PA) have been associated with poorer health outcomes and are increasingly prevalent in individuals working in sedentary occupations such as office jobs. Gamification and nudges have attracted attention as promising strategies to promote changes in health behavior. However, most effectiveness studies thus far lacked active controls, and few studies have tested interventions combining these strategies.

Objective: This study investigates the effectiveness of combining a gamified digital app with physical nudges to increase PA and reduce SB in Dutch office workers.

Methods: Employees in the municipality of Rotterdam (N=298) from two office locations were randomized at the location level to either a 10-week intervention, combining a 5-week gamification phase encompassing a gamified digital app with social support features and a 5-week physical nudges phase, or to an active control (ie, basic digital app with self-monitoring and goal setting). The primary outcome was the daily step count, objectively measured via accelerometers. Secondary outcomes were self-reported PA and SB measured at baseline and at 5, 10, and 14 weeks. Mixed effects models were used to analyze the effects of the intervention on the outcome measures.

Results: A total of 78.5% (234/298) of participants completed the study and provided accelerometer data, whereas 36.9% (110/298) participants completed the self-report measures at 14 weeks. In the gamification phase, step count data were missing for 13.5% (473/3492) of observations in the control and 11.4% (445/3888) in the intervention condition; however, these percentages increased to 39.6% (1154/2910) and 59.6% (1932/3492) at follow-up, respectively. During the gamification phase, intervention participants increased their number of daily steps by 634 (95% CI 154.2-1113.8; *P*=.01) more than participants in the control group, after controlling for relevant factors. Improvements were not sustained during the physical nudges phase (*P*=.76) or follow-up (*P*=.88).

Conclusions: A digital intervention with gamification and social support features significantly increased the step count of office workers compared with an active control. Physical nudges in the workplace were insufficient to promote the maintenance of behavioral changes achieved in the gamification phase. Future research should explore the long-term effectiveness of similar gamified digital interventions.

Trial Registration: International Standard Randomized Controlled Trial Number (ISRCTN) 49129401; https://www.isrctn.com/ISRCTN14881571

(J Med Internet Res 2021;23(4):e19875) doi:10.2196/19875

KEYWORDS

internet; eHealth; mHealth; mobile phone; lifestyle; obesity; social network; multilevel analysis; physical exercise

Introduction

Background

Ample evidence has demonstrated that moderate-to-vigorous physical activity (PA) is associated with improved health outcomes [1]. However, increasing evidence has demonstrated that light forms of PA are associated with decreased risks of cardiovascular disease and all-cause mortality, even after adjusting for levels of moderate-to-vigorous PA [2]. A systematic review has shown that light PA, such as walking (ie, objectively measured step count), is associated with lower risk of obesity and diabetes type 2 [3], and studies have found associations between light PA and lower levels of depression [4,5], stress, and burnout [6]. Moreover, sedentary behavior (SB) has been linked to higher risks of all-cause mortality [7] and reduced life expectancy [7]. SB can be defined as any waking behavior with an energy expenditure ≤1.5 metabolic equivalent of task while in a sitting or reclining posture [8]. SB can be reduced by frequently interrupting the sitting time with light PA, such as walking breaks, which have been shown to reduce the health risks related to SB [9]. Thus, even for individuals meeting guidelines for moderate-to-vigorous PA, regular engagement in light PA is recommended to further reduce all-cause mortality and improve mental and physical health.

Despite these findings, a sedentary lifestyle is an escalating epidemic. Most common occupations have become increasingly sedentary because of technological advancements, and particularly for office workers, workplace sitting patterns are largely responsible for decreases in light PA and increases in SB [10,11]. One study showed that highly educated office workers in the Netherlands spend less time in light PA and more time in SB than workers in other occupations [12], and a recent report found that Dutch workers sit on average for 10 hours per week day [13]. Given that workplace sitting is the largest contributor to decreases in light PA [12] and increases in SB [14], behavior change interventions in this setting can bring considerable benefits at both the individual and societal levels, for instance, through the prevention of health care costs associated with noncommunicable diseases [15].

Influencing Health Behaviors: Behavior Change Theory and Nudging

Noncompliance with health behaviors can be largely attributed to lack of motivation, insufficient capacity to self-regulate toward one's goals [16-18], or environmental and policy factors that may limit opportunities for healthy behavior [19-21]. According to the Social Cognitive Theory by Bandura [22], several factors influence motivation for health behavior change, such as whether people are confident in their capacity to change (ie, self-efficacy), which depends on goal characteristics (eg, difficulty), or whether people have social support to change. Social support may involve modeling by family and friends; feedback and support from peers; or social incentives enhancing accountability, competition, and cooperation. People's capacity to self-regulate can also be influenced by social support and self-efficacy and their ability to self-monitor and their use of planning strategies and reminders [17,18,23]. Socioecological models emphasize the importance of environmental and policy factors for PA, such as walkability and esthetics of the environment, and social norms. Several behavioral change techniques (BCTs) can be used to influence motivational, self-regulatory, or environmental factors to promote behavior change.

Nudging in the Environment to Promote Motivation and Self-Regulation for Light PA

Certain BCTs attempt to motivate individuals to change by providing information on the risks of their current behavior or on the future benefits of behavior change. However, although these strategies can influence people's self-reported intention to change, they have, at best, a modest effect on behavior change [24]. Recently, increased attention has been paid to using insights from behavior change theory to help people make better choices by modifying their physical and social environment [25,26] (ie, *nudging*). Nudges are typically BCTs that exploit behavioral and cognitive tendencies to promote a desired behavior, and various interventions have successfully used nudges to stimulate healthy choices in a workplace setting. Motivational nudges can increase motivation for light PA by conveying information on the benefits of walking through an authority figure (eg, a doctor). Nudges can also enhance motivation for PA by influencing the social environment related to PA; for instance, by describing the social norms regarding that behavior or through role models [27,28]. A systematic review found that motivational sign nudges were effective in promoting stair climbing in various settings, including the workplace [29].

Other nudges can help promote the self-regulation of PA goals. For instance, point-of-choice prompts are cues that function by interrupting maladaptive habitual behaviors, such as prolonged sitting, and by highlighting opportunities in the environment to engage in alternative health-enhancing behaviors, such as walking breaks. Point-of-choice prompts have been shown to be effective in promoting stair climbing instead of escalator use [30] and walking, thereby reducing SB in the workplace [31]. Workplace nudges typically involve modifications in the physical environment, ranging from changes in the default positions of desks to reduce sitting to the use of motivational or point-of-choice prompt signs incorporating various BCTs to promote walking. However, although nudges are increasingly popular, partially because of their promising cost-effectiveness [32], the effect sizes of nudging interventions tend to be modest [33], and evidence for their effectiveness is still mixed [34,35]. One way to increase the effectiveness of nudges is innovation in intervention delivery, for instance, by including nudges in interactive digital apps [31,36] or by combining it with physical

nudges in the work environment, an approach that has been largely overlooked [37].

Gamification: Improving Digital Interventions to Promote Health Behavior Change

Given the growing use of technology, digital apps are promising avenues for delivering behavior change interventions. Digital interventions provide an empirically supported, convenient, and potentially more cost-effective alternative for reaching large proportions of the public over long periods [38,39]. However, digital interventions still depend on active user engagement to promote behavior change, which is challenging to maintain. Recently, gamification has emerged as a promising persuasive strategy to increase users' engagement, motivation, and social interaction in digital behavior change interventions [40]. Gamification is an *umbrella term* that refers to the use of game design elements in a nongaming context [41,42]. Gamified digital intervention can flexibly implement a wide range of BCTs, including nudges, such as educational strategies, social support, social comparison, self-monitoring, goal setting, rewards (eg, badges), and personalized feedback, all of which have been associated with greater behavioral change [23,43,44]. Besides promoting self-regulation and motivation for the initiation and maintenance of PA [38,45], gamification can enhance social support and social comparison through competition, cooperation, and salient visualization of others' behavior (eg, leaderboards) [45,46]. Two recent studies found that digital interventions with elements such as gamification, social support, and social comparison increased light PA in office workers and promoted adequate engagement and adherence with the digital app [46,47].

Despite the promising potential of gamification, recent reviews of gamified digital interventions have highlighted lack of empirical studies comparing gamified digital interventions with active controls (ie, nongamified digital interventions) [45,48]. Although multiple BCTs and nudges can be flexibly incorporated in gamified digital interventions, the effectiveness of such interventions could still be further enhanced through complementary strategies that engage participants outside of the virtual environment. For instance, physical nudges in the workplace, such as *motivational* and *point-of-choice prompts* sign *nudges*, are easy to implement and could serve as a cost-effective way to improve maintenance of the initial behavior change promoted through gamified digital interventions. However, research is needed to explore whether these types of physical nudges could complement and increase the effectiveness of gamification to promote behavior change.

This Study: MoveMore

We conducted a cluster randomized controlled trial (RCT) to evaluate the effects of *MoveMore*, a 10-week multicomponent intervention, on the PA and SB of office workers. The MoveMore intervention consisted of a 5-week gamification phase that included a commercially available gamified digital app incorporating several BCTs and nudges, such as social support and social comparison, followed by a physical nudges phase for the last 5 weeks, in which physical *motivational* and *point-of-choice prompt nudges* were introduced in the workplace. Intervention effects were compared with an active control encompassing a basic version of the digital app. We hypothesized that during the gamification phase, participants in the intervention condition would increase their levels of objectively measured light PA (ie, daily step count), compared with the control. Similarly, we hypothesized that during the gamification phase, we would observe increases in self-reported light PA, increases in self-reported moderate-to-vigorous PA, and reductions in SB in participants in the intervention condition compared with the control. We expected that improvements achieved during the gamification phase would be maintained during the physical nudges phase and at a 1-month follow-up.

Methods

Study Design

To evaluate the effects of the MoveMore intervention, a 2-arm cluster RCT was conducted at 2 office locations in the municipality of Rotterdam, the second largest city in the Netherlands. Each office location was randomly allocated to either the control or the intervention condition to minimize treatment contamination. The study protocol was approved by the Ethics Review Committee of the Department of Psychology, Education and Child Studies, Erasmus University Rotterdam (application number 18-039). The study was registered in the International Standard Randomized Controlled Trial Number Register (ISRCTN 49129401). This study is conducted and described according to the CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth) checklist (Multimedia Appendix 1) [49].

Study Setting and Population

Participants were office workers (N=298) from 2 government workplaces in the city center of Rotterdam (office locations A and B). Office location A consisted of a tall office building, with 44 floors accommodating city management and urban development departments, whereas office B consisted of a wide office building, with 5 floors accommodating social development departments. In both locations, government employees belonged to several different occupational groups, including managers, administrative workers, and blue-collar workers. In this field study, it was only possible to recruit departments located in 2 office buildings in the municipality of Rotterdam. Given the limited number of departments involved in this study, our sample size was limited by the number of employees in these departments who were willing and eligible to participate in the research. Considering that the feasible sample size of this field study was larger than that of multiple other PA intervention trials in office workers with similar methodologies [46,50,51], our feasible sample size was considered sufficiently adequate to investigate the effects of this intervention.

Eligibility Criteria

Given that several components of the intervention were in Dutch, only individuals fluent in the Dutch language were eligible to participate. Additional eligibility criteria included working in a department that was not involved in another PA-related intervention, access to a smartphone capable of

running the required digital app, and provision of written informed consent for participation.

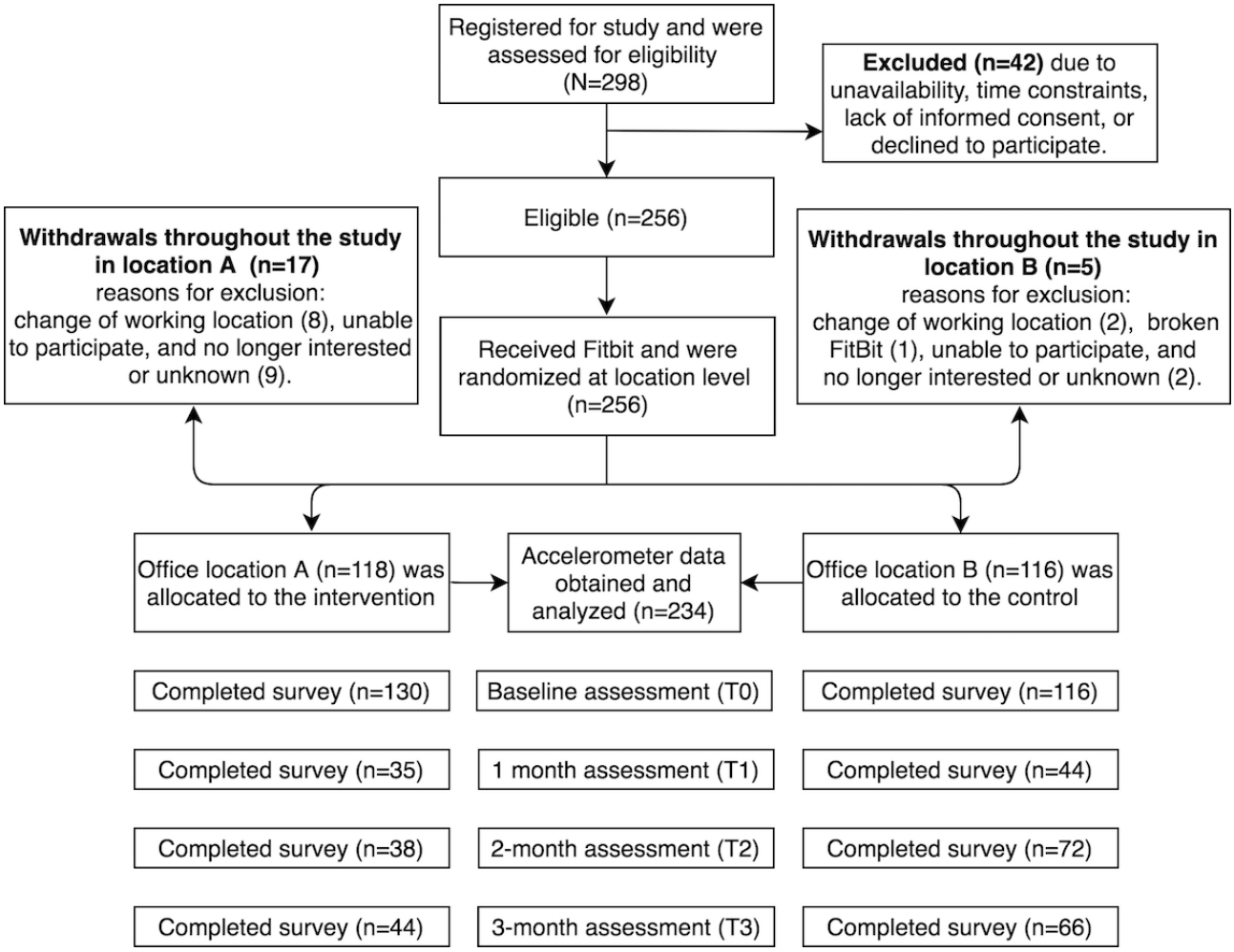
Recruitment

Potential participants were invited via email and social media and through their department team leaders to participate in a study on PA in office workers. Once approximately 150 office workers from each location responded, the invitation was closed. In total, 125 office workers from location A were included in the intervention arm, and 131 office workers from location B were included in the control arm. Participants were enrolled in October 2018 and were followed until February 2019.

Procedure

Participants were invited to attend an information session held by a study representative from the municipality, during which potential participants were screened for eligibility criteria, and participants' baseline measurements and informed consent were collected. Participants received written and verbal explanations of the intervention requirements before providing their consent. Subsequently, participants received a wrist-worn, triaxial accelerometer device (Fitbit Flex) to monitor step count [52] and were shown how to use it in combination with a digital app installed on their mobile phones during the session. The app was available for both iOS and Android operating system and was accessible through a website. Participants authorized their data to be captured for the study. Participants were told that the digital app was intended to support them in becoming more active and that they should use it throughout the day to help them increase their PA. Participants received subsequent questionnaires (Figure 1) via email during the interventions, at 5 weeks after baseline (T1), at 10 weeks after baseline (T2; postintervention), and at 14 weeks after baseline (T3; follow-up). Figure 1 illustrates the flow of participants throughout the study.

Figure 1. CONSORT (Consolidated Standards of Reporting Trials) diagram displaying the flow of participants throughout the study.



Intervention and Control

Participants in the control location were given a basic version of the app (Figure 2), whereas those in the location receiving the MoveMore intervention were given the full version of the digital app, which included additional features (Figure 3). Both versions of the app were linked to the accelerometer and provided users with a default daily step count goal. Participants were instructed on how to monitor their own daily step count and how to set more challenging daily step goals for themselves (Figure 2). The basic features available in both apps also included weekly personalized feedback to participants via email. In the first 5 weeks of the MoveMore intervention condition (ie, gamification phase), office workers were invited to participate in PA challenges through the digital app, which incorporated elements of gamification and social support and social comparison features. After the gamification phase, physical nudges were introduced to the workplace of participants

in the MoveMore intervention for another 5 weeks (ie, physical nudges phase). As one of this study's aims is to investigate whether physical nudges could promote maintenance of improvements in light PA achieved through the gamification phase, the order of the different study phases was not randomized. An overview of the study and intervention design is shown in Figure 4. The exact BCTs used in the MoveMore intervention and in the active control are described in the subsequent sections and in Table 1 according to the BCT version 1 taxonomy developed by Michie et al [53].

Figure 2. Screenshots of pages available in the basic version of the app used as active control.

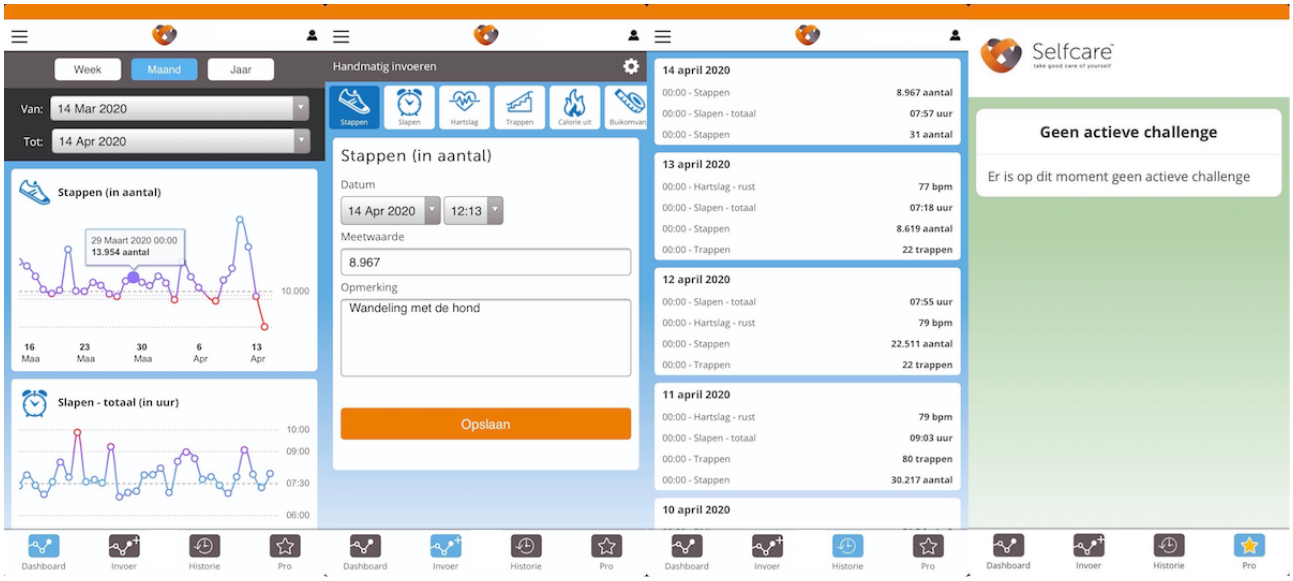


Figure 3. Screenshots of the additional features encompassed in the challenges offered to participants in the MoveMore intervention through the full gamified app.

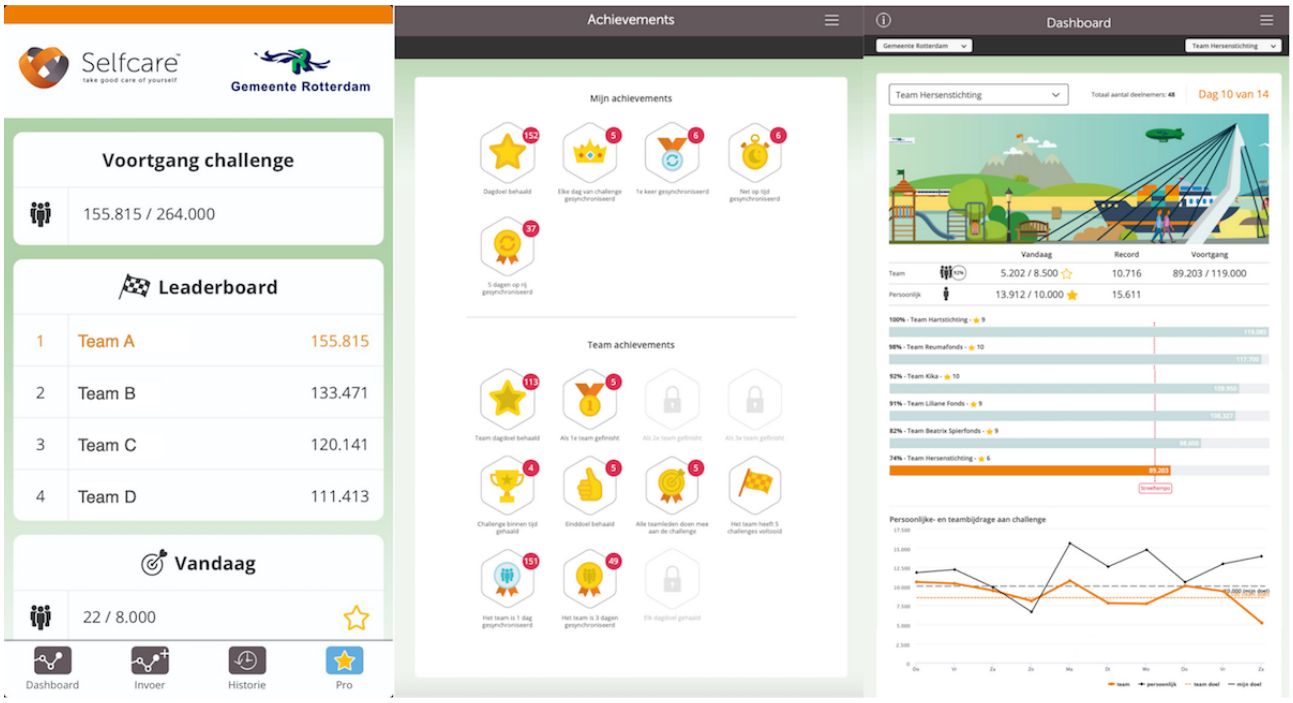


Figure 4. Illustration of study and intervention design. T0 to T3 represent the measuring moments.

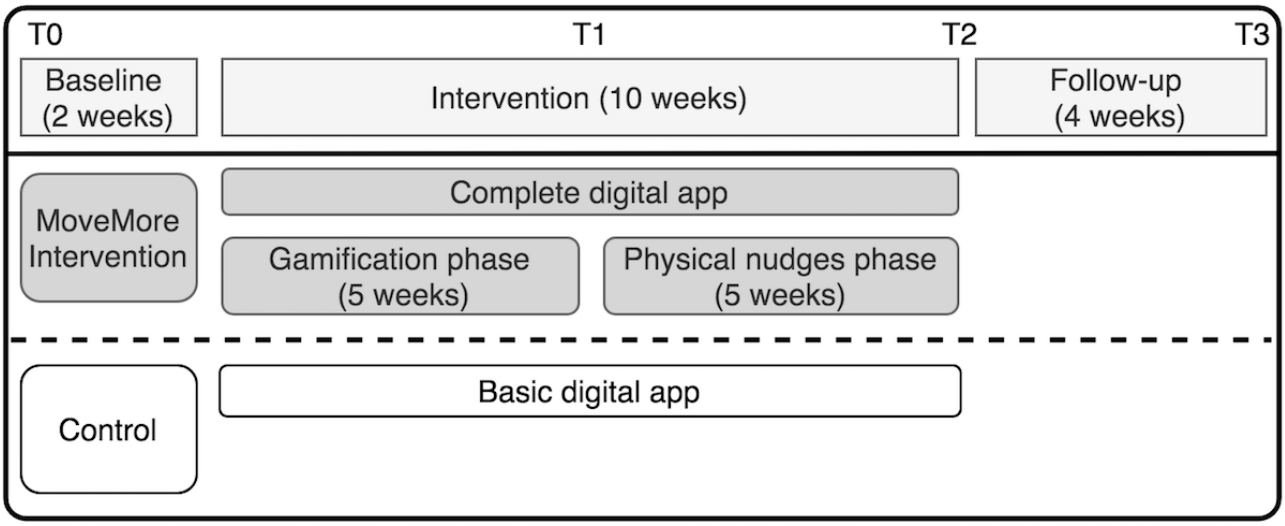


Table 1. Behavioral change techniques used in the gamified digital app and the physical nudges (MoveMore intervention) or in the basic digital app (control).

Behavioral components	MoveMore intervention		Control; basic digital app
	Gamified digital app	Physical nudges	
Self-monitoring (ie, accelerometer)	✓[a]	N/A[b]	✓
Information about health consequences	✓	✓	✓
Information about emotional consequences	✓	✓	✓
Self-monitoring	✓	N/A	✓
Goal setting (behavior)	✓	N/A	✓
Discrepancy between current behavior and goal	✓	N/A	✓
Personalized feedback on individual progress	✓	N/A	✓
Personalized feedback on team progress	✓	N/A	N/A
Graded tasks	✓	N/A	N/A
Reward (outcome)	✓	N/A	N/A
Social support	✓	N/A	N/A
Restructuring of the social environment	✓	N/A	N/A
Social comparison	✓	✓	N/A
Prompts and cues	✓	✓	✓
Present information from a credible source in favor of the desired behavior	N/A	✓	N/A

[a]Behavior change techniques applied through the gamified digital app, physical nudges, or basic digital app.

[b]N/A: not applicable.

Control Condition

The basic app used in the control condition allowed participants to self-monitor and to set their own daily step goal (Figure 2). The basic app gave participants a default daily step goal of 10,000 steps, which remained the same throughout the study. The basic app also provided participants with weekly personalized feedback on their step count progress via email (Multimedia Appendix 2). This basic app served as an active control because it allowed for an objective assessment of light PA, and its components (ie, self-monitoring, goal setting, and personalized feedback) are effective in promoting PA [54]. The same features provided in the basic apps were available to participants in the MoveMore intervention using the full version of the digital app. The gamified challenges provided to those in the MoveMore intervention were disabled in the basic app provided to those in the control condition (Figure 2).

Gamification Phase

In addition to the features included in the basic digital app, the full digital app allowed participants in the MoveMore intervention to participate in team walking challenges (Figure 3). During the first 5 weeks, office workers were invited to participate in 2 challenges, each lasting 2 weeks, with 1 week

in between them. To increase participants' light PA, the challenges incorporated elements of gamification and several BCTs to promote motivation and self-regulation for walking. During the challenge, participants in the MoveMore intervention were allocated to different teams (20-30 subjects), according to the department they worked in. The challenges consisted of a *virtual walking tour* (eg, a roundtrip across Rotterdam) representing a large goal that participants could achieve by attaining their daily goal for 2 weeks. In addition to progressing toward their daily step goals, participants' daily steps contributed toward their team step goal (ie, set as the number of participants in the team multiplied by their default daily step goal). A leaderboard served to enhance intrateam cooperation and individual accountability while promoting competition between teams (Figure 4). To enhance team identity and motivation, each team was allocated as a representative of a different charity. By earning points and climbing the leaderboard ranks, teams could win gradually bigger prizes for their charity, sponsored by the municipality. The first team earned €100 (US $120), the second team earned €90 (US $108), and so on, with the sixth and last team earning €50 (US $60). These components (ie, teams, leaderboards, and charity representations) were included to restructure the social environment regarding PA, thereby promoting social support and social comparison for walking in the workplace.

In addition to changes in social factors, the game elements of the app were also supposed to motivate participants to walk while increasing their ability to self-regulate their behavior. Similarly, in a game, the challenges started easy and became increasingly more difficult (ie, graded task) to enhance self-efficacy and, therefore, motivation for PA. The default goal for the first challenge was set at 8500 steps, which is easier than the default goal used in the basic version of the app provided to those in the control (ie, 10,000 steps), whereas for the second challenge, participants in the MoveMore intervention were encouraged to reach this more difficult default goal of 10,000 daily steps. Participants could also set more challenging daily step goals. The app rewarded participants with virtual awards for certain individual- (eg, "Daily step goal achieved!") and team-based achievements (eg, "Your team completed a challenge!"). On the website, participants could access detailed information on their achievements and on their progress with the team challenge, which was illustrated by their virtual avatars crossing the virtual tour scenarios (Figure 3). To promote self-regulation during the challenges, in addition to the weekly feedback on their personal step goals that is provided by the basic digital app, participants in the MoveMore intervention received biweekly newsletters during the challenges via email with updates on the competition and their team's progress (Multimedia Appendix 2).

Physical Nudges

After the gamification phase, physical nudges were introduced in the office workspace of participants in the intervention condition for 5 weeks to promote maintenance of behavior change. These nudges consisted of table signs aiming to (1) further motivate participants to engage in PA and reduce SB and (2) remind participants of the opportunities for PA in their work environment and routine. To achieve the former, motivational nudges incorporating several different behavioral insights were implemented. For example, 2 table sign posters portrayed an interaction between the office physician and an employee and presented the following messages from the physician: "walking breaks are healthy and increase work productivity!" or "My advice: Stand up every half an hour and move a little!" Another type of motivational nudge used social comparison to increase motivation for PA, with the following message: "Half of your colleagues try to move at least 10000 steps per day. What about you?" Motivational nudges were placed in visible locations (eg, on top of tables and eye-level closets) in office spaces and open areas of the intervention location. Complementarily, another type of nudge, namely, point-of-choice prompts, reminded participants of their PA goals, highlighting opportunities for PA in a timely manner and prompting cognitive and behavioral rehearsal. For instance, 2 point-of-choice prompt nudges were placed in the coffee and lunch areas of the workspace with the messages "Grabbing a drink? Perfect moment to be healthy and go for a walking break!" and "Lunch time? Perfect moment to move!" The messages reported have been translated from Dutch (Multimedia Appendix 3). The office workers participating in the intervention were spread across 22 floors. Approximately 5 table sign nudges, including at least one point-of-choice prompt, were placed on every floor in which participants of the intervention worked, for a total of over 110 nudges spread throughout the office building.

Measures

Demographics and Other Variables

The demographic information collected during baseline included participants' age, gender, weight, length, BMI (ie, calculated from self-reported length and weight), nationality, migrant background (ie, parental nationality), highest educational attainment, occupation in the municipality, weekly number of working days, and working hours.

Primary Outcome Measure

The primary outcome measure of walking behavior was the number of daily steps objectively measured using Fitbit Flex accelerometers (objective light PA). Previous studies have determined that the Fitbit Flex accelerometer has acceptable reliability and validity for step count measurements [52].

Secondary Outcome Measures

Secondary outcome measures included self-report measures of work time light PA, moderate-to-vigorous PA, and SB. SB at work was assessed using 2-item self-report measures of workplace sitting time and breaks in the sitting time. The self-report measures for assessing the duration of SB (Pearson r=0.44, 95% CI 0.24-0.60) and the frequency of breaks from sitting (Spearman r=0.26, 95% CI 0.11-0.44) were positively correlated with accelerometer measurements in a sample of desk workers [55]. The item on frequency of breaks, "During a typical work day how many breaks from sitting (such as standing up, or stretching or taking a short walk) during one hour of sitting would you take at work?," and the item on duration of SB at work, "Please estimate the total time during the last week that you spent sitting down as part of your job while at work or

working from home," were translated into Dutch and assessed for face validity. In our sample, the intraclass correlation coefficients (ICCs) for different measurements of SB duration and SB break frequency during work were 0.44 and 0.60, respectively, indicating poor to moderate test-retest reliability. To assess the intensity and levels of PA in various settings (ie, at work, at home, during active transport), the validated Dutch version of the Short Questionnaire for Assessing Health enhancing physical activity (SQUASH) was used [56,57]. The test-retest reliability of the SQUASH items was poor for assessing hours per week spent in light PA (ICC=0.35) and moderate for items assessing moderate-to-vigorous PA at work (ICC=0.60) and number of days or weeks engaging in at least 30 minutes of moderate-to-vigorous PA (ICC=0.55). Thus, the suboptimal test-retest reliability of some of our self-report measures may have hindered the assessment of intervention effects on secondary outcomes.

Data Management, Monitoring, and Safety

Except for the baseline, questionnaires were all administered electronically using the web survey platform Qualtrics [58]. Fitbit Flex accelerometer data were obtained through the company responsible for the gamified digital intervention [59] via the Fitbit app and were downloaded at the completion of the follow-up period. Data were exported into R statistical software version 3.5.2 and analyzed using the R package *lmer* [60]. Hardcopy consent forms were stored in locked filing cabinets, and electronic data were stored on password-protected drives accessible by study investigators.

Data Analysis

In this study, following a period of 2 weeks of baseline measurement, participants' daily number of steps (ie, primary outcome) was measured continuously for 14 consecutive weeks, resulting in a hierarchical data structure. Daily step count observations (level 1) were nested within participants (level 2), who, in turn, were nested within the departments (level 3). Recent statistical studies simulating variance in longitudinal data have shown that misspecification of the number of levels can lead to biased findings [61]. Therefore, we used a mixed effects model to account for the nested hierarchical structure of the data by including random intercepts for the different levels when the variance at that level was significantly different from zero [62]. As recommended by Haan-Rietdijk et al [61], we used autoregressive models (in combination with Akaike Information Criteria scores) to assess the variance at different levels and determine the levels needed to be included in the model. Given that the variance at the department level was not significant (χ^2_1=0; P=.99), models were only adjusted for the clustering of observations within participants (ie, level 2; χ^2_1=4934.5; P<.001). As we were interested in comparing daily step counts during the different study phases (ie, gamification, physical nudges, and follow-up) with baseline and in investigating potential interactions between intervention conditions and different phases, the study phase was not considered a level but rather included as a predictor in our models.

Considering that multilevel models can handle data missing at random, no missing data imputation was performed, and partially completed records were included in the model to avoid biases associated with a completer-only analysis [63]. In the primary analysis, the first 5 days of data were ignored when estimating the baseline step count to diminish the potential upward bias from estimating higher activity during initial accelerometer use. Observations with less than 1000 steps and more than 60,000 steps were considered missing because evidence indicates that these values are unlikely to represent actual activity [64-66]. Such observations were considered either extreme outliers or the result of forgetting to wear the accelerometer.

To assess the interaction between intervention condition and time and study phases, a repeated measures mixed effects model was employed following the intention-to-treat principles. Models were initially fit with a random intercept for participants, fixed effects of time, study phase (ie, baseline, gamification, nudging, and follow-up), and covariates. When the relationship between our outcome variable and time was quadratic or cubic, quadratic and cubic parameters of time were included in the model as fixed effects. To avoid convergence issues in the primary analysis, time was rescaled to represent 2-week intervals. Covariates initially included in the model were age, sex, parental nationality, work occupation, number of weekly working hours, education, and BMI (ie, calculated from self-reported weight and length). Covariates that were not significant predictors of outcome variables were excluded from the model. Next, the level 2 variable intervention condition (ie, intervention vs control) was included as a fixed effect. Random slopes of the study phase and time per participant were added to the model. When the random slope of the study phase per participant was significant, a 2-way cross-level interaction between the study phase and the intervention condition was included in the final model to investigate the effects of the intervention during each phase. In addition, in an exploratory analysis, we investigated whether intervention effects were influenced by individual differences by examining interactions between intervention effects and relevant covariates. The model used in the primary analysis was refit using secondary outcome measures: (1) the mean number of hours spent in light PA and moderate-to-vigorous PA during work and the number of days engaging in sufficient amount of moderate-to-vigorous PA, as assessed by the SQUASH questionnaire, and (2) 2 self-reported items assessing SB: the average number of sitting breaks taken per hour during work and the mean daily sitting time during work.

Results

Demographic Statistics

Table 2 presents baseline descriptive statistics of the study sample per intervention condition. Relative to the control condition, the intervention condition had a significantly higher proportion of male participants and participants of lower educational backgrounds. In addition, participants in the intervention condition weighted significantly more and had significantly higher BMI than participants in the control

condition. Participants in the intervention group also logged a lower number of daily steps at baseline, although this difference did not reach statistical significance.

Table 2. Sociodemographic and behavioral characteristics of participants at baseline.

Variable	Intervention (n=118)	Control (n=116)	*P* value
Behavioral characteristics			
Number of daily steps, mean (SD)	10,138 (4643.5)	10,403 (4191.6)	.22
Meeting physical activity guidelines (days per week)[a], mean (SD)	5.1 (2.0)	5.4 (1.0)	.20
Hour sitting per week[b], mean (SD)	30.1 (9.5)	29.2 (10.2)	.45
Breaks per hour[b], mean (SD)	1.8 (1.2)	1.9 (1.3)	.49
Demographic characteristics			
Age (years), mean (SD)	47.5 (9.6)	45.9 (10.2)	.25
Gender (female), n (%)	63 (55.3)	83 (72.8)	*.02* [c]
Weight (kg), mean (SD)	82 (17.8)	75.7 (13.9)	*.003*
BMI[d], mean (SD)	26.9 (5.0)	25.6 (4.5)	*.04*
Nationality (Dutch), n (%)	101 (89.4)	104 (92.0)	.57
Parental nationality (Dutch), n (%)	88 (77.9)	89 (78.1)	.63
Education (higher education), n (%)	78 (69.6)	99 (86.1)	*.005*
Work position (highly skilled)[e], n (%)	105 (97.2)	110 (96.5)	.99
Number of weekly working days, mean (SD)	4.4 (0.6)	4.3 (0.5)	.16

[a]Meeting daily physical activity guidelines was defined as self-reported engagement in at least 30 minutes of moderate-to-vigorous physical activity per day.

[b]The number of hours sitting per week and the number of sitting breaks per hour refer specifically to sedentary behavior during work time.

[c]Italics indicates statistical significance ($P<.05$).

[d]Calculated from self-reported height and weight.

[e]Nonmanual labor occupations, such as managers and administrative positions, were coded as highly skilled.

Primary Analysis: Daily Step Count

After controlling for relevant covariates and subject-specific differences, our mixed effects model investigated the effects of the intervention condition, time, study phase, and the interaction between study phase and intervention condition on the objectively measured step count of participants. The step count data included in the model were recorded for 109 days from baseline to follow-up. Participants, on average, wore their accelerometers and recorded at least 1000 daily steps for approximately 75 (68.8%; SD 27.8) days throughout the study. During the gamification phase, step data that were missing or had values less than 1000 steps per day represented 13.5% (473/3492) of observations for the control arm and 11.4% (445/3888) for the intervention arm. During the follow-up period, these percentages increased to 39.6% (1154/2910) in the control arm and 59.6% (1932/3240) in the intervention arm, indicating substantial missing data in our sample during the later study phases.

The mean number of daily steps of participants in each condition across the study phases is shown in Table 3 and Figure 5. The first repeated measures mixed model analysis (model 1; Table 3) included the effects of relevant covariates, study phase, time, intervention, and random slopes of time and study phase per participant. Model 1 revealed that the daily number of steps was negatively associated with BMI ($B=-178.03$; SE 38.94; $t_{168.61}=-4.57$; $P<.001$) and positively associated with age ($B=35.85$; SE 18.89; $t_{172.50}=1.90$; $P=.06$) in two-tailed t tests. These 2 predictors explained 12.6% of the variance at the participant level ($R^2=0.126$), with larger BMI and younger age being associated with lower daily step counts overall. The fixed effects of gender, work occupation, number of working hours, education, and nationality (both individual and parental) were removed from the model, as they were not significantly related to step count.

Table 3. Means of primary and secondary outcome variables across study phases.

Variable	Baseline (intervention: n=130; control: n=116)	Gamification (intervention: n=35; control: n=44)	Physical nudges (intervention: n=38; control: n=72)	Follow-up (intervention: n=44; control: n=66)
Intervention location, mean (SD)				
Number of daily steps[a]	10,138.3 (4643.5)	10,901.8 (5068.3)	9873.5 (5020.8)	10,481.1 (5035.9)
Meeting PA[b] guidelines[c]	5.2 (2.0)	5.8 (1.6)	5.8 (1.4)	5.6 (1.7)
Hours in light PA	26.1 (13.9)	28.4 (12.2)	22.7 (14.4)	20.7 (14.3)
Hours in moderate-to-vigorous PA	0.35 (2.2)	0.05 (0.23)	0.08 (0.41)	0.71 (2.24)
Hours sitting per week[d]	30.9 (10.7)	32.6 (10.6)	30.4 (9.8)	30.0 (6.8)
Number of breaks per hour[d]	1.7 (1.2)	1.5 (0.8)	1.6 (0.9)	1.4 (0.7)
Control location, mean (SD)				
Number of daily steps[a]	10,403.0 (4191.6)	10,618.6 (4377.3)	10,138.5 (4820.7)	10,279.3 (4387.8)
Meeting PA guidelines (days per week)[c]	5.4 (1.9)	5.7 (1.6)	5.8 (1.6)	5.7 (1.7)
Hours in light PA	26.4 (13.1)	23.2 (14.1)	23.1 (14.3)	24.7 (13.25)
Hours in moderate-to-vigorous PA	0.16 (0.75)	0.22 (0.71)	0.17 (0.81)	0.73 (2.66)
Hours sitting per week[d]	29.1 (10.2)	28.9 (8.2)	31.6 (10.7)	29.9 (10.7)
Number of breaks per hour[d]	2.0 (1.3)	1.6 (1.0)	1.6 (0.8)	1.6 (0.8)

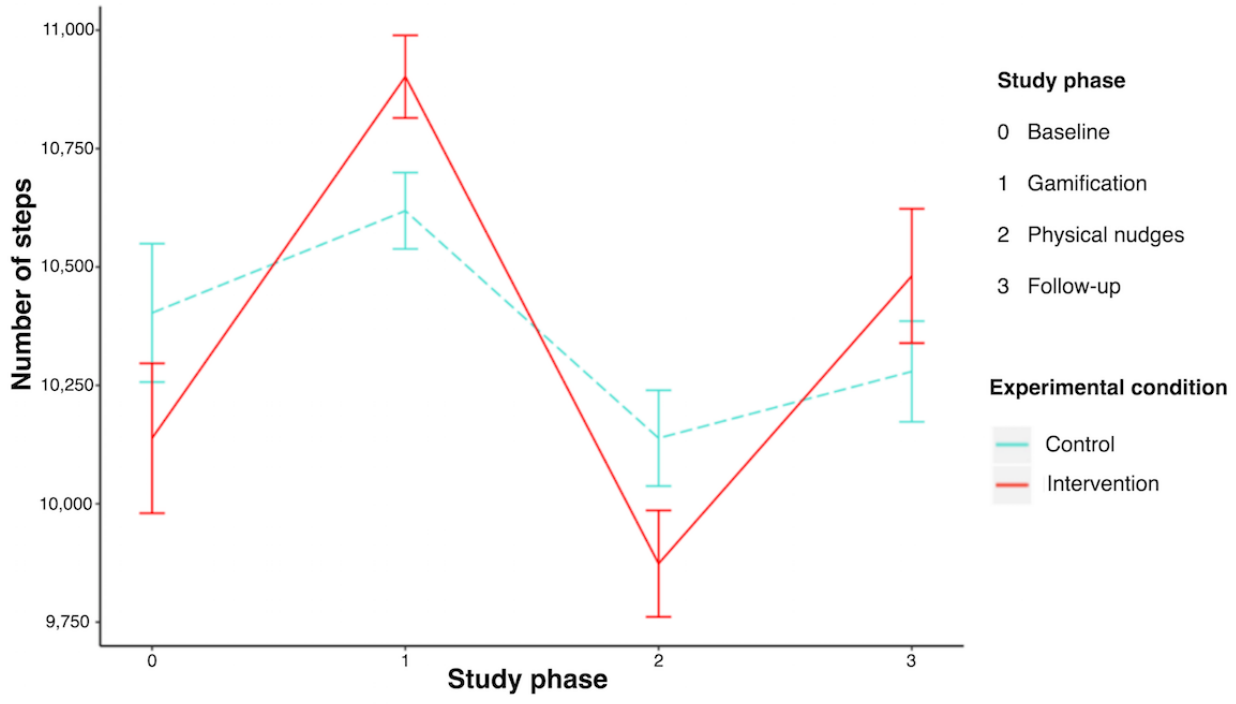
[a]The sample size shown refers to questionnaire assessments and does not apply to the number of daily steps measured using accelerometers. In total, 234 participants completed the accelerometer measurements, and on average, the accelerometer was worn on 68.8% of the days.

[b]PA: physical activity.

[c]Meeting daily physical activity guidelines was defined as self-reported engagement in at least 30 minutes of moderate-to-vigorous physical activity per day.

[d]The number of hours sitting per week and number of sitting breaks per hour refers specifically to sedentary behavior during work time.

Figure 5. Unadjusted differences in average daily step counts between intervention and control conditions across study phases.



Given the initial novelty of the gamification and physical nudges phase, which potentially wears off toward the end of each phase, it is plausible that daily step counts oscillated significantly within each study phase. In support of this interpretation, our

analysis revealed that linear, quadratic, and cubic effects of time were significant, suggesting that changes in step count within a study phase oscillated across time. Within-subject variability was partially modeled by the fixed effects of time and study phase, whereas the remaining unexplained variance was accounted for by the level 1 error term ε_{ij}. Model 1 revealed significant effects of the study phases on step count; however, the main effect of the intervention condition was not significant (model 1; Table 4). These effects of study phases reflect that, after controlling for the effects of time and other covariates, participants from both groups increased their number of daily steps during the novel gamification phase and decreased their number of daily steps during the nudging and follow-up phases, compared with baseline.

Table 4. Results of models with and without cross-level interactions predicting number of daily steps.

Parameter	Model 1[a]		Model 2[b]	
	Estimate (SE)	*P* value	Estimate (SE)	*P* value
Intercept	14,003.0 (1233.7)	N/A[c]	14,063.9 (1240.4)	N/A
BMI[d]	−178.0 (38.9)	*<.001* [e]	−177.1 (39.02)	*<.001*
Age	35.9 (18.9)	.06	36.0 (18.92)	.06
Time[f]	−1261.9 (308.0)	*<.001*	−1259.60 (307.95)	*<.001*
Time[g]	332.5 (81.4)	*<.001*	332.4 (81.36)	*<.001*
Time[h]	−21.2 (6.1)	*<.001*	−21.2 (6.08)	*<.001*
Gamification	856.9 (211.4)	*<.001*	541.9 (241.7)	*.03*
Nudging	−690.9 (326.6)	*.04*	−751.5 (359.0)	*.04*
Follow-up	−1475.3 (406.7)	*<.001*	−1520.2 (441.2)	*<.001*
Intervention	−16.0 (374.8)	.97	−222.3 (416.4)	.59
Intervention × gamification	N/A	N/A	634.0 (244.8)	*.005*
Intervention × nudging	N/A	N/A	98.2 (325.5)	.76
Intervention × follow-up	N/A	N/A	53.49 (381.7)	.89

[a]Model 1: $\gamma_{ij}=\beta 0_j + \beta 1\text{BMI}_j + \beta 2\text{Age}_j + \beta 3\text{Time}_{ij} + \beta 4\text{Time}_{ij}^2 + \beta 5\text{Time}_{ij}^3 + \beta 6\text{Gamification}_{ij} + \beta 7\text{Nudging}_{ij} + \beta 8\text{Follow-up}_{ij} + \beta 9\text{Intervention}_j + \varepsilon_{ij} + \mu_{ij}$. Model 1 (Akaike Information Criteria: 269682) refers to the model without cross-level interactions between the intervention and study phases.

[b]Model 2: $\gamma_{ij}=\beta 0_j + \beta 1\text{BMI}_j + \beta 2\text{Age}_j + \beta 3\text{Time}_{ij} + \beta 4\text{Time}_{ij}^2 + \beta 5\text{Time}_{ij}^3 + \beta 6\text{Gamification}_{ij} + \beta 7\text{Nudging}_{ij} + \beta 8\text{Follow-up}_{ij} + \beta 9\text{Intervention}_j + \beta 10\text{Intervention}_j\text{Gamification}_{ij} + \beta 11\text{Intervention}_j\text{Nudging}_{ij} + \beta 12\text{Intervention}_j\text{Follow-up}_{ij} + \varepsilon_{ij} + \mu_{0j} + \mu_{1j}\,\text{Gamification}_{ij} + \mu_{2j}\,\text{Nudging}_{ij} + \mu_{3j}\,\text{Follow-up}_{ij}$. Model 2 (Akaike Information Criteria: 269679) refers to the model with cross-level interactions between the intervention and study phases. Model 2 significantly improved the model fit (χ^2_3=8.6; *P*=.04).

[c]N/A: not applicable.

[d]Calculated from self-reported height and weight.

[e]Italics indicates statistical significance (*P*<.05).

[f]Time is rescaled to represent 2-week intervals.

[g]Represents the quadratic function of time.

[h]Represents the cubic function of time.

The changes in daily step count across study phases and time were different between participants, as evidenced by the significant random slopes of the study phase and time per participant detected in model 1. Adding cross-level interactions between the study phase and intervention phase (model 2; Table 4) significantly improved the model fit, indicating that differences between participants in changes in daily step counts across study phases could be explained by intervention effects. Findings from model 2 suggest that differences between participants in changes in daily steps are partially explained by significantly greater increases in daily steps during the gamification phase for participants in the intervention condition than in the control. This was evidenced by a significant interaction between the intervention condition and gamification phase (*B*=634.00; SE 244.81; $t_{167.52}$=2.59; *P*=.005) in a one-sided test, which explained 20.3% of the variance between participants in changes during the gamification phase (R^2=0.203). There were no differences in changes in daily steps between participants in the intervention and control conditions during the nudging phase (*B*=98.23; SE 325.52; $t_{163.28}$=0.30; *P*=.76) or follow-up (*B*=53.49; SE 381.67; $t_{143.61}$=0.14; *P*=.89) in a two-tailed test. Exploratory analysis showed that differences in changes in daily steps between participants could not be explained by individual differences, such as BMI, education, or gender. In essence, our findings indicate that the gamification phase of our intervention was effective in increasing the daily step count of office workers, compared with an active control.

However, improvements were not maintained during the physical nudges or during the follow-up phase.

Secondary Analysis: Self-Reported PA and SB

The sample size and mean values for the secondary measurements at each assessment point are listed in Table 3. Model 1 used in the primary analysis was refitted using secondary outcome measures. Higher BMI was associated with less time spent in SB (B=−0.21; SE 0.07; $t_{115.66}$=−2.96; P=.003) in a two-tailed test; however, no association was found between the intervention condition or study phase and the time spent in SB. The two-tailed tests revealed that participants in the intervention condition took fewer breaks from sitting than those in the control (B=−0.22; SE 0.11; $t_{229.12}$=−2.00; P=.046), and participants in both conditions took fewer breaks during follow-up compared with baseline (B=−0.34; SE 0.11; $t_{363.08}$=−3.01; P=.003). The two-tailed tests also showed that there was no effect of the intervention on hours spent in light PA in the workplace while controlling for parental nationality and age; however, overall, participants engaged in less light PA at the end of the nudging phase (B=−4.39; SE 1.45; $t_{333.73}$=−3.03; P=.002) and at follow-up (B=−3.47; SE 1.40; $t_{333.98}$=−2.48; P=.01), compared with baseline. After controlling for BMI, there was no main effect of the intervention for self-reported engagement (hours and days) in moderate-to-vigorous PA (ie, 150 minutes per week). However, a two-tailed test revealed a significant effect of study phase, with participants engaging in more hours of moderate-to-vigorous PA (B=0.55; SE 0.19; $t_{178.83}$=2.82; P=.005) and more days with sufficient moderate-to-vigorous PA (B=0.33; SE 0.16; $t_{304.05}$=2.14; P=.03) during follow-up, compared with baseline. Owing to low response rates for web-based questionnaires, it was not possible to refit model 2 with secondary outcomes to examine interactions between the intervention and study phases.

Discussion

Principal Findings

This study tested the effects of MoveMore, a multicomponent intervention designed to promote walking behavior (ie, light PA) and reduce SB in office workers. The MoveMore intervention consisted of an initial 5-week gamification phase encompassing a gamified digital app with social support features, followed by a 5-week physical nudges phase, including motivational and point-of-choice prompt nudges. By offering the gamification and physical nudges components separately, we could gain insights into their independent effects and explore whether physical nudges could promote maintenance of behavior change achieved during the gamification phase.

In line with our main hypothesis, significant increases in daily step counts were observed for participants in the intervention condition during the gamification phase compared with participants in the control group. However, contrary to our expectations, improvements in the daily step count for participants in the MoveMore intervention were not maintained during the physical nudges phase or at follow-up. We also hypothesized that similar improvements in secondary outcomes would be observed during the gamification phase for participants in the intervention. Although questionnaires administered in-person at baseline yielded high response rates, we could not investigate differences between intervention and control in changes of secondary outcome measures because of the low response rate in subsequent assessments via email. Nonetheless, overall, participants reported higher engagement in moderate-to-vigorous PA during work at follow-up compared with baseline. Unexpectedly, participants in both conditions reported engaging in less light PA during the physical nudges and follow-up phases than at baseline. Participants in the MoveMore intervention reported taking fewer breaks from sitting than those in the control, and participants in both groups reported taking fewer breaks during later study phases relative to baseline. However, given the poor reliability of the self-report measures in our sample, the validity and generalizability of these findings are limited. Future studies could address these limitations by using more sophisticated accelerometers that measure SB and by using more frequent and less time-consuming measurements that can be integrated in digital apps, such as ecological momentary assessments.

Nevertheless, our main findings suggest that adding gamification components with social support and social comparison features to a digital intervention seems to be an effective strategy for promoting PA in office workers, as evidenced by a significant interaction detected between the gamification phase and intervention condition. Exploratory analysis revealed that intervention effects were not influenced by individual differences, for example, BMI, education, or gender. The short-term effects of the gamification phase on step count were modest (ie, 763.5 increase in average number of daily steps for the participants in the intervention condition compared with 215.6 increase for those in the control) but comparable with previous RCTs evaluating this type of gamified digital intervention [46,67]. Similarly, these studies observed small but clinically significant effects on PA and/or SB. However, systematic reviews of pedometer-based interventions have found that these interventions typically increase PA by approximately 2000 steps per day [68]. Although the gamification phase of the MoveMore intervention resulted in considerable increases in step count, considering the increases reported in other RCTs assessing similar gamified interventions [46,69], these effects may seem underwhelming.

Several factors may explain the smaller effect sizes reported in this study, such as high daily step counts at baseline. Most similar intervention studies recruited inactive adults moving approximately 7000-7500 steps, which is far below the fairly disputed yet often recommended guideline of 10,000 steps per day [46,69,70]. Consequently, larger increases in step counts (ie, approximately 2000 steps) may have been observed in these samples precisely because low levels of PA at baseline allowed for and motivated participants to achieve greater improvements during the intervention period. Comparably, participants in this study walked on average 10,270 steps per day at baseline. Although increasing one's daily number of steps beyond the recommended guidelines is still beneficial [71], high rates of functioning at baseline may have hindered motivation and limited how many participants in our sample could improve (ie, ceiling effect). Given this ceiling effect, it is comprehensible

that increases in the step count of the highly active participants in our sample are lower and more difficult to maintain than those of interventions with inactive participants. Meta-analysis of PA interventions with healthy adults reported that studies with active adults reported lower effect sizes than those with sedentary adults [72]. Given that our sample was already highly active at baseline, the significant improvements in step count observed in participants in the intervention location during the gamification phase highlight the potential of gamified digital interventions to promote light PA.

In addition to the high rate of functioning at baseline, several other factors may partially explain our findings. For example, the use of an active control. Most studies exploring the effects of gamified digital interventions for PA and SB use either nonintervention controls [73] or self-monitoring controls [46,74]. To our knowledge, only a few studies investigating similar gamified digital interventions have used active controls with self-monitoring, goal setting, and personalized feedback [69]. In this study, implementing an active control allowed us to make stronger causal inferences about the effects of gamification and social support features on PA and SB. However, it is well established that the combination of self-monitoring and goal setting alone leads to the initiation of behavior change [54,71]. A meta-analysis of worksite PA interventions found that studies implementing active controls unsurprisingly reported lower effect sizes than those with no intervention controls [75]; thus, the use of active rather than no intervention control is another possible explanation for the relatively smaller effect sizes detected in this study. Another factor that may have influenced our results was the short duration of the gamification phase (5 weeks). A systematic review suggests that PA interventions with longer durations (ie, >24 weeks) are more likely to promote the maintenance of behavior change [76]. Thus, although behavior change was initiated during the gamification phase, 5 weeks may have been too short to form the habit of walking. Due to operational constraints, it was not possible to offer the gamification phase for longer durations in this study; however, future research should explore the effects of gamified digital interventions for longer durations to help establish habit formation.

We anticipated that physical nudges in the workplace could promote the maintenance of the behavior change achieved during the gamification phase. When designing the social norms nudges and point-of-choice prompts, fellow office workers from the intervention location served as role models, which could act as motivators and goal reinforcement [22,77]. Studies have found that combining motivational and point-of-choice prompt nudges was more effective than a control or either strategy alone in increasing stair climbing behavior [78]. However, in this study, combining motivational nudges based on social norms and authority with point-of-choice prompts for walking behavior resulted neither in increases in step count nor in maintenance of improvements achieved through the gamified digital intervention.

A possible explanation for these results may be the differences between the 2 locations. The effectiveness of certain BCTs, particularly nudges, is largely influenced by the context of implementation [33]. The 2 offices randomly allocated to the intervention and control groups differed considerably in terms of location and design. The control location was a wide building located in the city center, with large floor spaces and easy access to stairs and outside areas. Conversely, the intervention location was a tall building, with less surface area per floor and difficult access to outside areas. Participants in the control location had more space to walk indoors, whereas participants in the intervention location worked on multiple floors and may instead have had more opportunities to climb the stairs. Owing to operational constraints, no nudges for stair use could be implemented. Furthermore, weather conditions worsened (ie, lower temperature and more precipitation) throughout the course of the study, which may have discouraged participants from walking outside, maximizing the influence of the physical differences between the 2 locations. In addition, because of the collaborative nature of the study, the nudges used the colors and logos of the municipality of Rotterdam, which are used in other promotional materials found throughout the offices. This may have hindered the attractiveness of the nudges because they easily blended with other unrelated promotional materials.

The ineffectiveness of the nudges may also stem from the possibility that the physical nudges in the workplace in the form of table signs were not sufficiently engaging and motivating, especially when compared with the gamified digital intervention encompassing several BCTs. These findings, however, add support to the emerging evidence indicating that multicomponent interventions incorporating several nudges and BCTs, such as the gamified digital intervention used, are more effective at changing complex behaviors such as PA than interventions relying on only one or a few BCTs, such as the physical nudges used [23]. Most importantly, our findings suggest that gamification can be a useful complementary tool that can be flexibly incorporated to improve intervention effectiveness.

Strengths and Limitations

Given that reviews of gamified interventions have called for stronger empirical evaluations isolating the impacts of gamification [45], we consider the presence of an active control one of the strengths of this study. This study is the first to combine a gamified digital intervention with physical nudges to promote behavior change. An advantage of this study's design was the possibility to gain insight into the effects of gamification and physical nudges on initiation and maintenance of behavior change, respectively, although lack of randomization of the order of these components and order effects were the possible drawbacks. The use of objective measurements (ie, accelerometers) was another strength of this study, as self-reported measures of PA are often biased compared with objective measures, leading to false-positive findings [79]. In addition, the MoveMore intervention was implemented in the actual working environment of a large sample of Dutch office workers rather than in a controlled setting, which adds to the ecological and internal validity of our findings. As teams for the gamified challenges were formed according to office departments, a collective motivation for winning may have played a role. However, research has found that gamification is also effective in fostering motivation and promoting PA in adults when applied in individual settings [80]. Our findings, on the other hand, suggest that combining gamification with

changes in the social environment that promote social support and social comparison can effectively increase walking in office workers. As shown in a few previous studies, leveraging existing social structures through gamification by, for instance, allocating departments to different teams and stimulating cooperation and competition seems to be an effective strategy for promoting motivation and self-regulation for PA [46,47,81]. Although we recognize that the effects of gamification and social elements of the intervention could not be disentangled, the primary aim of this study is to design an effective intervention rather than isolating these effects.

Nevertheless, this study had several limitations inherent to field experiments, such as the inability to control for extraneous variables and operational constraints with regard to the physical nudges and the time frame of the study phases. Despite the large sample size, given that only 2 worksites were randomized to intervention or control, the limited number of clusters hindered the effectiveness of randomization and comparability between groups. Although we controlled for baseline differences between groups, differences between locations were possible confounders. However, given that the physical attributes of the control location facilitated PA, whereas the attributes of the intervention location hindered PA, it is highly unlikely that increases in step count of participants in the intervention location during the gamification phase can be attributed to physical differences between locations. Due to financial constraints, we could not opt for accelerometers that objectively measured SB. We relied on self-report measures of SB, which had poor reliability in our sample and have been shown to underestimate SB in adults [82,83], possibly explaining why the hypothesized intervention effects on SB were not observed even though they were supported by accelerometer step count data.

With regard to our statistical analysis, although we justified the quadratic and cubic effects of time in our model, we recognize that these higher-order effects are less stable and may be specific to our sample. In addition, dropout during physical nudges and follow-up phases resulted in a considerable amount of missing data, which hindered the assessment and interpretation of intervention effects during those phases. This dropout could simply be a result of timing, as physical nudges and follow-up phases occurred during the holiday season (ie, from December to February). Another possible explanation is that, as research has suggested, although gamification could be effective in enhancing engagement in the short term, users' motivations are unlikely to be sustained in the long term [48]. Another likely explanation for the high dropout rate in subsequent phases is that, as participants knew that the challenges were only available for a limited time, the gamification phase created a limited-time interest in the intervention, which then diminished once the gamified challenges ended. Owing to privacy agreements, it was not possible to collect data on user experience, functionality, and engagement with the app, which is a limitation of this study. Future research should investigate the long-term effects of gamified interventions while measuring user experience, engagement, and other possible mediators of intervention effectiveness.

Although our findings support the effectiveness of integrating gamification and social support features in digital interventions to promote light PA in active office workers, further research is needed to confirm the generalizability of our findings in more at-risk populations, such as inactive adults and adolescents from disadvantaged backgrounds. Given the unexpected effects of physical nudges, future studies should carefully consider the design and context of nudges. Nudges have gained attention because they are cost-effective in influencing people's behavior momentarily. However, further research is needed to explore how physical nudges can be more effectively combined with other interventions to promote the maintenance of complex behaviors, such as PA, in the short and long term. For example, offering gamified digital interventions and physical nudges simultaneously, rather than sequentially, may increase the intervention effectiveness. Research on gamification and nudging is still in its infancy. Future research with gamified digital interventions should continue innovating with the design of digital interventions by, for instance, testing different forms of social support and gamification components. Finally, future research on digital interventions should investigate complementary strategies to promote long-term maintenance of behavioral change, such as increasing engagement by tailoring the interventions to the participants' needs or using guided approaches in incorporating personal coaches or peer support.

Conclusions

Compared with an active control consisting of a digital app, including self-monitoring and goal setting, a gamified social support–based digital intervention was effective at promoting light PA (ie, objectively measured number of daily steps) in a sample of active office workers. Given the high prevalence of sedentary lifestyles and the associated health problems, both of which are costly to health care systems, even small improvements in light PA can have considerable effects at the population level. This study was one of the first to compare the effects of gamification with an active control and to test its effects on PA and SB of office workers. Our findings demonstrate that gamification can effectively complement the BCTs (eg, social support and social comparison) and nudges used in digital interventions to promote clinically significant improvements in PA, even beyond the recommended guidelines of 10,000 steps. Although more research is needed to establish its long-term effectiveness, policy makers should explore the use of gamified digital interventions with social support features as a promising strategy to promote behavior change and improve the health of the population. Physical nudges in the workplace were insufficient to promote the maintenance of behavior changes achieved during the gamification phase. Further research should explore how gamified digital interventions can be better leveraged to promote long-term behavior change, for instance, by investigating how to optimally tailor digital interventions to users' needs.

Acknowledgments

The authors would like to thank Ruth Plaggenborg and the team at SelfCare for designing and providing the MySelfCare and SelfCarePro apps used in the control and intervention locations, respectively, and for their contributions to the design and implementation of the study. The authors would also like to thank Berdi Christiaansen and other staff and participants from the municipality of Rotterdam for their involvement and contributions to the MoveMore study. Finally, the authors would like to thank Healthy'R [1], the expertise center on health behavior of the municipality of Rotterdam and Erasmus University Rotterdam. Healthy'R combines scientific knowledge in the field of behavior and health with practical knowledge and experiences to develop and investigate measures that lead to healthier choices of people in Rotterdam. The costs of the accelerometers, physical nudges, and digital apps used in this study were funded by the municipality of Rotterdam. The content is solely the responsibility of the authors and does not necessarily represent the official views and interests of the funders or authors' affiliated academic institutions. The funders had no role in the design of the study and were not involved in data analysis, decision to publish, or preparation of the manuscript.

Editorial Notice

This randomized study was only retrospectively registered; this is because, as explained by authors, at the time of data collection they did not believe that this study met the criteria for a clinical trial, due to the fact that the study only included behavioural outcomes, and did not involve any health outcomes. The editor granted an exception from ICMJE rules mandating prospective registration of randomized trials because the risk of bias appears low and the study was considered formative, guiding the development of the application. However, readers are advised to carefully assess the validity of any potential explicit or implicit claims related to primary outcomes or effectiveness, as retrospective registration does not prevent authors from changing their outcome measures retrospectively.

Authors' Contributions

AM designed the study, developed the intervention, implemented the intervention, managed data collection, conducted statistical analyses and interpretation, and wrote the manuscript. MS helped develop the intervention material, recruit participants, implement the intervention, and manage the data collection. JJ helped with the statistical analysis and interpretation and provided feedback on the manuscript. AS, GN, and SD contributed to the study design and provided guidance and consultation throughout the study and feedback on the manuscript. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT-EHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 2743 KB - jmir_v23i4e19875_app1.pdf]

Multimedia Appendix 2
Pictures displaying the components included in the basic and gamified versions of the digital apps used.
[DOCX File , 5498 KB - jmir_v23i4e19875_app2.docx]

Multimedia Appendix 3
Pictures displaying the original nudges (in Dutch) introduced in the workplace of participants in the MoveMore intervention condition during the physical nudges phase of the study.
[PDF File (Adobe PDF File), 550 KB - jmir_v23i4e19875_app3.pdf]

References

1. Hupin D, Roche F, Gremeaux V, Chatard J, Oriol M, Gaspoz J, et al. Even a low-dose of moderate-to-vigorous physical activity reduces mortality by 22% in adults aged ≥60 years: a systematic review and meta-analysis. Br J Sports Med 2015 Oct;49(19):1262-1267. [doi: 10.1136/bjsports-2014-094306] [Medline: 26238869]
2. Amagasa S, Machida M, Fukushima N, Kikuchi H, Takamiya T, Odagiri Y, et al. Is objectively measured light-intensity physical activity associated with health outcomes after adjustment for moderate-to-vigorous physical activity in adults? A systematic review. Int J Behav Nutr Phys Act 2018 Jul 09;15(1):65 [FREE Full text] [doi: 10.1186/s12966-018-0695-z] [Medline: 29986718]
3. Kraus WE, Janz KF, Powell KE, Campbell WW, Jakicic JM, Troiano RP, 2018 PHYSICAL ACTIVITY GUIDELINES ADVISORY COMMITTEE*. Daily Step Counts for Measuring Physical Activity Exposure and Its Relation to Health. Med Sci Sports Exerc 2019 Jun;51(6):1206-1212 [FREE Full text] [doi: 10.1249/MSS.0000000000001932] [Medline: 31095077]

4. Loprinzi PD. Objectively measured light and moderate-to-vigorous physical activity is associated with lower depression levels among older US adults. Aging Ment Health 2013 Sep;17(7):801-805. [doi: 10.1080/13607863.2013.801066] [Medline: 23731057]
5. Ku P, Steptoe A, Liao Y, Sun W, Chen L. Prospective relationship between objectively measured light physical activity and depressive symptoms in later life. Int J Geriatr Psychiatry 2018 Jan;33(1):58-65. [doi: 10.1002/gps.4672] [Medline: 28181713]
6. Jonsdottir IH, Rödjer L, Hadzibajramovic E, Börjesson M, Ahlborg G. A prospective study of leisure-time physical activity and mental health in Swedish health care workers and social insurance officers. Prev Med 2010 Nov;51(5):373-377. [doi: 10.1016/j.ypmed.2010.07.019] [Medline: 20691721]
7. Katzmarzyk PT, Lee IM. Sedentary behaviour and life expectancy in the USA: a cause-deleted life table analysis. BMJ Open 2012;2(4):1-8 [FREE Full text] [doi: 10.1136/bmjopen-2012-000828] [Medline: 22777603]
8. Tremblay M, Aubert S, Barnes J, Saunders T, Carson V, Latimer-Cheung AE, SBRN Terminology Consensus Project Participants. Sedentary Behavior Research Network (SBRN) - Terminology Consensus Project process and outcome. Int J Behav Nutr Phys Act 2017 Jun 10;14(1):75 [FREE Full text] [doi: 10.1186/s12966-017-0525-8] [Medline: 28599680]
9. Proper KI, Singh AS, van Mechelen W, Chinapaw MJM. Sedentary behaviors and health outcomes among adults: a systematic review of prospective studies. Am J Prev Med 2011 Feb;40(2):174-182. [doi: 10.1016/j.amepre.2010.10.015] [Medline: 21238866]
10. Miller R, Brown W. Steps and sitting in a working population. Int J Behav Med 2004;11(4):219-224. [doi: 10.1207/s15327558ijbm1104_5] [Medline: 15657022]
11. Clemes S, Patel R, Mahon C, Griffiths P. Sitting time and step counts in office workers. Occup Med (Lond) 2014 Apr;64(3):188-192. [doi: 10.1093/occmed/kqt164] [Medline: 24477502]
12. van Dommelen P, Coffeng JK, van der Ploeg HP, van der Beek AJ, Boot CRL, Hendriksen IJM. Objectively Measured Total and Occupational Sedentary Time in Three Work Settings. PLoS One 2016;11(3):e0149951 [FREE Full text] [doi: 10.1371/journal.pone.0149951] [Medline: 26937959]
13. Leefstijlmonitor. (CBS) Centraal Bureau voor de Statistiek. 2015. URL: https://www.volksgezondheidenzorg.info/sites/default/files/lsm_2015.pdf [accessed 2020-01-25]
14. Parry S, Straker L. The contribution of office work to sedentary behaviour associated risk. BMC Public Health 2013 Apr 04;13:296 [FREE Full text] [doi: 10.1186/1471-2458-13-296] [Medline: 23557495]
15. Chu A, Ng S, Tan C, Win A, Koh D, Müller-Riemenschneider F. A systematic review and meta-analysis of workplace intervention strategies to reduce sedentary time in white-collar workers. Obes Rev 2016 May;17(5):467-481. [doi: 10.1111/obr.12388] [Medline: 26990220]
16. Schwarzer R. Modeling Health Behavior Change: How to Predict and Modify the Adoption and Maintenance of Health Behaviors. Applied Psychology 2008 Jan;57(1):1-29. [doi: 10.1111/j.1464-0597.2007.00325.x]
17. Anderson ES, Winett RA, Wojcik JR, Williams DM. Social cognitive mediators of change in a group randomized nutrition and physical activity intervention: social support, self-efficacy, outcome expectations and self-regulation in the guide-to-health trial. J Health Psychol 2010 Jan;15(1):21-32. [doi: 10.1177/1359105309342297] [Medline: 20064881]
18. Anderson ES, Wojcik JR, Winett RA, Williams DM. Social-cognitive determinants of physical activity: the influence of social support, self-efficacy, outcome expectations, and self-regulation among participants in a church-based health promotion study. Health Psychol 2006 Jul;25(4):510-520. [doi: 10.1037/0278-6133.25.4.510] [Medline: 16846326]
19. Humpel N, Owen N, Leslie E. Environmental factors associated with adults' participation in physical activity: a review. Am J Prev Med 2002 Apr;22(3):188-199. [doi: 10.1016/s0749-3797(01)00426-3] [Medline: 11897464]
20. Yen YH, Li C. Correction: Determinants of physical activity: A path model based on an ecological model of active living. PLoS One 2019;14(9):e0222625 [FREE Full text] [doi: 10.1371/journal.pone.0222625] [Medline: 31513693]
21. Bauman AE, Reis RS, Sallis JF, Wells JC, Loos RJF, Martin BW, Lancet Physical Activity Series Working Group. Correlates of physical activity: why are some people physically active and others not? Lancet 2012 Jul 21;380(9838):258-271. [doi: 10.1016/S0140-6736(12)60735-1] [Medline: 22818938]
22. Bandura A. Health promotion by social cognitive means. Health Educ Behav 2004 Apr;31(2):143-164. [doi: 10.1177/1090198104263660] [Medline: 15090118]
23. Webb T, Joseph J, Yardley L, Michie S. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. J Med Internet Res 2010 Feb 17;12(1):e4 [FREE Full text] [doi: 10.2196/jmir.1376] [Medline: 20164043]
24. Webb T, Sheeran P. Does changing behavioral intentions engender behavior change? A meta-analysis of the experimental evidence. Psychol Bull 2006 Mar;132(2):249-268. [doi: 10.1037/0033-2909.132.2.249] [Medline: 16536643]
25. Marteau T, Ogilvie D, Roland M, Suhrcke M, Kelly MP. Judging nudging: can nudging improve population health? BMJ 2011 Jan 25;342:d228. [doi: 10.1136/bmj.d228] [Medline: 21266441]
26. Thaler RH, Sunstein CR. Nudge: improving decisions about health, wealth, and happiness. New Haven: Yale University Press; 2008.
27. McNeill LH, Kreuter MW, Subramanian S. Social environment and physical activity: a review of concepts and evidence. Soc Sci Med 2006 Aug;63(4):1011-1022. [doi: 10.1016/j.socscimed.2006.03.012] [Medline: 16650513]

28. Wally CM, Cameron LD. A Randomized-Controlled Trial of Social Norm Interventions to Increase Physical Activity. Ann Behav Med 2017 Oct;51(5):642-651. [doi: 10.1007/s12160-017-9887-z] [Medline: 28213634]
29. Bauman A, Milton K, Kariuki M, Fedel K, Lewicka M. Is there sufficient evidence regarding signage-based stair use interventions? A sequential meta-analysis. BMJ Open 2017 Nov 28;7(11):e012459 [FREE Full text] [doi: 10.1136/bmjopen-2016-012459] [Medline: 29183924]
30. Nocon M, Müller-Riemenschneider F, Nitzschke K, Willich SN. Review Article: Increasing physical activity with point-of-choice prompts--a systematic review. Scand J Public Health 2010 Aug;38(6):633-638. [doi: 10.1177/1403494810375865] [Medline: 20601438]
31. Evans RE, Fawole HO, Sheriff SA, Dall PM, Grant PM, Ryan CG. Point-of-choice prompts to reduce sitting time at work: a randomized trial. Am J Prev Med 2012 Sep;43(3):293-297. [doi: 10.1016/j.amepre.2012.05.010] [Medline: 22898122]
32. Benartzi S, Beshears J, Milkman K, Sunstein C, Thaler R, Shankar M, et al. Should Governments Invest More in Nudging? Psychol Sci 2017 Aug;28(8):1041-1055 [FREE Full text] [doi: 10.1177/0956797617702501] [Medline: 28581899]
33. Kremers S, Eves F, Andersen R. Environmental changes to promote physical activity and healthy dietary behavior. J Environ Public Health 2012;2012:470858 [FREE Full text] [doi: 10.1155/2012/470858] [Medline: 22991564]
34. Broers JV, De Breucker C, Van den Broucke S, Luminet O. A systematic review and meta-analysis of the effectiveness of nudging to increase fruit and vegetable choice. Eur J Public Health 2017 Oct 01;27(5):912-920. [doi: 10.1093/eurpub/ckx085] [Medline: 28655176]
35. Arno A, Thomas S. The efficacy of nudge theory strategies in influencing adult dietary behaviour: a systematic review and meta-analysis. BMC Public Health 2016 Jul 30;16:676 [FREE Full text] [doi: 10.1186/s12889-016-3272-x] [Medline: 27475752]
36. MacDonald B, Gibson A, Janssen X, Kirk A. A Mixed Methods Evaluation of a Digital Intervention to Improve Sedentary Behaviour Across Multiple Workplace Settings. Int J Environ Res Public Health 2020 Jun 24;17(12):4538 [FREE Full text] [doi: 10.3390/ijerph17124538] [Medline: 32599730]
37. Forberger S, Reisch L, Kampfmann T, Zeeb H. Nudging to move: a scoping review of the use of choice architecture interventions to promote physical activity in the general population. Int J Behav Nutr Phys Act 2019 Sep 03;16(1):77 [FREE Full text] [doi: 10.1186/s12966-019-0844-z] [Medline: 31481090]
38. Mohr DC, Burns MN, Schueller SM, Clarke G, Klinkman M. Behavioral intervention technologies: evidence review and recommendations for future research in mental health. Gen Hosp Psychiatry 2013;35(4):332-338 [FREE Full text] [doi: 10.1016/j.genhosppsych.2013.03.008] [Medline: 23664503]
39. Free C, Phillips G, Galli L, Watson L, Felix L, Edwards P, et al. The effectiveness of mobile-health technology-based health behaviour change or disease management interventions for health care consumers: a systematic review. PLoS Med 2013;10(1):e1001362 [FREE Full text] [doi: 10.1371/journal.pmed.1001362] [Medline: 23349621]
40. Hamari J, Koivisto J, Sarsa H. Does Gamification Work? -- A Literature Review of Empirical Studies on Gamification. 2014 Presented at: 47th Hawaii International Conference on System Sciences; 2014; Waikoloa, HI, USA URL: https://doi.org/10.1109/HICSS.2014.377 [doi: 10.1109/hicss.2014.377]
41. Cugelman B. Gamification: what it is and why it matters to digital health behavior change developers. JMIR Serious Games 2013 Dec 12;1(1):e3 [FREE Full text] [doi: 10.2196/games.3139] [Medline: 25658754]
42. Deterding S, Dixon D, Khaled R, Nacke L. From Game Design Elements to Gamefulness?: Defining ? Gamification. 2011 Presented at: MindTrek '11: Academic MindTrek 2011; 2011; Tampere, Finland p. 9-15. [doi: 10.1145/2181037.2181040]
43. Edwards E, Lumsden J, Rivas C, Steed L, Edwards L, Thiyagarajan A, et al. Gamification for health promotion: systematic review of behaviour change techniques in smartphone apps. BMJ Open 2016 Oct 04;6(10):e012447 [FREE Full text] [doi: 10.1136/bmjopen-2016-012447] [Medline: 27707829]
44. Kohl LFM, Crutzen R, de Vries NK. Online prevention aimed at lifestyle behaviors: a systematic review of reviews. J Med Internet Res 2013 Jul 16;15(7):e146 [FREE Full text] [doi: 10.2196/jmir.2665] [Medline: 23859884]
45. Johnson D, Deterding S, Kuhn K, Staneva A, Stoyanov S, Hides L. Gamification for health and wellbeing: A systematic review of the literature. Internet Interv 2016 Nov;6:89-106 [FREE Full text] [doi: 10.1016/j.invent.2016.10.002] [Medline: 30135818]
46. Gremaud A, Carr L, Simmering J, Evans N, Cremer J, Segre A, et al. Gamifying Accelerometer Use Increases Physical Activity Levels of Sedentary Office Workers. J Am Heart Assoc 2018 Jul 02;7(13):e007735 [FREE Full text] [doi: 10.1161/JAHA.117.007735] [Medline: 29967221]
47. Lowensteyn I, Berberian V, Berger C, Da Costa D, Joseph L, Grover S. The Sustainability of a Workplace Wellness Program That Incorporates Gamification Principles: Participant Engagement and Health Benefits After 2 Years. Am J Health Promot 2019 Jul;33(6):850-858. [doi: 10.1177/0890117118823165] [Medline: 30665309]
48. Sardi L, Idri A, Fernández-Alemán JL. A systematic review of gamification in e-Health. J Biomed Inform 2017 Jul;71:31-48 [FREE Full text] [doi: 10.1016/j.jbi.2017.05.011] [Medline: 28536062]
49. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res 2011 Dec 31;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]

50. Parry S, Straker L, Gilson ND, Smith AJ. Participatory workplace interventions can reduce sedentary time for office workers--a randomised controlled trial. PLoS One 2013;8(11):e78957 [FREE Full text] [doi: 10.1371/journal.pone.0078957] [Medline: 24265734]
51. Chae D, Kim S, Park Y, Hwang Y. The Effects of an Academic--Workplace Partnership Intervention to Promote Physical Activity in Sedentary Office Workers. Workplace Health Saf 2015 Jun;63(6):259-266. [doi: 10.1177/2165079915579576] [Medline: 26012515]
52. Diaz KM, Krupka DJ, Chang MJ, Peacock J, Ma Y, Goldsmith J, et al. Fitbit®: An accurate and reliable device for wireless physical activity tracking. Int J Cardiol 2015 Apr 15;185:138-140 [FREE Full text] [doi: 10.1016/j.ijcard.2015.03.038] [Medline: 25795203]
53. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95. [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
54. Michie S, Abraham C, Whittington C, McAteer J, Gupta S. Effective techniques in healthy eating and physical activity interventions: a meta-regression. Health Psychol 2009 Nov;28(6):690-701. [doi: 10.1037/a0016136] [Medline: 19916637]
55. Clark B, Thorp A, Winkler E, Gardiner P, Healy G, Owen N, et al. Validity of self-reported measures of workplace sitting time and breaks in sitting time. Med Sci Sports Exerc 2011 Oct;43(10):1907-1912. [doi: 10.1249/MSS.0b013e31821820a2] [Medline: 21926535]
56. Wendel-Vos G. Reproducibility and relative validity of the short questionnaire to assess health-enhancing physical activity. Journal of Clinical Epidemiology 2003 Dec;56(12):1163-1169 [FREE Full text] [doi: 10.1016/S0895-4356(03)00220-8]
57. de Hollander EL, Zwart L, de Vries SI, Wendel-Vos W. The SQUASH was a more valid tool than the OBiN for categorizing adults according to the Dutch physical activity and the combined guideline. J Clin Epidemiol 2012 Jan;65(1):73-81. [doi: 10.1016/j.jclinepi.2011.05.005] [Medline: 21840174]
58. Provo, Utah, USA. Qualtrics. 2020. URL: https://www.qualtrics.com [accessed 2019-03-01] [WebCite Cache ID www.qualtrics.com]
59. SelfCare. SelfCare Homepage Internet. SelfCare. 2019. URL: https://selfcare4me.com [accessed 2018-10-31]
60. Bates D, Mächler M, Bolker B, Walker S. Fitting Linear Mixed-Effects Models Using. J. Stat. Soft 2015;67(1):1-48. [doi: 10.18637/jss.v067.i01]
61. de Haan-Rietdijk S, Kuppens P, Hamaker E. What's in a Day? A Guide to Decomposing the Variance in Intensive Longitudinal Data. Front Psychol 2016;7:891 [FREE Full text] [doi: 10.3389/fpsyg.2016.00891] [Medline: 27378986]
62. Barr D, Levy R, Scheepers C, Tily H. Random effects structure for confirmatory hypothesis testing: Keep it maximal. J Mem Lang 2013 Apr;68(3):255-278 [FREE Full text] [doi: 10.1016/j.jml.2012.11.001] [Medline: 24403724]
63. Singer J, Willett J. Applied longitudinal data analysis: modeling change and event occurrence. In: Applied longitudinal data analysis: modeling change and event occurrence. Oxford: Oxford University Press; 2003:1-644.
64. Folta S, Kuder J, Goldberg J, Hyatt R, Must A, Naumova E, et al. Changes in diet and physical activity resulting from the Shape Up Somerville community intervention. BMC Pediatr 2013 Oct 04;13:157 [FREE Full text] [doi: 10.1186/1471-2431-13-157] [Medline: 24093936]
65. Van 't Riet J, Crutzen R, De Vries H. Investigating predictors of visiting, using, and revisiting an online health-communication program: a longitudinal study. J Med Internet Res 2010 Sep 02;12(3):e37 [FREE Full text] [doi: 10.2196/jmir.1345] [Medline: 20813716]
66. Kang M, Zhu W, Tudor-Locke C, Ainsworth B. Experimental Determination of Effectiveness of an Individual Information-Centered Approach in Recovering Step-Count Missing Data. Measurement in Physical Education and Exercise Science 2005 Oct;9(4):233-250 [FREE Full text] [doi: 10.1207/s15327841mpee0904_3]
67. van Dantzig S, Geleijnse G, van Halteren A. Toward a persuasive mobile application to reduce sedentary behavior. Pers Ubiquit Comput 2012 Jul 12;17(6):1237-1246 [FREE Full text] [doi: 10.1007/s00779-012-0588-0]
68. Kang M, Marshall S, Barreira TV, Lee J. Effect of Pedometer-Based Physical Activity Interventions: A Meta-Analysis. Res Quart Exerc Sport 2009 Sep 01;80(3):648-655 [FREE Full text] [doi: 10.5641/027013609x13088500160000]
69. Patel M, Benjamin E, Volpp K, Fox C, Small D, Massaro J, et al. Effect of a Game-Based Intervention Designed to Enhance Social Incentives to Increase Physical Activity Among Families: The BE FIT Randomized Clinical Trial. JAMA Intern Med 2017 Nov 01;177(11):1586-1593 [FREE Full text] [doi: 10.1001/jamainternmed.2017.3458] [Medline: 28973115]
70. Tudor-Locke C, Bassett DR. How many steps/day are enough? Preliminary pedometer indices for public health. Sports Med 2004;34(1):1-8. [doi: 10.2165/00007256-200434010-00001] [Medline: 14715035]
71. Tudor-Locke C, Craig C, Brown W, Clemes S, De Cocker K, Giles-Corti B, et al. How many steps/day are enough? For adults. Int J Behav Nutr Phys Act 2011 Jul 28;8:79 [FREE Full text] [doi: 10.1186/1479-5868-8-79] [Medline: 21798015]
72. Conn V, Hafdahl A, Mehr D. Interventions to increase physical activity among healthy adults: meta-analysis of outcomes. Am J Public Health 2011 Apr;101(4):751-758 [FREE Full text] [doi: 10.2105/AJPH.2010.194381] [Medline: 21330590]
73. Allam A, Kostova Z, Nakamoto K, Schulz PJ. The effect of social support features and gamification on a Web-based intervention for rheumatoid arthritis patients: randomized controlled trial. J Med Internet Res 2015 Jan 09;17(1):e14 [FREE Full text] [doi: 10.2196/jmir.3510] [Medline: 25574939]

74. Chen Y, Pu P. HealthyTogether? exploring Social Incentives for Mobile Fitness Applications. 2014 Apr Presented at: In the second International Symposium of Chinese; 2014; Toronto, Canada p. E. [doi: 10.1145/2592235.2592240]
75. Abraham C, Graham-Rowe E. Are worksite interventions effective in increasing physical activity? A systematic review and meta-analysis. Health Psychology Review 2009 Mar;3(1):108-144 [FREE Full text] [doi: 10.1080/17437190903151096]
76. Fjeldsoe B, Neuhaus M, Winkler E, Eakin E. Systematic review of maintenance of behavior change following physical activity and dietary interventions. Health Psychol 2011 Jan;30(1):99-109. [doi: 10.1037/a0021974] [Medline: 21299298]
77. Morgenroth T, Ryan M, Peters K. The Motivational Theory of Role Modeling: How Role Models Influence Role Aspirants' Goals. Review of General Psychology 2015 Dec;19(4):465-483 [FREE Full text] [doi: 10.1037/gpr0000059]
78. Lewis A, Eves F. Prompt before the choice is made: effects of a stair-climbing intervention in university buildings. Br J Health Psychol 2012 Sep;17(3):631-643. [doi: 10.1111/j.2044-8287.2011.02060.x] [Medline: 22248016]
79. Prince S, Adamo K, Hamel M, Hardt J, Connor Gorber S, Tremblay M. A comparison of direct versus self-report measures for assessing physical activity in adults: a systematic review. Int J Behav Nutr Phys Act 2008 Nov 06;5:56 [FREE Full text] [doi: 10.1186/1479-5868-5-56] [Medline: 18990237]
80. Kappen D. Gamification of Older Adults' Physical Activity: An Eight-Week Study. In: Proceedings of the 51st Hawaii International Conference on System Sciences. 2018 Presented at: 51st Hawaii International Conference on System Sciences; 2018; Hawaii p. A. [doi: 10.24251/hicss.2018.149]
81. Harris M. Maintenance of behaviour change following a community-wide gamification based physical activity intervention. Prev Med Rep 2019 Mar;13:37-40 [FREE Full text] [doi: 10.1016/j.pmedr.2018.11.009] [Medline: 30510892]
82. Chastin SFM, Culhane B, Dall PM. Comparison of self-reported measure of sitting time (IPAQ) with objective measurement (activPAL). Physiol Meas 2014 Nov;35(11):2319-2328. [doi: 10.1088/0967-3334/35/11/2319] [Medline: 25341050]
83. Aguilar-Farías N, Brown W, Olds T, Geeske Peeters GMEE. Validity of self-report methods for measuring sedentary behaviour in older adults. J Sci Med Sport 2015 Nov;18(6):662-666. [doi: 10.1016/j.jsams.2014.08.004] [Medline: 25172367]

Abbreviations

BCT: behavioral change technique
CONSORT-EHEALTH: Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth
ICC: intraclass correlation coefficient
PA: physical activity
RCT: randomized controlled trial
SB: sedentary behavior
SQUASH: Short Questionnaire for Assessing Health enhancing physical activity

Edited by G Eysenbach; submitted 27.05.20; peer-reviewed by J Alvarez Pitti, C Valle, N Khalili-Mahani, A Teles; comments to author 28.10.20; revised version received 22.12.20; accepted 17.01.21; published 12.04.21.

Please cite as:
Mamede A, Noordzij G, Jongerling J, Snijders M, Schop-Etman A, Denktas S
Combining Web-Based Gamification and Physical Nudges With an App (MoveMore) to Promote Walking Breaks and Reduce Sedentary Behavior of Office Workers: Field Study
J Med Internet Res 2021;23(4):e19875
URL: https://www.jmir.org/2021/4/e19875
doi:10.2196/19875
PMID:33843593

Original Paper

A Multimodality Machine Learning Approach to Differentiate Severe and Nonsevere COVID-19: Model Development and Validation

Yuanfang Chen[1,2*], MSc; Liu Ouyang[3*], MD; Forrest S Bao[4*], PhD; Qian Li[5*], MD; Lei Han[1,6*], MD, PhD; Hengdong Zhang[1,6], MD, PhD; Baoli Zhu[1,6,7], MSc; Yaorong Ge[8], PhD; Patrick Robinson[9], MD; Ming Xu[1,6,9], MD, PhD; Jie Liu[10], MD, PhD; Shi Chen[9,11], PhD

[1]Public Health Research Institute of Jiangsu Province, Nanjing, China
[2]Institute of HIV/AIDS/STI Prevention and Control, Jiangsu Provincial Center for Disease Control and Prevention, Nanjing, China
[3]Department of Orthopaedics, Union Hospital, Huazhong University of Science and Technology, Wuhan, China
[4]Department of Computer Science, Iowa State University, Ames, IA, United States
[5]Department of Pediatrics, Affiliated Kunshan Hospital of Jiangsu University, Kunshan, China
[6]Department of Occupational Disease Prevention, Jiangsu Provincial Center for Disease Control and Prevention, Nanjing, China
[7]School of Public health, Nanjing Medical University, Nanjing, China
[8]Department of Software and Information Systems, University of North Carolina at Charlotte, Charlotte, NC, United States
[9]Department of Public Health Sciences, University of North Carolina at Charlotte, Charlotte, NC, United States
[10]Department of Radiology, Union Hospital, Huazhong University of Science and Technology, Wuhan, China
[11]School of Data Science, University of North Carolina at Charlotte, Charlotte, NC, United States
[*]these authors contributed equally

Corresponding Author:
Ming Xu, MD, PhD
Department of Occupational Disease Prevention
Jiangsu Provincial Center for Disease Control and Prevention
172 Jiangsu Road
Nanjing, 210009
China
Phone: 86 85393210
Email: sosolou@126.com

Abstract

Background: Effectively and efficiently diagnosing patients who have COVID-19 with the accurate clinical type of the disease is essential to achieve optimal outcomes for the patients as well as to reduce the risk of overloading the health care system. Currently, severe and nonsevere COVID-19 types are differentiated by only a few features, which do not comprehensively characterize the complicated pathological, physiological, and immunological responses to SARS-CoV-2 infection in the different disease types. In addition, these type-defining features may not be readily testable at the time of diagnosis.

Objective: In this study, we aimed to use a machine learning approach to understand COVID-19 more comprehensively, accurately differentiate severe and nonsevere COVID-19 clinical types based on multiple medical features, and provide reliable predictions of the clinical type of the disease.

Methods: For this study, we recruited 214 confirmed patients with nonsevere COVID-19 and 148 patients with severe COVID-19. The clinical characteristics (26 features) and laboratory test results (26 features) upon admission were acquired as two input modalities. Exploratory analyses demonstrated that these features differed substantially between two clinical types. Machine learning random forest models based on all the features in each modality as well as on the top 5 features in each modality combined were developed and validated to differentiate COVID-19 clinical types.

Results: Using clinical and laboratory results independently as input, the random forest models achieved >90% and >95% predictive accuracy, respectively. The importance scores of the input features were further evaluated, and the top 5 features from each modality were identified (age, hypertension, cardiovascular disease, gender, and diabetes for the clinical features modality, and dimerized plasmin fragment D, high sensitivity troponin I, absolute neutrophil count, interleukin 6, and lactate dehydrogenase

for the laboratory testing modality, in descending order). Using these top 10 multimodal features as the only input instead of all 52 features combined, the random forest model was able to achieve 97% predictive accuracy.

Conclusions: Our findings shed light on how the human body reacts to SARS-CoV-2 infection as a unit and provide insights on effectively evaluating the disease severity of patients with COVID-19 based on more common medical features when gold standard features are not available. We suggest that clinical information can be used as an initial screening tool for self-evaluation and triage, while laboratory test results should be applied when accuracy is the priority.

(J Med Internet Res 2021;23(4):e23948) doi:10.2196/23948

KEYWORDS

COVID-19; clinical type; multimodality; classification; machine learning; machine learning; diagnosis; prediction; reliable; decision support

Introduction

COVID-19 is a pandemic disease caused by the novel SARS-CoV-2 virus. As of January 12, 2021, COVID-19 had spread through at least 220 countries and regions, resulting in more than 88 million cases and almost 2 million deaths [1]. It has become the single most severe pandemic in the 21st century, dwarfing other coronavirus-caused epidemics, such as severe acute respiratory syndrome (SARS) in 2003 and Middle East respiratory syndrome (MERS) in 2012. COVID-19 is especially challenging to health professionals and the general population. Unlike in the preceding SARS and MERS epidemics, patients with COVID-19 can be either asymptomatic or symptomatic, and the virus has been demonstrated to be transmissible in both states to varying degrees [2-5]. In addition, the distinct clinical types of COVID-19, nonsevere and severe, require different treatment and care plans [6]. In current studies, patients with COVID-19 can be differentiated from patients who do not have the disease; however, further detection of nonsevere or severe types of COVID-19 has not been comprehensively explored. Patients with nonsevere COVID-19 can be accommodated with less intensive clinical monitoring and intervention, including treating pre-existing conditions and preventing health care–associated infections and other comorbidities [7]. In contrast, patients with severe disease require close monitoring, usually in the intensive care unit (ICU), by more clinicians [6]. Therefore, effectively and efficiently classifying clinical types of COVID-19 is essential for triage, resource optimization, and care planning for frontline clinicians and health care systems as well as for the patients [6,8].

Currently, nonsevere and severe COVID-19 types are classified based on only a few clinical features in China, including shortness of breath, O_2 saturation, and PaO_2 [9]. Because of the complexity of the pathological, physiological, and immunological response of COVID-19, these three features do not sufficiently characterize the difference between nonsevere and severe types in patients with COVID-19 [9-11]. Although shortness of breath can be self-monitored, O_2 saturation and PaO_2 cannot be accurately self-evaluated and may not be readily assessed in clinical settings, especially for socioeconomically disadvantaged patients. In addition, some patients with severe disease may not present shortness of breath initially. However, without proper medical intervention, their clinical course will worsen abruptly, often resulting in respiratory failure with high mortality [6]. Therefore, these gold standard features bear the risk of misclassification and misdiagnosis. Misclassification of COVID-19 clinical types can result in inappropriate early treatment decisions; this can place patients at risk of progression due to insufficiently aggressive supportive therapy or expose other patients to overly invasive treatment, both of which have negative clinical consequences. In addition, the three defining features may not be readily available during initial diagnosis when resources are inadequate.

It is therefore critical to provide a rapid, accurate, and efficient method to determine the severity of COVID-19 infection and identify the clinical type using alternative features. This determination will enable optimization of treatment plans for patient care and improve utilization of health care resources and staff. We suggest that additional readily available medical features, including the patient's comorbidities (eg, hypertension and diabetes) and symptoms (eg, fever and chest pain), as well as laboratory test results, can be used to develop an effective method to determine the clinical type and severity of COVID-19 [12,13]. Angiotensin-converting enzyme 2 (ACE-2) receptors, which facilitate SARS-CoV-2 infiltration, are distributed across multiple organs and systems in the human body [14]. More recent discoveries have found that in addition to the respiratory system, SARS-CoV-2 can invade digestive, reproductive, and even neural systems [15-18]. In other words, all clinical and laboratory test information of patients with COVID-19 could be consequences or risk factors of SARS-CoV-2 infection. In clinical practice to treat COVID-19, clinicians not from respiratory units or ICUs may rely only on the referenced features [9] while neglecting diverse and important clinical features of COVID-19, and they may miss critical signs leading to undesirable prognosis.

The potential power of clinical and laboratory testing features, as well as their combinations, to determine COVID-19 clinical type is currently being explored [19-24]. To use such diverse multimodality information as alternative evidence to facilitate accurate classifications, we propose a data mining and machine learning (ML) framework as an alternative to commonly used hypothesis-driven parametric models. The goal of this study is to provide reliable data-driven support for clinicians, even those who do not have comprehensive experience in diagnosing the emerging disease COVID-19. We aim to explore and contrast the distributions of clinical and laboratory testing features between nonsevere and severe COVID-19 types. We will identify key features that differ substantially between the two clinical types. Next, we will investigate whether a single

modality or specific combination of features across modalities are able to provide accurate classification models via ML techniques. Specifically, we aim to identify a small and practical set of input features that can accurately differentiate COVID-19 clinical types. The insights gained from this study, as well as the developed end-to-end multimodal data analysis and ML framework, will enable us to better understand the comprehensive pathology of COVID-19, further distinguish COVID-19 from other respiratory infections, and apply the framework to other diseases with multimodal medical data in the future.

Methods

Data Source and Clinical Feature Extraction

In this study, we recruited 362 patients with COVID-19 from January to March 2020, including 148 patients presenting with severe disease and 214 patients lacking criteria for severe disease during admission, from Wuhan Union Hospital, China. The definitions of nonsevere and severe cases were mainly adopted from the official COVID-19 Diagnosis and Treatment Plan from the National Health Commission of China, and we also consulted guidelines from the American Thoracic Society [9-11]. Patients with severe COVID-19 should present any one of the following features: (1) respiratory rate >30 breaths per minute; (2) oxygen saturation <93% at rest; or (3) PaO_2/fraction of inspired oxygen <300 mm Hg (40 kPa). Each patient with COVID-19 was confirmed by two independent quantitative reverse transcriptase–polymerase chain reaction tests before being included in this study. All patients or their responsible surrogates signed informed consent forms prior to study inclusion. The patients' symptoms were evaluated and blood samples were drawn upon admission to perform laboratory testing. No pediatric patients aged less than 18 years were included.

The patients' deidentified medical information include two major modalities of features, both of which were assessed at the time of admission. The first modality was a total of 26 pre-existing comorbidities and symptoms, referred to as "clinical features" hereinafter. These features included gender, age, hypertension, coughing, and different types of fever. A detailed description of these 26 features is provided in Table S1 in Multimedia Appendix 1. All clinical features were coded as 0-1 binary variables (age was dichotomized using 50 years as the threshold).

In addition, we collected the patients' laboratory test results. The laboratory tests were plasma, serum, or whole blood assays for commonly obtained biochemistry tests, complete blood counts with differential counts and percentages, immunologic markers, such as interleukin 6 (IL-6), dimerized plasmin fragment D (D-dimer) and high-sensitivity C-reactive protein (hsCRP). After initial screening, several features with too many missing data, such as calcitonin, were excluded. In addition, respiratory rate, oxygen saturation, and PaO_2 without supplemental oxygen were excluded because they are type-defining features according to the official National Diagnosis and Treatment Plan of China [9]. We used 26 laboratory test features in this study. Detailed descriptions and units of these features are provided in Table S2 in Multimedia Appendix 1. All these laboratory testing features were continuous features, in contrast to the binary features used in the clinical feature modality.

Patient-specific identifying information (eg, name and address of residence) was removed from the data collected for this study. This study was evaluated and approved by the IRB committee of Union Hospital, Wuhan, China (approval number: 2020-IEC-J-345).

Data Mining on Multimodal Features

Initial data mining on the multimodal COVID-19 data was conducted. The patients' clinical data were complete. Approximately 5% of the laboratory testing data were missing. Predictive mean matching (PMM) was applied to impute the missing data. To evaluate the effectiveness of PMM, we used a subset of the original data set with no data missing, randomly dropped 5% data to simulate potential data loss, re-extrapolated the data with PMM, and evaluated the root mean square error (RMSE) between the original and imputed data sets. The RMSE was less than 0.05, indicating that the extrapolation was feasible and reliable. The imputed data were then passed on to successive data mining and ML steps.

The prevalence of each clinical feature was calculated as the number of positive test results divided by the number of patients in the nonsevere and severe groups as defined by the Diagnosis and Treatment Plan [9]. The *z* test was applied to detect any statistically significant differences in the features between the two types. In addition, a forest plot of the odds ratios (ORs) and 95% confidence intervals of the clinical features between severe and nonsevere COVID-19 types was graphed.

For the continuous laboratory testing features, we characterized and contrasted the distribution of each feature between the two types. Because the values of most features were not normally distributed, we applied a 2-sided Kolmogorov-Smirnov test instead of the Student *t* test to determine whether distributions of the feature values differed significantly between the two clinical types.

COVID-19 Clinical Type Classification via ML

Commonly used hypothesis-driven parametric models rely heavily on human decisions of how features interact with each other (eg, interaction terms in the logistic regression model), which may not reflect the underlying medical reality. In addition, these models have strict prerequisites to perform correctly, including normality of residuals, homoscedasticity, and independence of input features. Our initial exploratory analyses showed the that input features in both the clinical and laboratory testing modalities had nonnormality and high collinearity among the features. Another technical challenge to logistic regression in this study was the mixture of binary clinical and continuous laboratory testing input features.

Due to these problems, logistic regression would not be a preferred modeling approach to accurately classify and predict COVID-19 clinical types. Our exploratory analysis showed that logistic regression could only achieve average predictive accuracies of 68% and 77% on an 80-20 training-testing split

using clinical and laboratory testing feature data sets, respectively (Table S3, Multimedia Appendix 1). Thus, logistic regression is less feasible in clinical settings, where high accuracy, sensitivity, and specificity are required to differentiate COVID-19 clinical types.

On the other hand, state-of-the-art ML classification models work directly with data to avoid possible human bias. In addition, ML models do not have restrictions on how input data should be distributed or related. Therefore, in this study, we determined that ML classification would be a more appropriate modeling approach to predict COVID-19 clinical type with a complicated data structure. We developed an end-to-end ML analytical framework to accurately predict the clinical type of patients with COVID-19 based on clinical and laboratory testing modality features. We built random forest (RF) classification models, as RF enables excellent interpretability of the relative importance of an input variable to provide a more comprehensive understanding of the pathobiology of COVID-19. RF is a widely used ML model based on decision theory and the decision tree approach. Due to the internal validation process with out-of-bag error measurement, RF is especially accurate and reliable. Unlike other commonly used ML models (eg, support vector machine or k-nearest neighbor), which usually require a separate cross-validation set, the RF model performs internal validation and is especially suitable when the data set is not large. In addition, RF is robust against data loss and data unbalancing (eg, there are more patients with nonsevere than severe disease in our study [25-29]). Because the major goal of this study was not to compare the performance of different ML models, we focused on RF to deliver the most accurate classification possible.

For the single modality RF model, we used 50% randomly selected data for both clinical and laboratory testing blood biochemistry features. In this step, 107 patients with nonsevere COVID-19 and 74 patients with severe COVID-19 were randomly chosen, while the other patients' information was held to build the multimodal RF model. We assigned severe cases as "positive" and nonsevere cases as "negative" in the classification. The goal of ML classification through RF was to accurately predict the patient's COVID-19 type, either positive (severe) or negative (nonsevere), based on features from different clinical modalities. In this part of the study, we first used a single modality of features, either clinical or laboratory testing, as the input. The detailed RF modeling and validation processes are provided in Multimedia Appendix 2. We trained the model with 100 independent runs; in each run, a different set of 80% of the data was randomly selected for training, while the remaining 20% of the data were held for testing only. This step was performed to explore whether the RF model was robust against different input data and to assess the generalizability of the model. Hyperparameters in this RF model include using Gini impurity to determine the decision tree split, a minimum of 2 samples for tree spit, a minimum of 1 sample at any leaf node, and a total of 8 trees for the model ensemble [26]. Important ML performance metrics, including accuracy, sensitivity, specificity, F1 score, and area under the curve (AUC) value based on the receiver operating characteristic (ROC) curve, were computed for the testing set only.

In addition, RF can evaluate the relative importance of the input variables based on their Gini importance scores [28]. We further quantified the Gini impurity importance scores of the input features in each RF run during the model development stage on the training set (80% randomly selected data). We then identified the top contributing features based on the Gini importance scores in each of the 100 runs, aggregated over 100 runs; identified the overall top contributing features; and explored the clinical relevance and interpretability of these features for COVID-19. Note that the Gini impurity importance was calculated from the RF model based on the training set only and not on the testing set. In addition, each run of the RF model was based on a completely different, randomly sampled, and independent set of 80% training data, from which the Gini importance was calculated. Therefore, this approach avoided potential issues of overfitting and inflated performance [30]. If an RF model is robust, important input features should be consistent with the different 80% portions of the data used as the training set to develop the model. The most important features to differentiate COVID-19 clinical types were also cross-checked with our results from exploratory data mining, including the prevalence of the clinical features and the distribution of the laboratory testing features.

COVID-19 Clinical Type Classification With Multimodal ML

More importantly, we explored whether and how combining features across feature modalities improved classification performance. We developed another RF model with the same hyperparameter setting that incorporated features from both modalities. The modeling process using a single modality was similar. Instead of putting all 52 features into the model, we selected only the top 5 features from each of the two modalities as new inputs to reduce the increase model feasibility in case certain features in the total 52-feature pool would not be readily available. These top features were identified from the Gini importance of the single modality RF models (highlighted in Table S1 and Table S2 in Multimedia Appendix 1). The data set for developing the multimodal RF model was a completely new data set, as in, the other 50% of the original data was based on 107 additional patients with nonsevere COVID-19 and 74 additional patients with severe COVID-19 whose data were not used in the development of the single modality RF model.

We explored whether only 10 important features from different modalities could perform sufficiently well to address the clinical challenge of differentiating COVID-19 clinical types. This study can serve as alternative and supplemental tool to the gold standard features, which may not be readily available at the time of diagnosis.

All statistical analyses and ML models were built in R 4.0.2 (R Project) and Python 3.7 with additional supporting packages. The complete codes and fully deidentified data are freely available on GitHub [31].

Results

Clinical Findings in Nonsevere and Severe COVID-19

The prevalence of clinical features in patients with nonsevere and severe COVID-19 at the time of entry into the study were calculated and compared (Figure 1). A more detailed comparison of the clinical features between the two COVID-19 clinical types is provided in Table S1 in Multimedia Appendix 1, which shows the ORs, confidence intervals, and associated *P* values. For patients with the two clinical types of COVID-19, the prevalence was distinct for a number of different features. Patients with severe COVID-19 were statistically much more likely to be older (aged ≥50 years, OR 13.77, 95% CI 7.33-25.86, *P*<.001) and male (OR 1.89, 95% CI 1.24-2.90, *P*=.003) and to have renal diseases (OR 8.51, 95% CI 1.86-38.99, *P*<.001), cardiovascular diseases (OR 5.61, 95% CI 2.81-11.20, *P*<.001), hypertension (OR 5.37, 95% CI 3.36-8.56, *P*<.001), diabetes (OR 4.61, 95% CI 2.53-8.38, *P*<.001), loss of appetite and taste (OR 3.20, 95% CI 1.70-6.01, *P*<.001), chills (OR 2.21, 95% CI 1.16-4.22, *P*=.01), and chest congestion (OR 1.88, 95% CI 1.22-2.89, *P*=.003) than their counterparts with nonsevere COVID-19. The only exception was sore throat, which patients with severe COVID-19 were significantly much less likely to develop (OR 0.30, 95% CI 0.14-0.61, *P*<.001). These discoveries are further demonstrated in the forest plot of the ORs and confidence intervals in Figure 2, which shows the differences between the two clinical types. Therefore, these relatively easily measured and acquired clinical features could be used to clinically evaluate the disease severity of patients with COVID-19. Our findings, especially for patients with severe COVID-19, echoed the US Centers for Disease Control and Prevention's recently updated list of symptoms of COVID-19 [32] and more recent characterizations of patients with COVID-19 in the United States [33]. Our findings showed that older male patients with COVID-19 who had cardiovascular disease, respiratory disease, renal disease, and diabetes were at much higher risk of developing serious complications of COVID-19, such as acute respiratory distress syndrome (ARDS) and even death [20,21]. In addition, we discovered that Chinese patients with renal diseases were significantly more likely to develop severe COVID-19, which has not been widely reported. Clinical evidence has shown that ACE-2 expression is associated with kidney diseases; thus, kidney disease is a potential complication of SARS-CoV-2 infection [34,35]. This finding would inform clinicians that they should also monitor kidney dysfunction, such as acute kidney injury, as a clinical sign or consequence of severe COVID-19 complications.

Figure 1. Comparison of clinical features of patients with nonsevere and severe COVID-19. Note that because these features were binary, the y-axis indicates the prevalence of a positive result. CAR: cardiovascular disease; CHL: chills and shaking; CNC: cancer; CON: contact with patients with COVID-19; COU: coughing; CPD: chronic obstructive pulmonary disease; DIA: diabetes; DIR: diarrhea; FAM: family members with COVID-19; FEV: fever; FTG: fatigue; HED: headache; HIF: high fever; HYP: hypertension; KID: renal disease; LOF: low fever; MOF: medium fever; MSA: muscle ache; MUC: phlegm; NAP: loss of appetite; OLD: older age; PREV: prevalence; SEX: male sex; SHB: chest congestion; SMK: history of smoking; SOR: sore throat; VOM: vomiting.

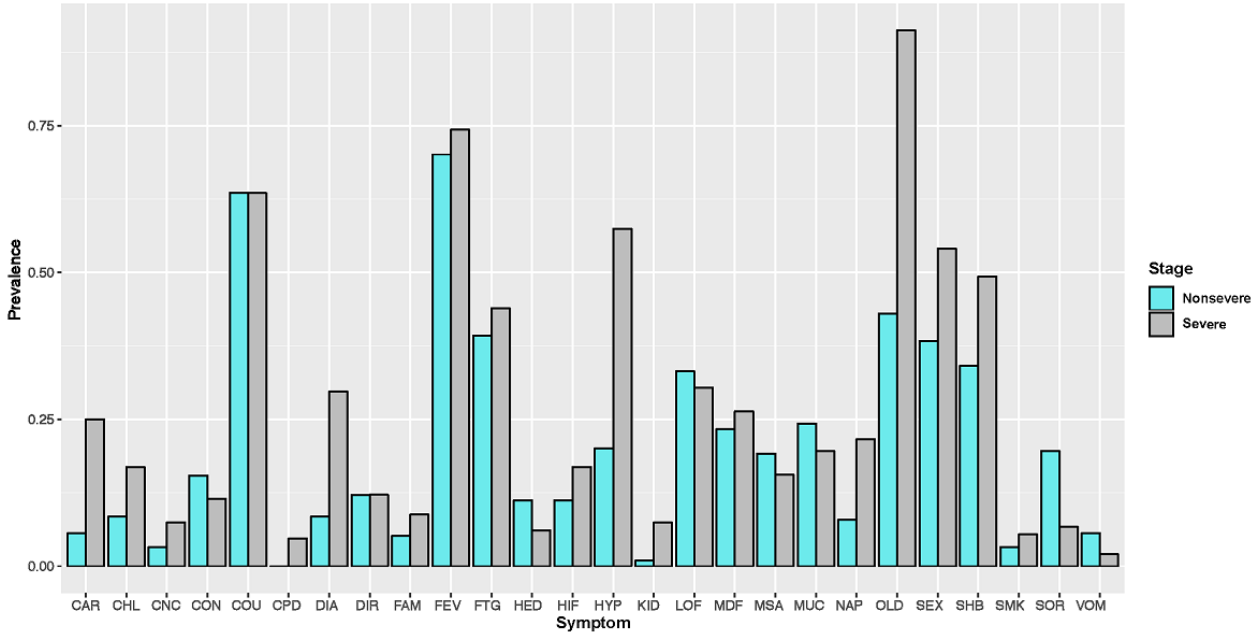
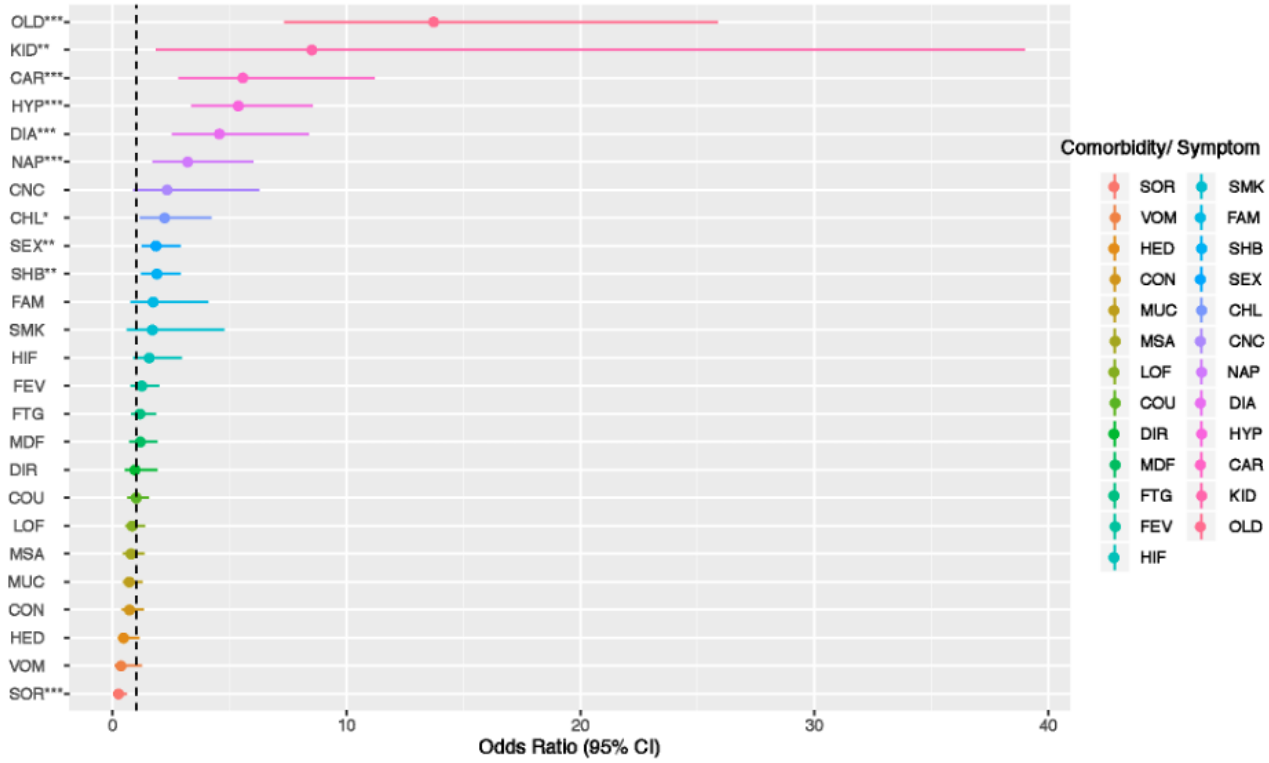
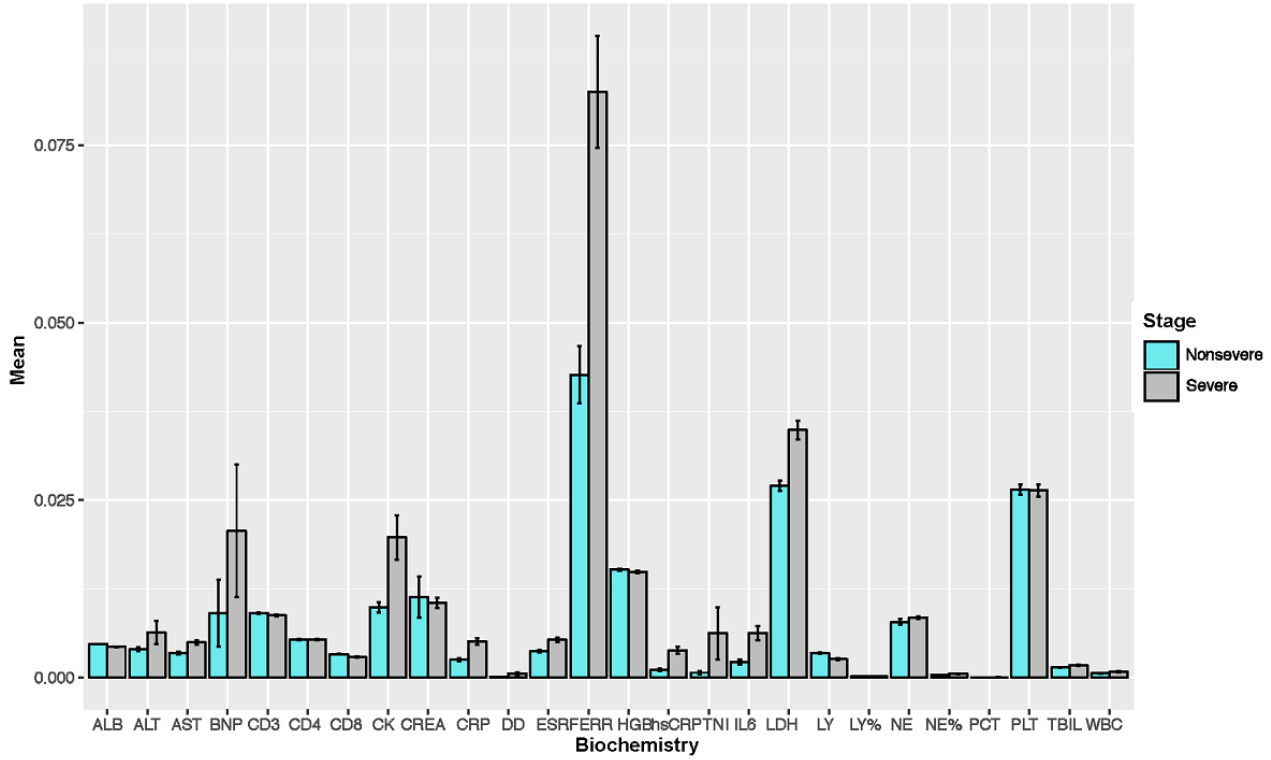


Figure 2. Forest plot of the importance of clinical features of patients with nonsevere and severe clinical types of COVID-19. Chronic obstructive pulmonary disease is not included because only patients with severe COVID-19 showed it as a comorbidity. The threshold for a feature to be positively or negatively associated with severe COVID-19 was 1 (dashed line), not 0. CAR: cardiovascular disease; CHL: chills and shaking; CNC: cancer; CON: contact with patients with COVID-19; COU: coughing; CPD: chronic obstructive pulmonary disease; DIA: diabetes; DIR: diarrhea; FAM: family members with COVID-19; FEV: fever; FTG: fatigue; HED: headache; HIF: high fever; HYP: hypertension; KID: renal disease; LOF: low fever; MOF: medium fever; MSA: muscle ache; MUC: phlegm; NAP: loss of appetite; OLD: older age; SEX: male sex; SHB: chest congestion; SMK: history of smoking; SOR: sore throat; VOM: vomiting. *P<.05, **P<.01; ***P<.001 from the 2×2 contingency table for each feature.



For the laboratory testing modality features, we compared the distributions of the continuous features between nonsevere and severe COVID-19. The results are demonstrated in Figure 3. A more detailed comparison of these 26 laboratory testing features between the two clinical types is provided in Table S2 in Multimedia Appendix 1, which shows the *P* values from the Kolmogorov-Smirnov tests. Based on the 2-sided Kolmogorov-Smirnov test, severe and nonsevere COVID-19 types differed significantly in most laboratory features, except for platelet (PLT), hemoglobin (HGB), CD3, and CD4. Among all laboratory features, IL-6, high-sensitivity troponin I (hsTNI), and D-dimer had the most significant differences between nonsevere and severe COVID-19 types.

Figure 3. Comparison of laboratory testing features of patients with nonsevere and severe COVID-19. Values shown on the y-axis were obtained after feature scaling and are between 0 and 1. The error bars represent the standard error of each laboratory testing feature. ALB: albumin; ALT: alanine transaminase; AST: aspartate aminotransferase; BNP: Brain natriuretic peptide; CK: creatine kinase; CREA: creatinine; CRP: C-reactive protein; DD: dimerized plasmin fragment D; ESR: erythrocyte sedimentation rate; FERR: ferritin; HGB: hemoglobin; hsCRP: high-sensitivity C-reactive protein; TNI: troponin I; IL6: interleukin 6; LDH: lactate dehydrogenase; LY: lymphocyte; LY%: percent of lymphocytes; NE: neutrophil; NE% percent of neutrophils; PCT: procalcitonin; PLT: platelet; TBIL: total bilirubin; WBC: white blood cell.



In conclusion, after extensive clinical feature extraction and data mining, we obtained strong qualitative and quantitative evidence that nonsevere and severe COVID-19 types differ substantially with regard to clinical features and laboratory test results. These findings pave the way toward creating an effective ML classifier to accurately differentiate these two COVID-19 types in clinical practice.

Clinical Type Classification via ML

Comorbidity and Symptom (Clinical) Modality

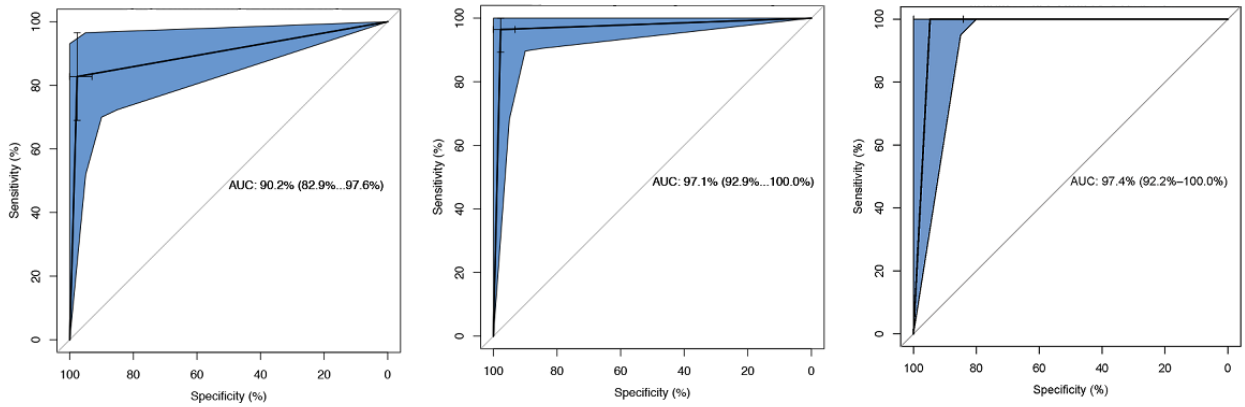
We first explored whether relatively simple binary features could provide accurate insights in identifying COVID-19 disease severity. The performance of this model is summarized in the upper section of Table 1. Based on 100 independent runs, the RF model reached a median of >99% and 94% accuracy for the training and testing sets, respectively (Table 1). Median is reported instead of mean value because the performance metrics were not normally distributed. The AUC was 90.2% (range 82.9%-97.6%) based on the ROC curve (Figure 4, left panel). The model performed better in detecting true positives (ie, severe clinical type) than true negatives (ie, nonsevere type). In other words, clinical features alone in the RF models were very unlikely to misclassify a severe case as a nonsevere case but had a higher likelihood of predicting a nonsevere case to be a severe case. In clinical practice, this would be a lesser concern, as a false positive (failure to detect nonsevere type) would be more tolerable than a false negative (failure to detect severe type).

Table 1. Performance of the random forest model with multimodal features. The results are based on 100 runs. In each run, 80% of the data was randomly selected as the training set and 20% as the testing set. The table shows the model performance on the testing set only.

Feature and performance metric	Median	Minimum	Maximum
Clinical (%)			
Accuracy	94.59	81.08	>99
Sensitivity	>99	80.95	>99
Specificity	93.75	75.00	>99
F1 score	97.30	82.93	>99
AUC[a]	90.20	82.90	97.60
Laboratory testing (%)			
Accuracy	97.22	93.06	>99
Sensitivity	>99	94.59	>99
Specificity	96.97	83.33	>99
F1 score	97.89	94.74	>99
AUC	97.10	92.90	>99
Multimodal (%)			
Accuracy	97.22	91.67	>99
Sensitivity	>99	90.00	>99
Specificity	94.44	75.00	>99
F1 score	97.78	97.22	>99
AUC	97.40	92.20	>99

[a]AUC: area under the curve

Figure 4. ROC curves from the random forest models based on clinical, laboratory testing, and multimodal features. Left: the symptom feature as the sole input; middle: the laboratory testing feature as the sole input; right: both features combined as the input. AUC: area under the curve; ROC: receiving operator characteristic.



Our RF model also identified the major influential features to differentiate COVID-19 types based on their contributions to the Gini importance in the training set. The top influential clinical features, in descending order, were age, gender, hypertension, diabetes, and cardiovascular diseases, in accordance with existing literature reports [36]. Other important clinical features included fatigue, chest congestion, sore throat, phlegm, and fever. Most of these findings aligned well with our parametric data mining with OR comparison (Figure 2, Table S1 in Multimedia Appendix 1) but showed much higher accuracy (94% accuracy on the testing set of the RF model compared to 68% accuracy from logistic regression). The only exception was renal disease, which was not considered to be a major differentiating factor based on its Gini importance (Table S1, Multimedia Appendix 1).

Clinically, older male patients with pre-existing comorbidities, especially hypertension, diabetes, and cardiovascular diseases, are much more vulnerable to COVID-19 and have a much higher risk of developing severe disease [19,21]. Therefore, we suggested using the comorbidity and symptom features of patients with COVID-19 as the first round of evaluation of severity with reasonable accuracy.

Laboratory Testing Modality

The RF model with 26 laboratory testing features was highly effective in differentiating nonsevere and severe COVID-19. The RF model achieved >99% and >95% accuracy for the training and testing data sets, respectively. The sensitivity, specificity, and F1 scores were all >95% when using only 8 trees in the RF model (Table 1, middle section). The AUC was 97% based on the ROC curve (Figure 4, middle panel). Although this study focused on ML methods, we evaluated the model performance of non-ML logistic regression in Table S3 (Multimedia Appendix 1) as a reference point to show the improvement that state-of-the-art ML models could achieve.

The top differentiating features in the laboratory testing modality were D-dimer, hsTNI, neutrophil, IL-6, lactate dehydrogenase (LDH), and hsCRP, in descending order. The clinical interpretation of their important roles was that patients with severe COVID-19 experience more intense immune responses and hyperinflammation, such as cytokine storm syndrome, with substantially increased IL-6 [37]. Research has also shown that SARS-CoV-2 can infect many organs other than the lungs, including the heart, and induce dysfunction of these organs [38,39]. Increasing hsTNI was found to be a sign of heart tissue damage from SARS-CoV-2 infection [40]. In addition, patients with severe COVID-19 may have microthrombosis, which induces higher D-dimer levels [19,21,41-43]. Abnormal levels of neutrophils may be responsible for cytokine storms and ARDS in patients with severe COVID-19 [13,44]. hsCRP, a biomarker of acute inflammation, cardiovascular disease, and ischemic events, was also confirmed to be a major contributing factor of COVID-19 mortality [19]. LDH is a biomarker of tissue damage and has been used to predict the clinical course of patients with COVID-19 [45]. These findings add further clinical insights to how multiple organs and systems, not just the lungs, respond to SARS-CoV-2 infection in different clinical types [14,46,47].

Multimodal Features

We further developed a multimodal RF model that incorporated both clinical and laboratory testing modalities with a completely new data set that was not used for the single modality model development. We used only the 5 most important features from the clinical and laboratory testing modalities, based on their Gini importance scores. The results showed that the top 10 of a total of 52 features from both modalities achieved almost >95% in every model performance metric, including accuracy, sensitivity, specificity, and F1 score (Table 1). The AUC was >97% as well (Figure 4, right panel). Therefore, we concluded that a two-step evaluation and triaging process would be feasible to differentiate the clinical types of patients with COVID-19 when the gold standard type-defining features were not readily available.

These findings reflect our clinical understanding that SARS-CoV-2 attacks multiple organs and systems, and the human body reacts in a unity against infection. Different features (eg, comorbidity, symptom, and laboratory testing results) complemented each other to provide a more comprehensive characterization of how the human body as a united entity, not only the respiratory system, reacted to SARS-CoV-2 infection [14]. In addition, the decent model performance supports the feasibility of multimodal data mining in detecting and differentiating patients with nonsevere COVID-19 from patients with severe disease.

Comparing the original 52 features in both modalities, which may not be all available at the same time during COVID-19 diagnosis, the top 10 most differentiating multimodal features provided a more practical input combined with the highly accurate ML model. Therefore, we concluded that our work would help effectively optimize health care operations during the pandemic and avoid overloading of the health care system [8].

Discussion

Principal Findings

This study provides a novel analytical framework that combines the power of multiple clinical features from different modalities to differentiate COVID-19 clinical types via ML techniques. Practically, it enables the delivery of a more comprehensive understanding of the pathobiology of COVID-19. It can aid the development of optimal treatment plans for individual patients, such as sending them to a mobile cabin hospital or admitting to a hospital with an ICU [7]. In addition, it will enable more effective triaging and optimization of health care system resources and personnel. This will substantially reduce the risk of overloading the health care system by admitting all patients with COVID-19 to hospital, decrease potential health care–associated infections, and improve clinical outcomes for the patients, especially during the COVID-19 pandemic [8].

In addition to accurately detecting vulnerable patients with COVID-19 who are likely to have severe disease, this study also provides insights on why these patients may have severe disease. ML models work directly with data and therefore are generally not good at providing clear interpretations. In this study, we combined the power of both hypothesis-driven and data-driven ML models. The highest-contributing comorbidities, symptoms, and biochemical features help predict and explain potential COVID-19 clinical courses and prognoses. Our research echoes recent studies that characterize and predict the clinical course, critical illness, and mortality of patients with COVID-19 [13,19,21]. In particular, another decision tree–based algorithm, extreme gradient boosting (XGBoost), showed promising performance in predicting the mortality of patients with COVID-19 [19]. RF is technically similar to XGBoost, and our results were consistent in identifying the key differentiating features, including LDH and hsCRP.

A continuous-valued risk score calculator for predicting risk of transitioning to critical-type COVID-19 (an even more severe type that requires ICU hospitalization, an invasive ventilator, or extracorporeal membrane oxygenation, and has a mortality rate as high as 50%) has been developed for patients with COVID-19 [21]. As a comparison, although our RF model predicts a 0-1 binary outcome for nonsevere and severe type disease, the internal RF modeling process through decision tree approach actually calculates an intermediate score between 0 and 1. By using a cutoff threshold, the RF model reports a final

dichotomized 0-1 outcome. Therefore, our analytical framework can also be readily adjusted to provide a continuous risk score for clinical evaluation and triaging of patients with COVID-19, if needed.

Many patients with severe COVID-19 present symptoms in lungs, especially ground-glass opacity (GGO), which can be detected by biomedical imaging techniques such as computed tomography (CT). However, a major clinical challenge of COVID-19 lies in the asymptomatic patient problem, which creates far worse difficulties than other coronavirus epidemics, including the original SARS and MERS epidemics. These patients show few or no classic symptoms related to viral pneumonia, and they present no GGO; however, they are almost as capable of transmitting the virus as symptomatic patients [4-6]. We suggest that the term "asymptomatic" is used due to lack of a comprehensive characterization and understanding of this novel pathogen and the pathophysiology of the host; we also suggest that these patients are not truly "asymptomatic," as in, without any clinical symptoms or signs. Approximately 10% of the patients with mild COVID-19 in our study cohort did not show typical respiratory symptoms, including fever, coughing, and chest pain, upon admission. However, they showed other symptoms from the more comprehensive modality of 26 clinical features.

Future Work

The next step of this study is to further include a biomedical imaging modality. A technical barrier is that a CT scan is a high-dimensional feature set, while clinical and laboratory test data have relatively low dimensionality. Therefore, the CT scan, in its original form of imaging, cannot be effectively combined with other modalities. We will evaluate the feasibility of using a convolution neural network (CNN, another ML technique) first to reduce the feature space in CT scans and extract a fully connected layer in the CNN as a representation of the CT scan feature. A fully connected layer is a 1D vector and has the same dimensionality as the other two modalities. Therefore, in theory, we would be able to further combine CT scans with other clinical features and investigate the association between these features with regard to COVID-19.

COVID-19 is a complex disease in which the pathogen not only attacks the respiratory system but other organs and systems that possess ACE-2 receptors as well [14,33]. Our findings reveal the complicated pathological, physiological, and immunological responses to SARS-CoV-2 infection and shed light in understanding the complex interactions between the virus and the human body. Although our multimodal data mining and ML framework was developed with data from patients with severe and nonsevere COVID-19, we suggest that the end-to-end framework is applicable to many disease systems in which multimodal inputs are common, including demographic information, comorbidity, laboratory testing, imaging, and -omics data. Having a more holistic viewpoint and approach will enable us to understand and respond to these emerging diseases, especially the unprecedented COVID-19, more readily in the field. Another feasible analytical solution is ensembling. Each input feature modality can be used independently to train a specific model, and the final prediction of COVID-19 clinical type can be made through ensembling. We will further explore this analytical framework and transfer our insights to future clinical studies, such as differentiating healthy patients from patients with non–COVID-19 viral pneumonia, nonsevere COVID-19, and severe COVID-19.

Limitations

In this study, we recruited participants from a single hospital in Wuhan, the first epicenter of COVID-19. There will inevitably be selection bias, as the ethnic group is currently limited to Chinese participants who are mostly of Han ethnicity. It is possible that ethnicity and race and their confounding risk factors (eg, socioeconomic status, nutrition conditions, accessibility of care, and other social determinants of health) are different in various studies. Therefore, we wish to share our findings with our colleagues worldwide and determine whether different demographic backgrounds influence feature distributions between nonsevere and severe COVID-19 in patients. Some of our findings of the top contributing clinical and laboratory testing features were supported in other COVID-19 studies across different ethnic groups, while others were not [32,40]. For example, while we found male gender to be a strong influencing factor of severe COVID-19, other studies did not reach a similar conclusion [36]. The findings in this study on Chinese ethnicity could actually complement other existing studies on other ethnic groups and reveal the clinical and epidemiological complexity of this unprecedented ongoing pandemic.

Another limitation of this study is that the patients were evaluated at the time of admission; therefore, the study was a cross-sectional instead of a longitudinal study. Future studies could examine both diagnosis and prognosis and further explore how and why some patients with nonsevere COVID-19 may transition to a severe disease state and whether ML techniques are able to identify critical predictive features to undesirable prognoses such as death.

Additionally, different subtypes of SARS-CoV-2, their specific pathogenicity and virulence, and their host-pathogen interactions should be taken into consideration when conducting and comparing studies across different regions of the world. The other factors that this study did not include are behavioral and societal aspects, such as whether and how using mobile cabin hospitals to treat patients with nonsevere COVID-19 reduces the rate of transition to severe type. The COVID-19 epidemic, like all infectious disease epidemics, has individual clinical, epidemiological, behavioral and societal factors. Therefore, we will also explore cross-scale individual clinical course and population-level epidemics in future studies.

Conclusion

We trained and validated ML RF models to predict COVID-19 severity based on 26 comorbidity and symptom features and 26 laboratory testing features from a cohort of 214 patients with nonsevere COVID-19 and 148 patients with severe COVID-19. We identified the top features from both feature modalities to differentiate the clinical types, and we achieved predictive accuracies of >90%, >95%, and >99% when clinical features, laboratory test data, and the top 5 features from each modality

combined were used as inputs, respectively. The results will help patients with COVID-19 self-evaluate their condition, help clinicians to evaluate disease severity and triage patients, and optimize health care resource utilization during the COVID-19 pandemic.

Acknowledgments

This study is supported by the National Science Foundation for Young Scientists of China (81703201, 81602431, and 81871544), the North Carolina Biotechnology Center Flash Grant on COVID-19 Clinical Research (2020-FLG-3898), the Natural Science Foundation for Young Scientists of Jiangsu Province (BK20171076, BK20181488, BK20181493, and BK20201485), the Jiangsu Provincial Medical Innovation Team (CXTDA2017029), the Jiangsu Provincial Medical Youth Talent program (QNRC2016548 and QNRC2016536), the Jiangsu Preventive Medicine Association program (Y2018086 and Y2018075), the Lifting Program of Jiangsu Provincial Scientific and Technological Association, and the Jiangsu Government Scholarship for Overseas Studies.

Authors' Contributions

YC designed the study. LO and JL derived and processed the data. LO, FSB, QL, LH, BZ, JL, PR, and SC interpreted the results. MX and SC performed analyses. YG and SC supervised the study. PR and SC developed the manuscript. YC, LO, and FSB contributed equally. JL (liu_jie0823@163.com), MX (sosolou@126.com), and SC (schen56@uncc.edu) serve as corresponding authors of this study equally.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary tables S1-S3.
[DOCX File , 128 KB - jmir_v23i4e23948_app1.docx]

Multimedia Appendix 2
Supplementary information regarding the model development and complete feature distributions.
[DOCX File , 25 KB - jmir_v23i4e23948_app2.docx]

References

1. Weekly epidemiological update - 12 January 2021. World Health Organization. 2021 Jan 12. URL: https://www.who.int/publications/m/item/weekly-epidemiological-update---12-january-2021 [accessed 2021-03-18]
2. Gandhi M, Yokoe DS, Havlir DV. Asymptomatic transmission, the Achilles' heel of current strategies to control covid-19. N Engl J Med 2020 May 28;382(22):2158-2160. [doi: 10.1056/nejme2009758]
3. Wölfel R, Corman VM, Guggemos W, Seilmaier M, Zange S, Müller MA, et al. Virological assessment of hospitalized patients with COVID-2019. Nature 2020 May;581(7809):465-469 [FREE Full text] [doi: 10.1038/s41586-020-2196-x] [Medline: 32235945]
4. Luo Y, Trevathan E, Qian Z, Li Y, Li J, Xiao W, et al. Asymptomatic SARS-CoV-2 Infection in Household Contacts of a Healthcare Provider, Wuhan, China. Emerg Infect Dis 2020 Aug;26(8):1930-1933 [FREE Full text] [doi: 10.3201/eid2608.201016] [Medline: 32330112]
5. Ye F, Xu S, Rong Z, Xu R, Liu X, Deng P, et al. Delivery of infection from asymptomatic carriers of COVID-19 in a familial cluster. Int J Infect Dis 2020 May;94:133-138 [FREE Full text] [doi: 10.1016/j.ijid.2020.03.042] [Medline: 32247826]
6. Wu Z, McGoogan JM. Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China: summary of a report of 72,314 cases from the Chinese Center for Disease Control and Prevention. JAMA 2020 Apr 07;323(13):1239-1242. [doi: 10.1001/jama.2020.2648] [Medline: 32091533]
7. Chen S, Zhang Z, Yang J, Wang J, Zhai X, Bärnighausen T, et al. Fangcang shelter hospitals: a novel concept for responding to public health emergencies. Lancet 2020 Apr;395(10232):1305-1314. [doi: 10.1016/s0140-6736(20)30744-3]
8. Truog RD, Mitchell C, Daley GQ. The toughest triage - allocating ventilators in a pandemic. N Engl J Med 2020 May 21;382(21):1973-1975. [doi: 10.1056/NEJMp2005689] [Medline: 32202721]
9. Novel coronavirus pneumonia diagnosis and treatment plan (provisional 7th edition). Webpage in Chinese. National Health Commission of China. 2020 Mar 04. URL: http://www.nhc.gov.cn/yzygj/s7652m/202003/a31191442e29474b98bfed5579d5af95.shtml [accessed 2021-03-18]
10. Metlay JP, Waterer GW, Long AC, Anzueto A, Brozek J, Crothers K, et al. Diagnosis and treatment of adults with community-acquired pneumonia. An official clinical practice guideline of the American Thoracic Society and Infectious Diseases Society of America. Am J Respir Crit Care Med 2019 Oct 01;200(7):e45-e67. [doi: 10.1164/rccm.201908-1581st]

11. Neill AM, Martin IR, Weir R, Anderson R, Chereshsky A, Epton MJ, et al. Community acquired pneumonia: aetiology and usefulness of severity criteria on admission. Thorax 1996 Oct 01;51(10):1010-1016 [FREE Full text] [doi: 10.1136/thx.51.10.1010] [Medline: 8977602]
12. The Novel Coronavirus Pneumonia Emergency Response Epidemiology Team. Vital surveillances: the epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (COVID-19) — China, 2020. China CDC Wkly 2020;2(8):113-122. [doi: 10.46234/ccdcw2020.032]
13. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet 2020 Feb;395(10223):497-506. [doi: 10.1016/s0140-6736(20)30183-5]
14. Wadman M. How does coronavirus kill? Clinicians trace a ferocious rampage through the body, from brain to toes. Science 2020 Apr 18. [doi: 10.1126/science.abc3208]
15. Liu H, Wang R, Qu G. A general report on the systematic anatomy of COVID-19. Article in Chinese. J Forensic Med 2020;36:1-3. [doi: 10.12116/j.issn.1004-5619.2020.01.005]
16. Cholankeril G, Podboy A, Aivaliotis VI, Tarlow B, Pham EA, Spencer SP, et al. High prevalence of concurrent gastrointestinal manifestations in patients with severe acute respiratory syndrome coronavirus 2: early experience from California. Gastroenterology 2020 Aug;159(2):775-777 [FREE Full text] [doi: 10.1053/j.gastro.2020.04.008] [Medline: 32283101]
17. Jing Y, Run-Qian L, Hao-Ran W, Hao-Ran C, Ya-Bin L, Yang G, et al. Potential influence of COVID-19/ACE2 on the female reproductive system. Mol Hum Reprod 2020 Jun 01;26(6):367-373 [FREE Full text] [doi: 10.1093/molehr/gaaa030] [Medline: 32365180]
18. Moriguchi T, Harii N, Goto J, Harada D, Sugawara H, Takamino J, et al. A first case of meningitis/encephalitis associated with SARS-Coronavirus-2. Int J Infect Dis 2020 May;94:55-58 [FREE Full text] [doi: 10.1016/j.ijid.2020.03.062] [Medline: 32251791]
19. Yan L, Zhang H, Goncalves J, Xiao Y, Wang M, Guo Y, et al. An interpretable mortality prediction model for COVID-19 patients. Nat Mach Intell 2020 May 14;2(5):283-288 [FREE Full text] [doi: 10.1038/s42256-020-0180-7]
20. Liu J, Li S, Liu J, Liang B, Wang X, Wang H, et al. Longitudinal characteristics of lymphocyte responses and cytokine profiles in the peripheral blood of SARS-CoV-2 infected patients. EBioMedicine 2020 May;55:102763 [FREE Full text] [doi: 10.1016/j.ebiom.2020.102763] [Medline: 32361250]
21. Liang W, Liang H, Ou L, Chen B, Chen A, Li C, China Medical Treatment Expert Group for COVID-19. Development and validation of a clinical risk score to predict the occurrence of critical illness in hospitalized patients with COVID-19. JAMA Intern Med 2020 Aug 01;180(8):1081-1089 [FREE Full text] [doi: 10.1001/jamainternmed.2020.2033] [Medline: 32396163]
22. Wynants L, Van Calster B, Collins GS, Riley RD, Heinze G, Schuit E, et al. Prediction models for diagnosis and prognosis of covid-19 infection: systematic review and critical appraisal. BMJ 2020 Apr 07;369:m1328 [FREE Full text] [doi: 10.1136/bmj.m1328] [Medline: 32265220]
23. Shamsoddin E. Can medical practitioners rely on prediction models for COVID-19? A systematic review. Evid Based Dent 2020 Sep 25;21(3):84-86 [FREE Full text] [doi: 10.1038/s41432-020-0115-5] [Medline: 32978532]
24. Struyf T, Deeks JJ, Dinnes J, Takwoingi Y, Davenport C, Leeflang MM, Cochrane COVID-19 Diagnostic Test Accuracy Group. Signs and symptoms to determine if a patient presenting in primary care or hospital outpatient settings has COVID-19 disease. Cochrane Database Syst Rev 2020 Jul 07;7:CD013665 [FREE Full text] [doi: 10.1002/14651858.CD013665] [Medline: 32633856]
25. Quinlan J. Induction of decision trees. Mach Learn 1986 Mar;1(1):81-106. [doi: 10.1007/BF00116251]
26. Ho TK. Random decision forests. In: Proceedings of 3rd International Conference on Document Analysis and Recognition. 1995 Presented at: 3rd International Conference on Document Analysis and Recognition; 1995; Montréal, QC p. 278-282. [doi: 10.1109/icdar.1995.598994]
27. Lundberg SM, Erion G, Chen H, DeGrave A, Prutkin JM, Nair B, et al. From Local Explanations to Global Understanding with Explainable AI for Trees. Nat Mach Intell 2020 Jan 17;2(1):56-67 [FREE Full text] [doi: 10.1038/s42256-019-0138-9] [Medline: 32607472]
28. Tolosi L, Lengauer T. Classification with correlated features: unreliability of feature ranking and solutions. Bioinformatics 2011 Jul 15;27(14):1986-1994. [doi: 10.1093/bioinformatics/btr300] [Medline: 21576180]
29. Strobl C, Boulesteix A, Augustin T. Unbiased split selection for classification trees based on the Gini Index. Comput Stat Data Anal 2007 Sep;52(1):483-501. [doi: 10.1016/j.csda.2006.12.030]
30. Hastie T, Tibshirani R, Friedman J. The elements of statistical learning: data mining, inference, and prediction. 2nd edition. Berlin, Germany: Springer; 2009.
31. Bao F. Triaging moderate COVID-19 and other viral pneumonia from routine blood tests. GitHub. URL: https://github.com/forrestbao/covid-19 [accessed 2021-03-18]
32. Symptoms of coronavirus. US Centers for Disease Control and Prevention. URL: https://www.cdc.gov/coronavirus/2019-ncov/symptoms-testing/symptoms.html [accessed 2021-03-18]
33. Garg S, Kim L, Whitaker M, O'Halloran A, Cummings C, Holstein R, et al. Hospitalization rates and characteristics of patients hospitalized with laboratory-confirmed coronavirus disease 2019 - COVID-NET, 14 states, March 1-30, 2020.

MMWR Morb Mortal Wkly Rep 2020 Apr 17;69(15):458-464 [FREE Full text] [doi: 10.15585/mmwr.mm6915e3] [Medline: 32298251]
34. Mizuiri S, Ohashi Y. ACE and ACE2 in kidney disease. World J Nephrol 2015 Feb 06;4(1):74-82 [FREE Full text] [doi: 10.5527/wjn.v4.i1.74] [Medline: 25664248]
35. Lieben L. Alport syndrome: ACE2 administration slows kidney damage. Nat Rev Nephrol 2017 May 13;13(5):261-261. [doi: 10.1038/nrneph.2017.36] [Medline: 28287108]
36. Yang X, Yu Y, Xu J, Shu H, Xia J, Liu H, et al. Clinical course and outcomes of critically ill patients with SARS-CoV-2 pneumonia in Wuhan, China: a single-centered, retrospective, observational study. Lancet Respir Med 2020 May;8(5):475-481. [doi: 10.1016/s2213-2600(20)30079-5]
37. Li Y, Hu Y, Yu J, Ma T. Retrospective analysis of laboratory testing in 54 patients with severe- or critical-type 2019 novel coronavirus pneumonia. Lab Invest 2020 Jun;100(6):794-800 [FREE Full text] [doi: 10.1038/s41374-020-0431-6] [Medline: 32341519]
38. Chen L, Li X, Chen M, Feng Y, Xiong C. The ACE2 expression in human heart indicates new potential mechanism of heart injury among patients infected with SARS-CoV-2. Cardiovasc Res 2020 May 01;116(6):1097-1100 [FREE Full text] [doi: 10.1093/cvr/cvaa078] [Medline: 32227090]
39. Shi S, Qin M, Shen B, Cai Y, Liu T, Yang F, et al. Association of cardiac injury with mortality in hospitalized patients with COVID-19 in Wuhan, China. JAMA Cardiol 2020 Jul 01;5(7):802-810 [FREE Full text] [doi: 10.1001/jamacardio.2020.0950] [Medline: 32211816]
40. Li J, Han T, Woodward M, Anderson CS, Zhou H, Chen Y, et al. The impact of 2019 novel coronavirus on heart injury: a systematic review and meta-analysis. Prog Cardiovasc Dis 2020 Apr;63(4):518-524 [FREE Full text] [doi: 10.1016/j.pcad.2020.04.008] [Medline: 32305557]
41. Wu C, Chen X, Cai Y, Xia J, Zhou X, Xu S, et al. Risk factors associated with acute respiratory distress syndrome and death in patients with coronavirus disease 2019 pneumonia in Wuhan, China. JAMA Intern Med 2020 Jul 01;180(7):934-943 [FREE Full text] [doi: 10.1001/jamainternmed.2020.0994] [Medline: 32167524]
42. Guan W, Ni Z, Hu Y, Liang W, Ou C, He J, China Medical Treatment Expert Group for Covid-19. Clinical characteristics of coronavirus disease 2019 in China. N Engl J Med 2020 Apr 30;382(18):1708-1720 [FREE Full text] [doi: 10.1056/NEJMoa2002032] [Medline: 32109013]
43. Griffin DO, Jensen A, Khan M, Chin J, Chin K, Saad J, et al. Pulmonary embolism and increased levels of d-dimer in patients with coronavirus disease. Emerg Infect Dis 2020 Aug;26(8):1941-1943 [FREE Full text] [doi: 10.3201/eid2608.201477] [Medline: 32348233]
44. Barnes BJ, Adrover JM, Baxter-Stoltzfus A, Borczuk A, Cools-Lartigue J, Crawford JM, et al. Targeting potential drivers of COVID-19: neutrophil extracellular traps. J Exp Med 2020 Jun 01;217(6):e20200652 [FREE Full text] [doi: 10.1084/jem.20200652] [Medline: 32302401]
45. Zhou F, Yu T, Du R, Fan G, Liu Y, Liu Z, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. Lancet 2020 Mar;395(10229):1054-1062. [doi: 10.1016/s0140-6736(20)30566-3]
46. Zaim S, Chong JH, Sankaranarayanan V, Harky A. COVID-19 and multiorgan response. Curr Probl Cardiol 2020 Aug;45(8):100618 [FREE Full text] [doi: 10.1016/j.cpcardiol.2020.100618] [Medline: 32439197]
47. Wang T, Du Z, Zhu F, Cao Z, An Y, Gao Y, et al. Comorbidities and multi-organ injuries in the treatment of COVID-19. Lancet 2020 Mar;395(10228):e52. [doi: 10.1016/s0140-6736(20)30558-4]

Abbreviations

ACE-2: angiotensin-converting enzyme 2
ARDS: acute respiratory distress syndrome
AUC: area under the curve
CT: computed tomography
D-dimer: dimerized plasmin fragment D
GGO: ground-glass opacity
hsCRP: high-sensitivity C-reactive protein
hsTNI: high-sensitivity troponin I
ICU: intensive care unit
IL-6: interleukin 6
LDH: lactate dehydrogenase
MERS: Middle East respiratory syndrome
ML: machine learning
OR: odds ratio
PMM: predictive mean matching
RF: random forest

RMSE: root mean square error
ROC: receiver operating characteristic
SARS: secure acute respiratory syndrome
XGBoost: extreme gradient boosting

Edited by C Basch; submitted 30.08.20; peer-reviewed by F Agakov; comments to author 22.09.20; revised version received 11.10.20; accepted 11.03.21; published 07.04.21.

Please cite as:
Chen Y, Ouyang L, Bao FS, Li Q, Han L, Zhang H, Zhu B, Ge Y, Robinson P, Xu M, Liu J, Chen S
A Multimodality Machine Learning Approach to Differentiate Severe and Nonsevere COVID-19: Model Development and Validation
J Med Internet Res 2021;23(4):e23948
URL: https://www.jmir.org/2021/4/e23948
doi:10.2196/23948
PMID:33714935

Original Paper

Applying A/B Testing to Clinical Decision Support: Rapid Randomized Controlled Trials

Jonathan Austrian[1,2], MD; Felicia Mendoza[3], BA; Adam Szerencsy[1,2], DO; Lucille Fenelon[2], BSN, MSN, MHA; Leora I Horwitz[1,3], MD, MHS; Simon Jones[3], PhD; Masha Kuznetsova[3], MPH; Devin M Mann[1,2,3], MD

[1]Department of Medicine, NYU Grossman School of Medicine, New York, NY, United States

[2]Medical Center Information Technology, NYU Langone Health, New York, NY, United States

[3]Department of Population Health, NYU Grossman School of Medicine, New York, NY, United States

Corresponding Author:
Jonathan Austrian, MD
Department of Medicine
NYU Grossman School of Medicine
360 Park Avenue South
New York, NY
United States
Phone: 1 646 524 0359
Email: Jonathan.Austrian@nyulangone.org

Abstract

Background: Clinical decision support (CDS) is a valuable feature of electronic health records (EHRs) designed to improve quality and safety. However, due to the complexities of system design and inconsistent results, CDS tools may inadvertently increase alert fatigue and contribute to physician burnout. A/B testing, or rapid-cycle randomized tests, is a useful method that can be applied to the EHR in order to rapidly understand and iteratively improve design choices embedded within CDS tools.

Objective: This paper describes how rapid randomized controlled trials (RCTs) embedded within EHRs can be used to quickly ascertain the superiority of potential CDS design changes to improve their usability, reduce alert fatigue, and promote quality of care.

Methods: A multistep process combining tools from user-centered design, A/B testing, and implementation science was used to understand, ideate, prototype, test, analyze, and improve each candidate CDS. CDS engagement metrics (alert views, acceptance rates) were used to evaluate which CDS version is superior.

Results: To demonstrate the impact of the process, 2 experiments are highlighted. First, after multiple rounds of usability testing, a revised CDS influenza alert was tested against usual care CDS in a rapid (~6 weeks) RCT. The new alert text resulted in minimal impact on reducing firings per patients per day, but this failure triggered another round of review that identified key technical improvements (ie, removal of dismissal button and firings in procedural areas) that led to a dramatic decrease in firings per patient per day (23.1 to 7.3). In the second experiment, the process was used to test 3 versions (financial, quality, regulatory) of text supporting tobacco cessation alerts as well as 3 supporting images. Based on 3 rounds of RCTs, there was no significant difference in acceptance rates based on the framing of the messages or addition of images.

Conclusions: These experiments support the potential for this new process to rapidly develop, deploy, and rigorously evaluate CDS within an EHR. We also identified important considerations in applying these methods. This approach may be an important tool for improving the impact of and experience with CDS.

Trial Registration: Flu alert trial: ClinicalTrials.gov NCT03415425; https://clinicaltrials.gov/ct2/show/NCT03415425. Tobacco alert trial: ClinicalTrials.gov NCT03714191; https://clinicaltrials.gov/ct2/show/NCT03714191

(J Med Internet Res 2021;23(4):e16651) doi:10.2196/16651

KEYWORDS

AB testing; randomized controlled trials; clinical decision support; clinical informatics; usability; alert fatigue

Introduction

Clinical decision support (CDS) is a valuable feature of electronic health records (EHRs). CDS can generate several forms of decision support that have been extensively studied, including alerts, calculators, reminders, and order sets [1-3]. Successful examples of CDS have led to reductions in prescribing brand-name antibiotics [4], improved lipid management in renal transplant patients [5], improved compliance with guidelines for treating HIV [6-8], reduced ordering of tests when costs were displayed [9], and age-specific alerts that reduced inappropriate prescribing in the elderly [10-15]. Although conceptually straightforward, successful CDS alerts must deliver accurate information, in clinical context and at the point of care, and must be well-integrated into the clinical workflow. These complexities have contributed to inconsistent results [16,17] that reflect the enormous heterogeneity in system design, workflow integration, usability, simplicity, and content [18,19].

Despite the complexity of designing effective CDS and the inconsistent results, the sheer volume of CDS in modern EHRs has increased dramatically. In the past 5 years, the number of alerts released in our institution's EHR increased from 13 to 117. This volume has amplified the underlying challenges of clunky user interfaces, poor usability, and workflow integration and has fueled the rapid escalation of alert fatigue [20-22]. Alert fatigue has yielded an environment in which providers ignore clinical alerts at rates as high as 70% of the time or more [23].

Consequently, while many CDS have been effective in promoting better care, they do so with generally poor efficiency. Moreover, EHRs burden providers with tedious, cognitively draining documentation requirements, leaving little time or energy to engage with CDS tools that have the potential to improve care but are not necessary requirements to complete a visit [24]. Alert fatigue contributes to provider burnout, a highly publicized phenomenon that is a major threat to patient safety and physician satisfaction.

CDS designers' capabilities are often constrained by the limitations of commercial EHR systems and the operational challenges inherent to local customizations; new CDS tools are frequently released with minimal modification of the standard tools. Similarly, traditional randomized controlled trials (RCTs) require a long, static experimental design to determine the potential statistical advantage of the intervention. These approaches stand in contrast to how other industries launch new products and determine their value. Software development and other industries leverage rapid experimentation processes, often termed "A/B trials," to efficiently evaluate multiple design choices using live environments and users [25]. In this paradigm, there is an a priori acknowledgement that the "best" version of the product is unknown until it is empirically tested. This approach is applied to big and small design choices, from what features to include in a new smartphone app to where to place icons on a web page. To enable this process, agile procedures have been developed to quickly implement versions, assess the relative impact, and then modify the product for the next round of A/B testing. Highly agile companies like Amazon use this rapid experimentation to deliver the optimal customer experience and release code modifications every 11.7 seconds to accommodate their rapid learning [26]. We believe CDS developers can deploy A/B testing methods to evaluate many aspects of evaluation including promotion of positive outcomes, minimizing unintended consequences on safety, and poor user experience.

The inpatient influenza alert was intended to promote vaccination for eligible patients. The interruptive alert triggered at the time the nurse documented responses to the influenza screening question flowsheets. Acceptance was defined as placing the order for the influenza vaccine. The alert continued to trigger from flowsheet documentation until the patient was ordered for the vaccine or the patient no longer met eligibility criteria.

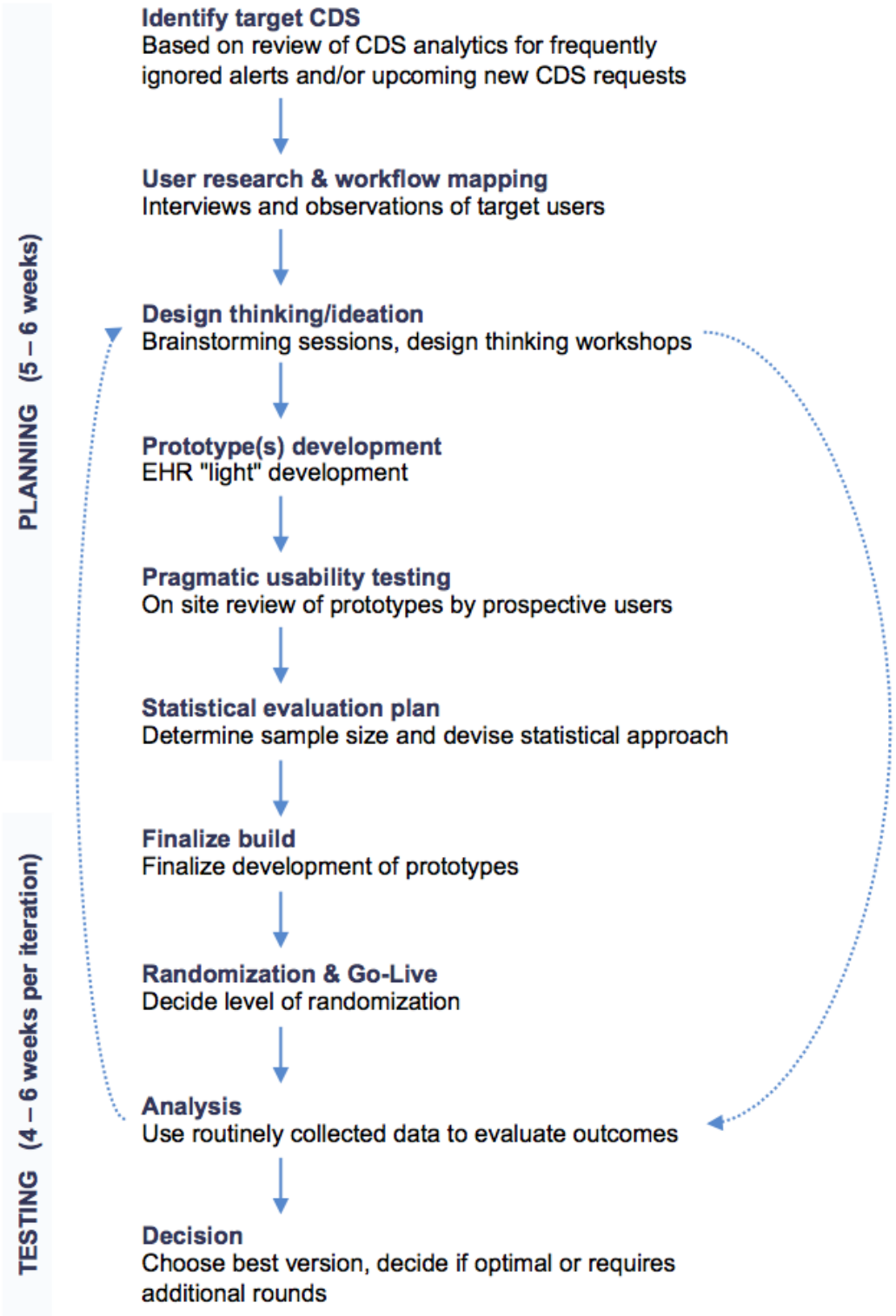
The outpatient tobacco cessation alert was designed to promote counseling and treatment for patients using tobacco products. The noninterruptive alert displayed in an alert section for outpatient providers at chart opening when active smokers had not received counseling within the previous 3 months. Acceptance was defined as documentation of counseling, prescription of therapy, or referral to the state tobacco quit line. We hypothesized that A/B testing methods would enable our CDS development teams to quickly evaluate CDS designs and iteratively modify them to maximize their acceptance rates and impact on clinical outcomes while minimizing their firing rates.

Methods

Creating an A/B Testing Framework

The CDS optimization initiative was a partnership between our CDS team and the Rapid RCT lab at NYU Langone Health [27]. The Rapid RCT lab uses elements of the traditional RCT methodology in a series of rapid cycle experiments to test and optimize systems interventions at our institution. Together, the CDS and Rapid RCT lab teams collaboratively defined a new process combining A/B methods with traditional RCTs to optimize our CDS alerts. To achieve this synergy, a multidisciplinary team was assembled including physician and nurse informaticists, implementation scientists, EHR analysts, data analyst, statistician, project manager, and project associate. A multistep process was developed that combined tools from user-centered design, A/B testing, and implementation science (Figure 1).

Figure 1. Clinical decision support (CDS)/randomized controlled trial (RCT) process map. EHR: electronic health record.



This process was supported by 3 key "enablers." One was the work done by the Rapid RCT lab to engage the local institutional review board (IRB) and create sufficient understanding of how A/B testing works so that these projects would typically be self-certified as quality improvement research and not require IRB review. This enabled our team to develop and test each iteration of a CDS alert at a rapid pace. Second, our CDS team engaged the relevant health system leadership early in the process for each alert so that there was sufficient buy-in for this novel approach to CDS development and optimization. This required multiple presentations on our process to develop trust in how it would improve the clinician experience with CDS and acceptance that there would be multiple versions of each CDS, creating heterogeneity in the user experience. This effort has been ongoing and critical to educating clinicians and administrators about this highly novel work that touches many important clinical processes. Third, our CDS team developed its technical ability to randomize experiences in the EHR at the

patient level (Multimedia Appendix 1). This work leveraged native EHR functionality. However, this logic did not provide for automated randomization at the practice level. To overcome this limitation, our team developed a process for manual randomization of clinical practices and then created logic to assign the appropriate CDS version to the appropriate target group. Fourth, we created a robust ability to report on CDS tool engagement that gave us near real-time access to the impact of each new version. Our team developed reporting capabilities beyond the standard EHR tools to drill down on what parts of each CDS tool were being "clicked" versus "ignored," what orders were being placed from what exact location, and trends in utilization. These tools allowed the team to quickly ascertain if new trends were emerging. Last, we prespecified safety metrics before each go-live. Along with our CDS outcomes, balancing clinical outcomes were evaluated to track performance and ensure patient quality and safety after each iteration. These capabilities unlocked our team's ability to scale rapid, randomized CDS experiments.

Application of the Framework

Based on the A/B testing process described in the previous section, our team was now positioned to implement our CDS A/B testing process more widely across our enterprise. We use 2 examples, our inpatient, nurse-facing influenza vaccine alert and our ambulatory, provider-facing tobacco cessation counseling alert, to detail the experimental methods involved in implementing our CDS A/B testing process.

The first step involved rapid user research. Both of these alerts were already active, with low rates of acceptance by the relevant users: 2% for the influenza vaccine alert and 31% for the tobacco cessation alert. Our process involved conducting interviews and observations (usually by our clinician informaticists and a research team member) of users interacting with the alerts using our pragmatic usability testing method to quickly summarize field notes that could be reviewed by the design team [28]. With these data in hand, the team conducted ideation sessions to develop potential alternatives to the current CDS that could enable higher adoption rates. These included modifications based on CDS best practices to reduce cognitive load (embedding images, simplifying text, consistency of formatting) [29-31] and other motivational tools ("nudges," such as highlighting financial incentives [eg, relative value units] or using more authoritative messaging) [32].

Once the team identified the versions of interest, the EHR development team created lightweight (eg, paper-based illustrations) prototypes of each version for further review by stakeholders. After iterative rounds of refinement, the prototypes were fully built within our EHR testing environment for additional usability testing and feedback from target users. With these data, the multiple CDS versions were then built and randomized to control for unmeasured confounders (eg, changes in patient volumes, staffing changes, seasonal changes), and deployed into the live environment. The inpatient influenza alert and the tobacco alert were randomized at the patient and practice level, respectively.

Statistical Analysis

Once deployed, our team used relevant CDS metrics (alert views and follow-up actions such as placing a suggested order) and the predetermined statistical approach to evaluate which CDS version was superior. Initially, we used only simple statistics (*t* tests or Chi square tests) to compare outcomes between versions. Once the testing cycle was completed, the team reconvened and evaluated if the improvement (or lack thereof) was sufficient or if additional rounds of A/B testing were warranted.

Post-hoc analysis initiated based on result irregularities in the cluster randomization tobacco experiment revealed unanticipated behaviors in the 3 groups that resulted in different baseline rates by randomization group. To control for these underlying irregularities, we modified our methods to employ a more sophisticated analysis. A multilevel logistic regression model was developed to predict alert acceptance, using the lme4 library from R [33]. Ambulatory practice was included as a random effect, and randomization group and study rounds were included as fixed effects. This method produced odds ratios that compared acceptance rates of different alert modification groups to a reference group. We obtained odds ratios from the models and profiled confidence intervals for the odds ratios using the approach of Venables and Ripley [34].

Results

Flu Alert Experiment

Our week-long user feedback sessions highlighted several issues, including confusion about the need for the ordering nurse to obtain a physician cosignature and difficulty of appropriately documenting patient refusal to prevent future triggering of the alert. This feedback was used to tailor the alert to the appropriate setting and user. We first conducted a 5-week RCT to address the most frequently mentioned nursing barrier: a misconception that ordering the influenza vaccine was outside of the nurse's scope of practice. We tested whether explicitly stating nurses were empowered to order the vaccine with no cosignature requirement would improve alert acceptance. We randomized patients into 2 arms: the existing alert (Figure 2) and the new alert that stated that nurses can order the vaccine without cosignature (Figure 3). The new message resulted in a negligible reduction in firings per patient per day.

Figure 2. Original version of the flu alert tested in the randomized controlled trial (RCT). Copyright 2020, Epic Systems Corporation.

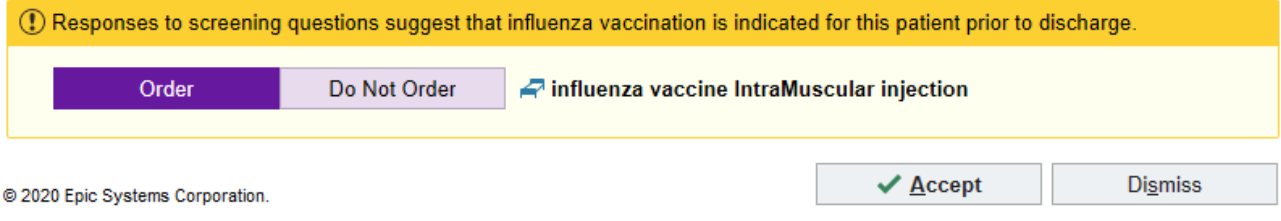
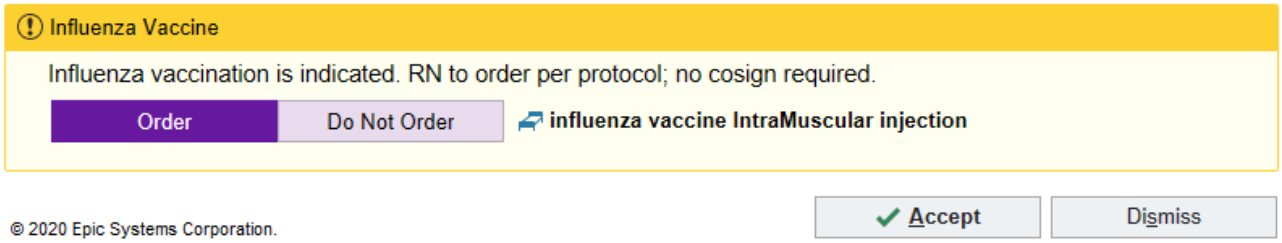


Figure 3. New version of the flu alert tested in the randomized controlled trial (RCT), with a simpler header and more directed verbiage that states "RN to order. Per Protocol; no cosign required". Copyright 2020, Epic Systems Corporation.



This limited impact encouraged the team to conduct a more intensive review of the triggering logic and user response options. This review had surfaced fundamental flaws inherent in both versions of the alert. These enhancements were prioritized, and the change most likely to succeed was implemented in version 3. Specifically, the team identified that nurses did not appreciate that when they dismissed the alert with no action, the alert would appear automatically again at next flowsheet filing. Consequently, for version 3, we eliminated the ability to "dismiss" the alert without ordering the vaccine or selecting the acknowledgment reason. When nurses attempted to rapidly satisfy the alert with "accept," the influenza order would automatically be placed, satisfying the alert and suppressing future firings (Figure 4). The CDS design team concluded that testing a new alert against an original version in a rapid RCT was no longer justifiable as we did not believe the 2 options had equipoise and did not want to delay enhancements that had high likelihood of reducing nurse fatigue. Version 3 resulted in a 64% reduction in firings per patient per day (Table 1).

Figure 4. Version 3 of the flu alert, in which an acknowledgement reason button was added and the "dismiss" button was removed. Copyright 2020, Epic Systems Corporation.

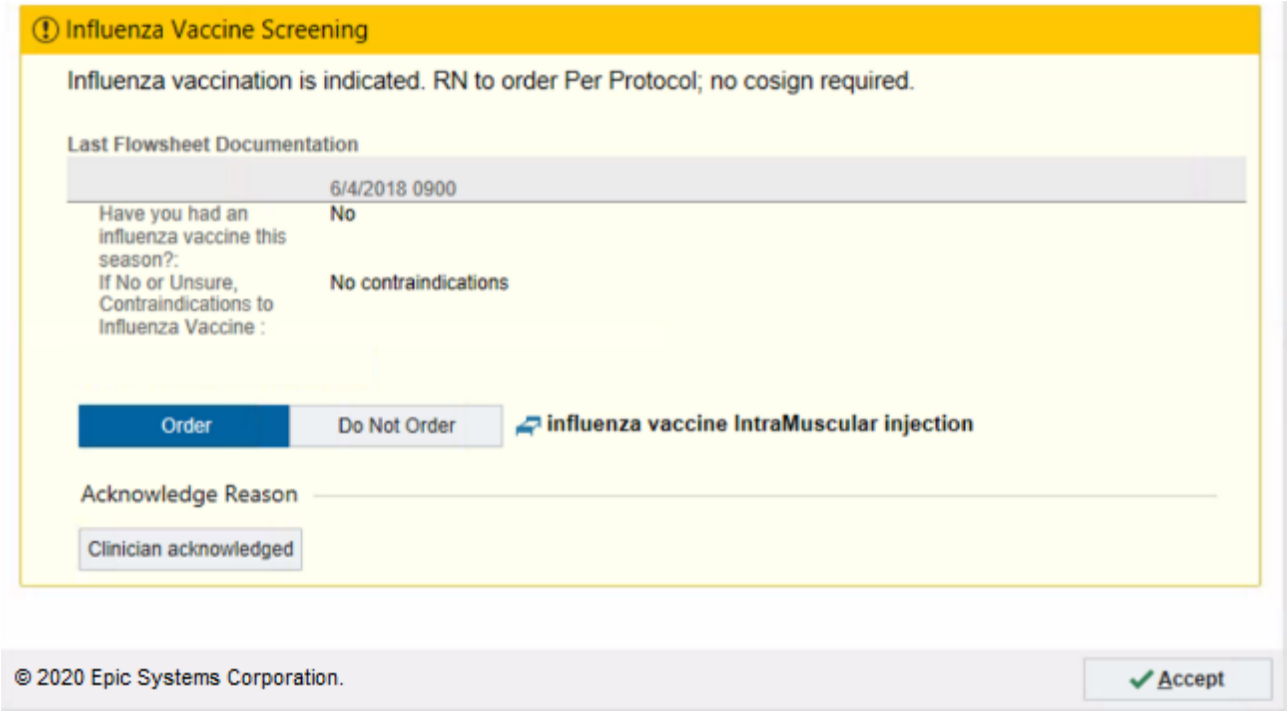


Table 1. Flu alert results.

Round	Alert version	Firings per patient per day	*P* value	Vaccination compliance rate at discharge, %
Baseline (Nov 2018 - Jan 2018)	1 (n=8296)	23.0	N/A[a]	N/A
1 (Feb 2018 - Apr 2018)	1 (n=2025)	23.6	.521[b]	90.8
1 (Feb 2018 - Apr 2018)	2 (n=2039)	23.1		
2 (Sep 2018 - Dec 2018)	3 (n=8777)	8.4	<.001[c]	87.7
3 (Jan 2019 - Feb 2019)	4 (n=1952)	7.3	<.001[c]	89.0

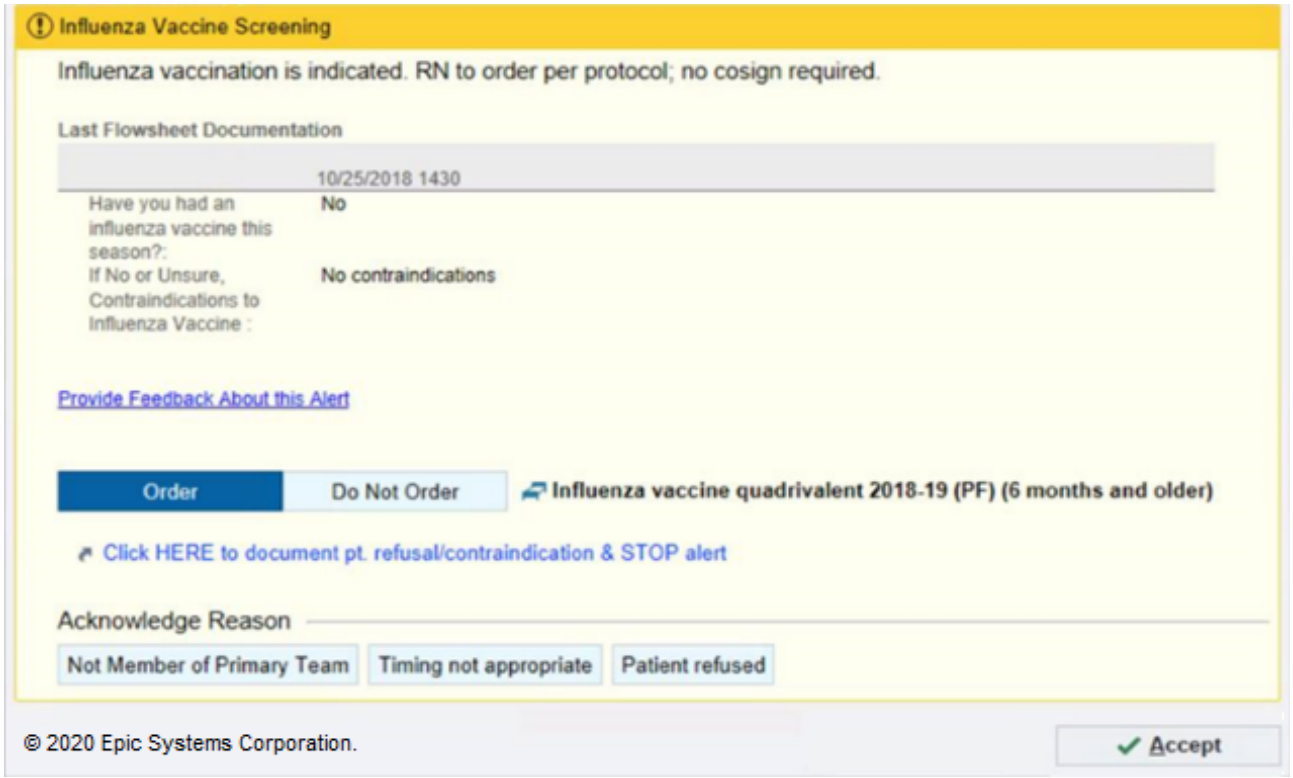
[a]N/A: not applicable.

[b]Comparing 1 vs 2 in the randomized controlled trial.

[c]Compared to baseline.

The CDS design team implemented the next prioritized enhancements in version 4. For the fourth cycle, the team identified that nurses disproportionally dismissed alerts in specific hospital units, including the Post-Anesthesia Care Unit. Upon learning about this finding, nursing leadership determined that these units were not appropriate locations to trigger the alert. Furthermore, our usability testing identified that the alert did not offer nurses sufficient locus of control. Specifically, nurses could not suppress the alert for a period of time when they felt the timing or recipient of the alert was inappropriate without placing the order. Consequently, we added targeted acknowledgement buttons to the alert based on these reasons that prevented the alert from firing for an acceptable period of time (Figure 5). This improvement resulted in a 13% reduction in alert firings per patient per day compared to version 3 (Table 1). These mixed methods served as a reminder that while A/B testing is a helpful tool, employing A/B testing in all situations is not the goal.

Figure 5. Version 4 of the flu alert, in which new acknowledgement reason buttons with lockout periods and a jumplink to update flowsheet documentation was added, and inappropriate units were excluded. Copyright 2020, Epic Systems Corporation.



Tobacco Alert Experiment

In the tobacco alert experiment, our week-long design thinking exercises produced 3 potential improvements with varying message framing (financial, evidence-based, regulatory) and complementary images (Figures 6-12). A financial framing indicated the additional revenue that a physician could generate by performing tobacco cessation counseling and gave them tools to document appropriately and create the relevant billing charge. The evidence framing highlighted that tobacco cessation was a part of providing high-quality care, and the regulatory framing indicated that tobacco cessation counseling was integral to the institution's expectations and policies. Additionally, in Round 2, images were added to reinforce the message framing.

Figure 6. Baseline tobacco cessation alert. Copyright 2020, Epic Systems Corporation.

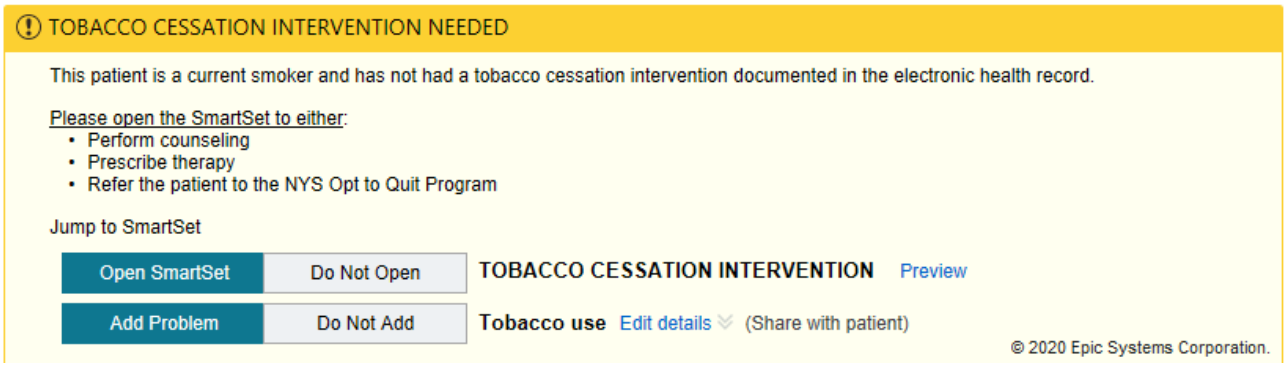


Figure 7. Tobacco financial messaging alert with images tested in Round 2 of the randomized controlled trial. Copyright 2020, Epic Systems Corporation.

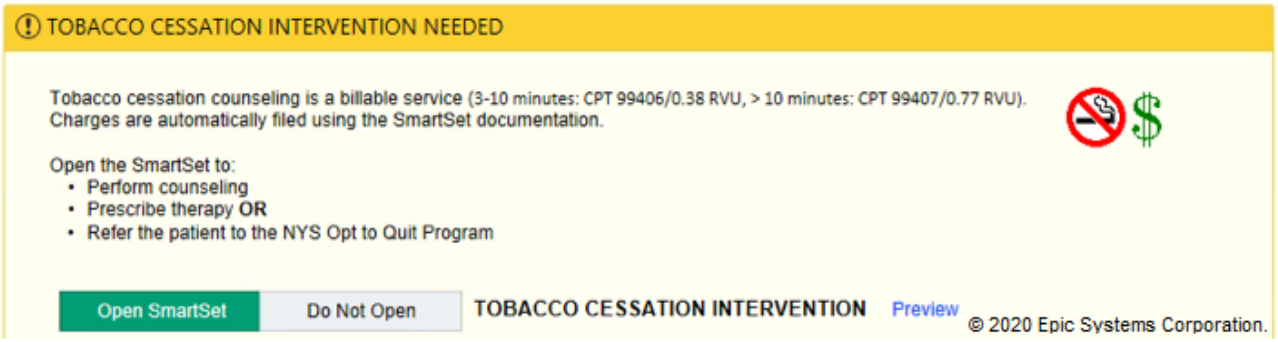


Figure 8. Tobacco evidence-based messaging alert with images tested in Round 2 of the randomized controlled trial. Copyright 2020, Epic Systems Corporation.

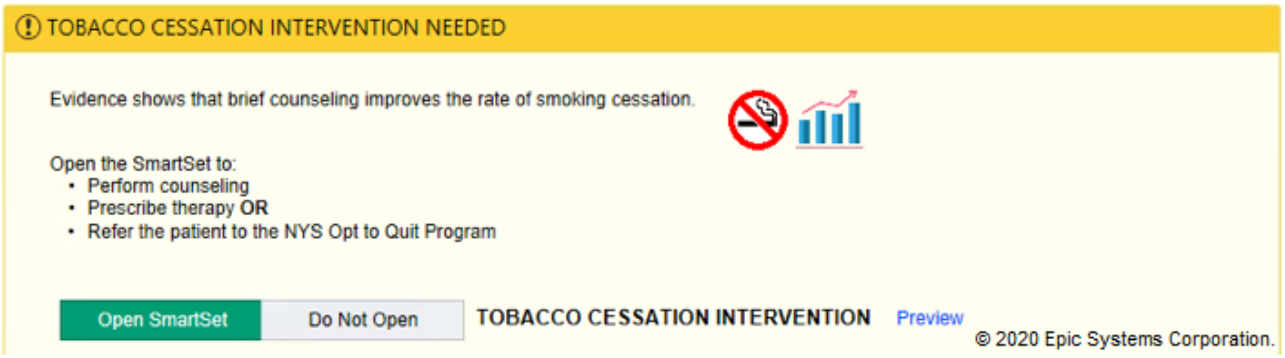


Figure 9. Tobacco regulatory messaging alert with images tested in Round 2 of the randomized controlled trial. Copyright 2020, Epic Systems Corporation.

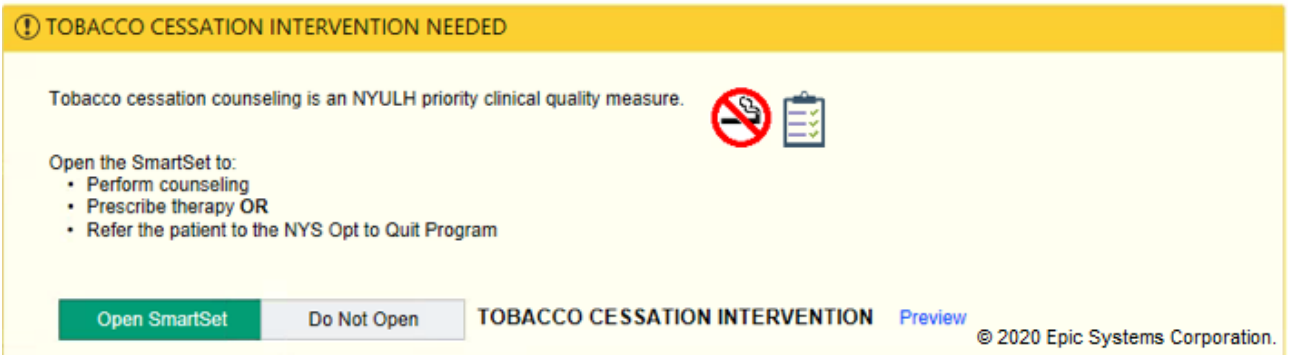


Figure 10. Tobacco financial messaging alert with no images tested in Round 3 of the randomized controlled trial. Copyright 2020, Epic Systems Corporation.

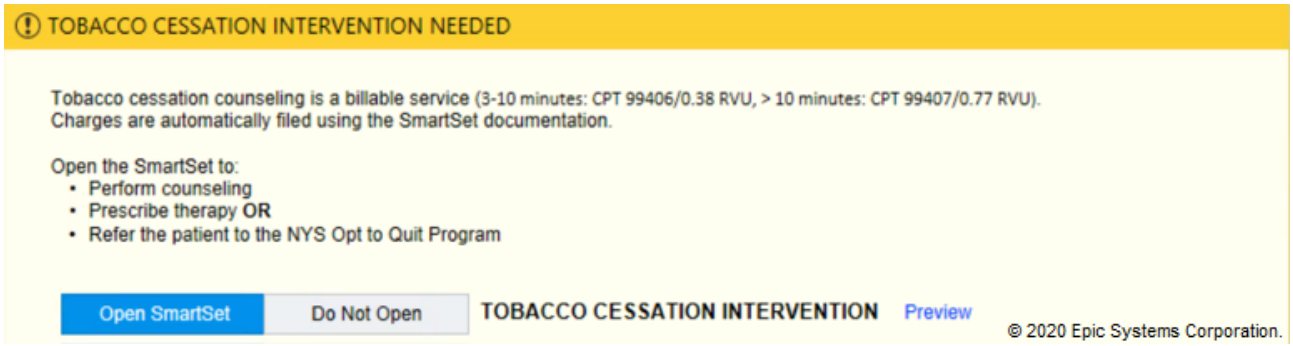


Figure 11. Tobacco financial messaging alert with both images (no smoking sign and dollar sign) tested in Round 3 of the randomized controlled trial. Copyright 2020, Epic Systems Corporation.

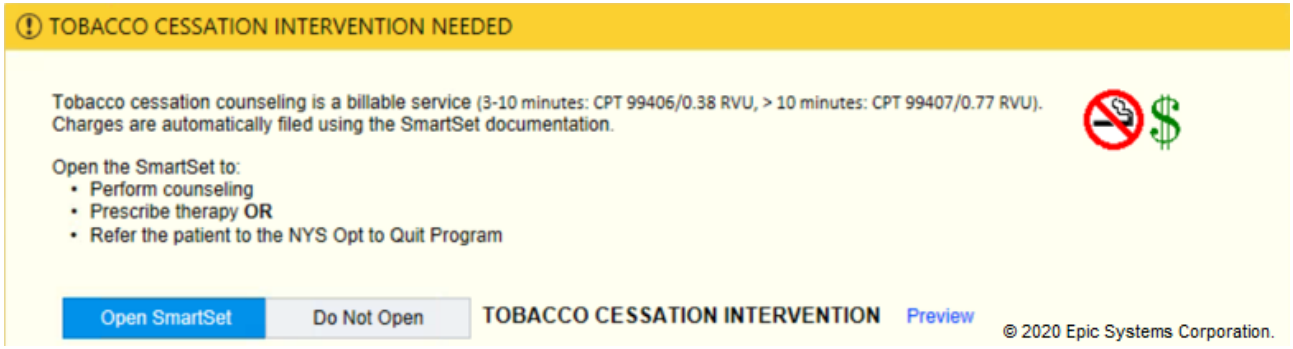
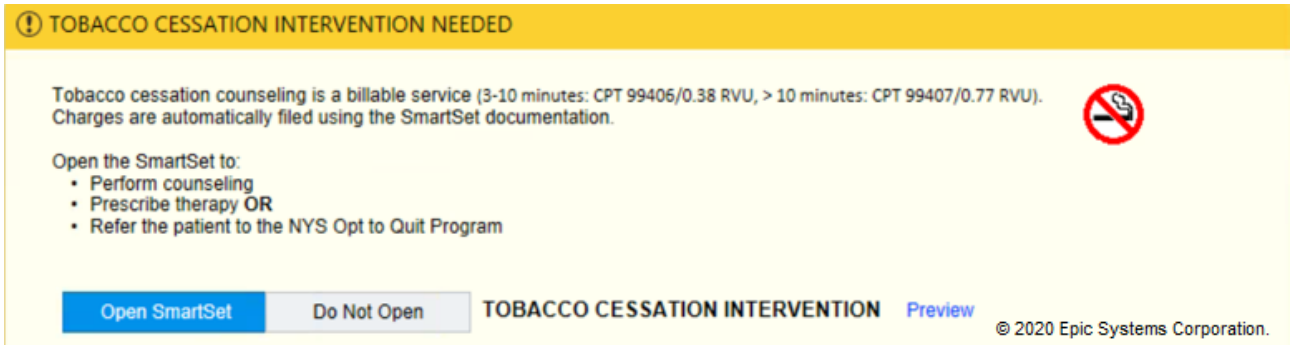


Figure 12. Tobacco financial messaging alert with image of no smoking sign tested in Round 3 of the randomized controlled trial. Copyright 2020, Epic Systems Corporation.



Each version was randomized at the ambulatory practice level for 4-5 weeks. In a series of 3 A/B testing experiments over a period of 8 months, our team observed that neither the framing method nor the addition of multiple or single images resulted in significant differences in acceptance rates for the tobacco alert (Table 2).

Table 2. Acceptance rates by alert type (encounter level acceptance rates; acceptance includes any positive action — perform counseling, prescribe therapy, or refer to state tobacco quit line).

Randomization group	Baseline acceptance rate (N=31650), number accepted/number displayed (%)	Round 1, Oct 2018 - Nov 2018 (N=26,975), number accepted/number displayed (%)	OR[a] (95% CI)	Round 2, Jan 2019 - Feb 2019 (N=11,631), number accepted/number displayed (%)	OR (95% CI)	Round 3, Apr 2019 - May 2019 (N=15,811), number accepted/number displayed (%)	OR (95% CI)
A	2327/8782 (26)	2045/7621 (27), financial incentive message	0.89 (0.48-1.67)	817/3335 (24), financial message with images	0.87 (0.47-1.59)	1036/4395 (24), financial message with no images	0.90 (0.49-1.65)
B	2171/10,585 (21)	1832/8821 (21), evidence-based message	0.90 (0.50-1.63)	705/3702 (19), evidence-based messages with images	0.88 (0.49-1.58)	903/4991 (18), financial message with images	0.90 (0.50-1.64)
C	2619/12,283 (21)	2682/10,533 (25), institutional priority message	1.00	1122/4594 (24), institutional priority message with images	1.00	1295/6425 (20), financial message with tobacco sign only	1.00

[a]OR: odds ratio.

Discussion

Our experience and results provide significant insights into the opportunities and challenges in transitioning to a strategy of optimization that has worked so successfully in other areas of information technology.

By creating an infrastructure to support this approach, we successfully tested multiple versions of alerts in a short period of time with outcomes that are supported by rigorous methods. This capability stands in contrast to the common approach of releasing single versions of CDS alerts with no empiric data supporting the relative efficacy of their design and then relying on weaker pre-post statistical methods for assessment.

This success is predicated on having assembled a multidisciplinary team, institutional support, and a repeatable process that can be applied to the rapid improvement of any CDS tool. It draws heavily on both the user-centered, agile approach from software and other quality improvement philosophies and traditional RCT methodologies. In our experience, standing up this approach is the most challenging phase. It requires effort to foster buy-in from the IRB, operational leadership, enhanced EHR randomization capabilities, and reporting to rapidly test CDS variations. The creation of the process drew heavily on our prior experience in usability and user-centered design [35] and a flexible randomization schema that could be tailored to the dynamic research methodology.

While we were successful at developing this infrastructure, the outcomes from our initial experiments offer guidance to other institutions embarking on a similar approach. For our first A/B testing experiment in influenza, we intentionally chose a less invasive design change (verbiage and mild display modifications) that would not be controversial to operational leaders and minimized risk of unintended consequences in our deployment. We recommend this approach as it allowed the information technology team members and clinical members to focus their efforts on creating the robust testing infrastructure without the distraction of complex intervention changes. The first experiment confirmed that a simple solution would not result in a dramatic improvement in alert acceptance. While strategizing the next experiment, the team uncovered more fundamental design flaws with clear remedies. Rather than subjecting these remedies to a round of RCTs, we opted to follow a traditional pre-post evaluation given the extremely high likelihood of success. This approach proved successful as we had dramatic increases in alert acceptance rates. All nurses benefited from this enhancement rather than having to wait for the result of a second RCT. Consequently, CDS teams have to consider the appropriate use cases in deploying randomized A/B testing vs more traditional heuristics-based approaches to improving CDS [21,36]. In the future, A/B testing could be more beneficial after these basic heuristic constructs have been satisfied.

In addition to the right use case, there are outstanding questions related to the right volume of usability testing vs rapid A/B testing to deploy when refining the CDS. Usability testing is significantly more resource and time intensive as compared to rapid A/B testing. Consequently, we will continue to explore how to balance a priori usability testing versus empiric A/B tests.

Similarly, despite repeated cycles, no intervention improved the acceptance rate of the tobacco alert. This result is likely because clinical practice behaviors are challenging to modify and changes to CDS presentation displays might not be sufficiently impactful. While small tests of change could be helpful to large internet sites with millions of users per day, these changes might not translate as well to CDS with a smaller, highly trained population. With our growing confidence in our underlying framework and approach, we will begin to experiment with larger structural changes to our CDS that are more likely to have impact.

Despite the successes in our implementation strategy, there are several constraints to this approach. Randomizing at the clinician level continues to be a challenge as our EHR does not easily support it and manual randomization of thousands of providers is suboptimal. Moreover, due to the pragmatic nature of the research, contamination is always a risk since clinicians and clinical workflows are not static. For instance, the flu alert was randomized at the patient level, which meant that nurses likely

interacted with both versions of the alert even on the same day, possibly reducing the potential effect size.

Our cluster randomization, while minimizing contamination bias, was challenging given underlying unanticipated irregularities in the practice sites. Specifically, operational reorganization of the ambulatory network during the trials distorted the original randomization. Furthermore, we assumed that randomization had successfully distributed key characteristics across arms and initially did not check baseline rates in randomized groups. We corrected for this using statistical procedures and now examine baseline rates of all key variables prior to randomization. We also did not include a baseline control arm, which made interpretation of our results more challenging. In the future, to mitigate inevitable changes in organization structures, cluster randomization should be stratified by practice groups in addition to by individual practices. Nonetheless, organizational changes may disrupt pragmatic experiments reinforcing the importance of a rapid, iterative evaluation framework.

Finally, post-hoc analysis revealed that our statistical team had misinterpreted data retrieved from the EHR. Data from each round were captured at the EHR department level. However, our analysis was conducted at the clinical practice level. This disconnect required our team to manually map EHR departments to clinical practices, which were not always in a 1:1 relationship. This allocation was further complicated by operational changes where practices were removed and combined. For the future, we would recommend randomizing at the level data will be reported (ie, at the EHR department level for cluster randomization).

These experiences highlight the importance of having the capability to make quick, data-driven decisions and a process to rapidly remediate mistakes and apply these learnings to future A/B cycles.

Fundamental differences between health care organizations and software companies provide additional challenges to prioritizing A/B experiments. Unlike large software companies whose primary key performance indicators are dependent on user acceptance of decision support (eg, clicking on an advertisement), alert burden is not yet the major priority of clinical institutions, though it is gaining importance given the rising appreciation of physician burnout [37].

Rapid A/B testing of CDS alerts in combination with RCT methods is a promising approach to efficiently, rapidly, and rigorously evaluating the impact of the tools and the clinicians' experience using them. Our experience also highlights the evaluative challenges associated with cluster randomization and the outstanding need for collaboration with EHR vendors to design scalable randomization approaches at various levels. Applied broadly, this approach could also help reduce the amount of CDS "noise" in the system, by both reducing the number of alerts and making each more impactful. If this proves true, the application of rapid A/B testing and RCT methods to CDS alerts could be a potential intervention for alert fatigue and improve the EHR experience.

Acknowledgments

Funding for the Rapid RCT Lab program was provided by Allen Thorpe. The authors thank Allen Thorpe for funding the Rapid RCT Lab program and Mark Samia RN, Christopher Mahoney, Lani Albania RN, Dawn Feldthouse RN, Rose Ann Alfaro RN, and Catherine Snapp at NYU Langone Medical Center Information Technology who worked on project implementation.

Authors' Contributions

All authors made substantial contributions to (1) conception and design, acquisition of data, or analysis or interpretation of data and (2) drafting or critical revisions of the manuscript. All listed authors approved the final version of the manuscript to be published.

Conflicts of Interest

None declared.

Multimedia Appendix 1
A tip sheet on how to randomize within Epic created by our institution's Epic Research team.
[PDF File (Adobe PDF File), 236 KB - jmir_v23i4e16651_app1.pdf]

Multimedia Appendix 2
Consort-EHealth V1.6 publication submission form.
[PDF File (Adobe PDF File), 3719 KB - jmir_v23i4e16651_app2.pdf]

References

1. Hunt DL, Haynes RB, Hanna SE, Smith K. Effects of computer-based clinical decision support systems on physician performance and patient outcomes: a systematic review. JAMA 1998 Oct 21;280(15):1339-1346. [doi: 10.1001/jama.280.15.1339] [Medline: 9794315]

2. Randolph AG, Haynes RB, Wyatt JC, Cook DJ, Guyatt GH. Users' Guides to the Medical Literature: XVIII. How to use an article evaluating the clinical impact of a computer-based clinical decision support system. JAMA 1999 Jul 07;282(1):67-74. [doi: 10.1001/jama.282.1.67] [Medline: 10404914]
3. Overhage JM, Tierney WM, Zhou XH, McDonald CJ. A randomized trial of "corollary orders" to prevent errors of omission. J Am Med Inform Assoc 1997;4(5):364-375 [FREE Full text] [doi: 10.1136/jamia.1997.0040364] [Medline: 9292842]
4. Bernstein SL, Whitaker D, Winograd J, Brennan JA. An electronic chart prompt to decrease proprietary antibiotic prescription to self-pay patients. Acad Emerg Med 2005 Mar;12(3):225-231 [FREE Full text] [doi: 10.1197/j.aem.2004.09.021] [Medline: 15741585]
5. Garthwaite EA, Will EJ, Bartlett C, Richardson D, Newstead CG. Patient-specific prompts in the cholesterol management of renal transplant outpatients: results and analysis of underperformance. Transplantation 2004 Oct 15;78(7):1042-1047. [doi: 10.1097/01.tp.0000137340.22880.c8] [Medline: 15480172]
6. Safran C, Rind DM, Davis RB, Ives D, Sands DZ, Currier J, et al. Guidelines for management of HIV infection with computer-based patient's record. The Lancet 1995 Aug 05;346(8971):341-346. [doi: 10.1016/s0140-6736(95)92226-1] [Medline: 7623532]
7. Safran C, Rind DM, Davis RM, Currier J, Ives D, Sands DZ, et al. An electronic medical record that helps care for patients with HIV infection. Proc Annu Symp Comput Appl Med Care 1993:224-228 [FREE Full text] [Medline: 8130466]
8. Safran C, Rind DM, Sands DZ, Davis RB, Wald J, Slack WV. Development of a knowledge-based electronic patient record. MD Comput 1996;13(1):46-54, 63. [Medline: 8569464]
9. Tierney WM, Miller ME, McDonald CJ. The effect on test ordering of informing physicians of the charges for outpatient diagnostic tests. N Engl J Med 1990 May 24;322(21):1499-1504. [doi: 10.1056/NEJM199005243222105] [Medline: 2186274]
10. Simon SR, Smith DH, Feldstein AC, Perrin N, Yang X, Zhou Y, et al. Computerized prescribing alerts and group academic detailing to reduce the use of potentially inappropriate medications in older people. J Am Geriatr Soc 2006 Jun;54(6):963-968. [doi: 10.1111/j.1532-5415.2006.00734.x] [Medline: 16776793]
11. Shah NR, Seger AC, Seger DL, Fiskio JM, Kuperman GJ, Blumenfeld B, et al. Improving acceptance of computerized prescribing alerts in ambulatory care. J Am Med Inform Assoc 2006;13(1):5-11 [FREE Full text] [doi: 10.1197/jamia.M1868] [Medline: 16221941]
12. Tamblyn R, Huang A, Perreault R, Jacques A, Roy D, Hanley J, et al. The medical office of the 21st century (MOXXI): effectiveness of computerized decision-making support in reducing inappropriate prescribing in primary care. CMAJ 2003 Sep 16;169(6):549-556 [FREE Full text] [Medline: 12975221]
13. Gaikwad R, Sketris I, Shepherd M, Duffy J. Evaluation of accuracy of drug interaction alerts triggered by two electronic medical record systems in primary healthcare. Health Informatics J 2007 Sep;13(3):163-177 [FREE Full text] [doi: 10.1177/1460458207079836] [Medline: 17711879]
14. Smith DH, Perrin N, Feldstein A, Yang X, Kuang D, Simon SR, et al. The impact of prescribing safety alerts for elderly persons in an electronic medical record: an interrupted time series evaluation. Arch Intern Med 2006 May 22;166(10):1098-1104 [FREE Full text] [doi: 10.1001/archinte.166.10.1098] [Medline: 16717172]
15. Seidling HM, Schmitt SPW, Bruckner T, Kaltschmidt J, Pruszydlo MG, Senger C, et al. Patient-specific electronic decision support reduces prescription of excessive doses. Qual Saf Health Care 2010 Oct;19(5):e15. [doi: 10.1136/qshc.2009.033175] [Medline: 20427312]
16. Austrian JS, Adelman JS, Reissman SH, Cohen HW, Billett HH. The impact of the heparin-induced thrombocytopenia (HIT) computerized alert on provider behaviors and patient outcomes. J Am Med Inform Assoc 2011;18(6):783-788 [FREE Full text] [doi: 10.1136/amiajnl-2011-000138] [Medline: 21712374]
17. Austrian JS, Jamin CT, Doty GR, Blecker S. Impact of an emergency department electronic sepsis surveillance system on patient mortality and length of stay. J Am Med Inform Assoc 2018 May 01;25(5):523-529 [FREE Full text] [doi: 10.1093/jamia/ocx072] [Medline: 29025165]
18. Garg AX, Adhikari NKJ, McDonald H, Rosas-Arellano MP, Devereaux PJ, Beyene J, et al. Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: a systematic review. JAMA 2005 Mar 09;293(10):1223-1238. [doi: 10.1001/jama.293.10.1223] [Medline: 15755945]
19. Krall MA, Sittig DF. Clinician's assessments of outpatient electronic medical record alert and reminder usability and usefulness requirements. Proc AMIA Symp 2002:400-404 [FREE Full text] [Medline: 12463855]
20. Footracer KG. Alert fatigue in electronic health records. JAAPA 2015 Jul;28(7):41-42. [doi: 10.1097/01.JAA.0000465221.04234.ca] [Medline: 26107794]
21. Bates DW, Kuperman GJ, Wang S, Gandhi T, Kittler A, Volk L, et al. Ten commandments for effective clinical decision support: making the practice of evidence-based medicine a reality. J Am Med Inform Assoc 2003;10(6):523-530 [FREE Full text] [doi: 10.1197/jamia.M1370] [Medline: 12925543]
22. Khalifa M, Zabani I. Improving Utilization of Clinical Decision Support Systems by Reducing Alert Fatigue: Strategies and Recommendations. Stud Health Technol Inform 2016;226:51-54. [Medline: 27350464]

23. McCullagh LJ, Sofianou A, Kannry J, Mann DM, McGinn TG. User centered clinical decision support tools: adoption across clinician training level. Appl Clin Inform 2014;5(4):1015-1025 [FREE Full text] [doi: 10.4338/ACI-2014-05-RA-0048] [Medline: 25589914]
24. Press A, Khan S, McCullagh L, Schachter A, Pardo S, Kohn N, et al. Avoiding alert fatigue in pulmonary embolism decision support: a new method to examine 'trigger rates'. Evid Based Med 2016 Dec;21(6):203-207. [doi: 10.1136/ebmed-2016-110440] [Medline: 27664174]
25. Kohavi R, Thomke S. The Surprising Power of Online Experiments. Harvard Business Review. 2017 Sep. URL: https://hbr.org/2017/09/the-surprising-power-of-online-experiments [accessed 2021-03-27]
26. Novak A. Going to Market Faster: Most Companies Are Deploying Code Weekly, Daily, or Hourly. New Relic. 2016 Feb 04. URL: https://blog.newrelic.com/technology/data-culture-survey-results-faster-deployment/ [accessed 2021-03-27]
27. Horwitz LI, Kuznetsova M, Jones SA. Creating a Learning Health System through Rapid-Cycle, Randomized Testing. N Engl J Med 2019 Sep 19;381(12):1175-1179. [doi: 10.1056/NEJMsb1900856] [Medline: 31532967]
28. Mann DM, Chokshi SK, Kushniruk A. Bridging the Gap Between Academic Research and Pragmatic Needs in Usability: A Hybrid Approach to Usability Evaluation of Health Care Information Systems. JMIR Hum Factors 2018 Nov 28;5(4):e10721 [FREE Full text] [doi: 10.2196/10721] [Medline: 30487119]
29. Payne TH, Hines LE, Chan RC, Hartman S, Kapusnik-Uner J, Russ AL, et al. Recommendations to improve the usability of drug-drug interaction clinical decision support alerts. J Am Med Inform Assoc 2015 Nov;22(6):1243-1250. [doi: 10.1093/jamia/ocv011] [Medline: 25829460]
30. Russ AL, Zillich AJ, Melton BL, Russell SA, Chen S, Spina JR, et al. Applying human factors principles to alert design increases efficiency and reduces prescribing errors in a scenario-based simulation. J Am Med Inform Assoc 2014 Oct;21(e2):e287-e296 [FREE Full text] [doi: 10.1136/amiajnl-2013-002045] [Medline: 24668841]
31. Horsky J, Phansalkar S, Desai A, Bell D, Middleton B. Design of decision support interventions for medication prescribing. Int J Med Inform 2013 Jun;82(6):492-503. [doi: 10.1016/j.ijmedinf.2013.02.003] [Medline: 23490305]
32. Chokshi SK, Belli HM, Troxel AB, Blecker S, Blaum C, Testa P, et al. Designing for implementation: user-centered development and pilot testing of a behavioral economic-inspired electronic health record clinical decision support module. Pilot Feasibility Stud 2019;5:28 [FREE Full text] [doi: 10.1186/s40814-019-0403-z] [Medline: 30820339]
33. Bates D, Mächler M, Bolker B, Walker S. Fitting Linear Mixed-Effects Models Using lme4. J. Stat. Soft 2015;67(1):1. [doi: 10.18637/jss.v067.i01]
34. Venables WN, Ripley BD. Modern Applied Statistics with S, 4th edition. New York, NY: Springer Publishing Company; 2002.
35. Chokshi SK, Mann DM. Innovating From Within: A Process Model for User-Centered Digital Development in Academic Medical Centers. JMIR Hum Factors 2018 Dec 19;5(4):e11048 [FREE Full text] [doi: 10.2196/11048] [Medline: 30567688]
36. Osheroff J, editor. Improving Medication Use and Outcomes with Clinical Decision Support: A Step by Step Guide. Chicago, IL: HIMSS Publishing; 2009.
37. Downing NL, Bates DW, Longhurst CA. Physician Burnout in the Electronic Health Record Era: Are We Ignoring the Real Cause? Ann Intern Med 2018 Jul 03;169(1):50-51. [doi: 10.7326/M18-0139] [Medline: 29801050]

Abbreviations

CDS: clinical decision support
EHR: electronic health record
IRB: institutional review board
RCT: randomized controlled trial

Edited by G Eysenbach; submitted 14.10.19; peer-reviewed by A Al, D Newman, Y Leykin; comments to author 24.04.20; revised version received 14.08.20; accepted 11.03.21; published 09.04.21.

Please cite as:
Austrian J, Mendoza F, Szerencsy A, Fenelon L, Horwitz LI, Jones S, Kuznetsova M, Mann DM
Applying A/B Testing to Clinical Decision Support: Rapid Randomized Controlled Trials
J Med Internet Res 2021;23(4):e16651
URL: https://www.jmir.org/2021/4/e16651
doi:10.2196/16651
PMID:33835035

Original Paper

Generalizability of an Automatic Explanation Method for Machine Learning Prediction Results on Asthma-Related Hospital Visits in Patients With Asthma: Quantitative Analysis

Gang Luo[1], DPhil; Claudia L Nau[2], DPhil; William W Crawford[3], MD; Michael Schatz[2,4], MD, MS; Robert S Zeiger[2,4], MD, DPhil; Corinna Koebnick[2], DPhil

[1]Department of Biomedical Informatics and Medical Education, University of Washington, Seattle, WA, United States
[2]Department of Research & Evaluation, Kaiser Permanente Southern California, Pasadena, CA, United States
[3]Department of Allergy and Immunology, Kaiser Permanente South Bay Medical Center, Harbor City, CA, United States
[4]Department of Allergy, Kaiser Permanente Southern California, San Diego, CA, United States

Corresponding Author:
Gang Luo, DPhil
Department of Biomedical Informatics and Medical Education
University of Washington
UW Medicine South Lake Union, 850 Republican Street,
Box 358047, Building C
Seattle, WA, 98195
United States
Phone: 1 2062214596
Fax: 1 206 221 2671
Email: gangluo@cs.wisc.edu

Abstract

Background: Asthma exerts a substantial burden on patients and health care systems. To facilitate preventive care for asthma management and improve patient outcomes, we recently developed two machine learning models, one on Intermountain Healthcare data and the other on Kaiser Permanente Southern California (KPSC) data, to forecast asthma-related hospital visits, including emergency department visits and hospitalizations, in the succeeding 12 months among patients with asthma. As is typical for machine learning approaches, these two models do not explain their forecasting results. To address the interpretability issue of black-box models, we designed an automatic method to offer rule format explanations for the forecasting results of any machine learning model on imbalanced tabular data and to suggest customized interventions with no accuracy loss. Our method worked well for explaining the forecasting results of our Intermountain Healthcare model, but its generalizability to other health care systems remains unknown.

Objective: The objective of this study is to evaluate the generalizability of our automatic explanation method to KPSC for forecasting asthma-related hospital visits.

Methods: Through a secondary analysis of 987,506 data instances from 2012 to 2017 at KPSC, we used our method to explain the forecasting results of our KPSC model and to suggest customized interventions. The patient cohort covered a random sample of 70% of patients with asthma who had a KPSC health plan for any period between 2015 and 2018.

Results: Our method explained the forecasting results for 97.57% (2204/2259) of the patients with asthma who were correctly forecasted to undergo asthma-related hospital visits in the succeeding 12 months.

Conclusions: For forecasting asthma-related hospital visits, our automatic explanation method exhibited an acceptable generalizability to KPSC.

International Registered Report Identifier (IRRID): RR2-10.2196/resprot.5039

(J Med Internet Res 2021;23(4):e24153) doi:10.2196/24153

KEYWORDS

asthma; forecasting; patient care management; machine learning

Introduction

Background

Asthma affects 8.4% of the US population [1], resulting in over 3000 deaths, almost 500,000 hospitalizations, and over 2,000,000 emergency department (ED) visits every year [1,2]. The state-of-the-art method for reducing asthma-related hospital visits, including ED visits and hospitalizations, is to use a model to forecast which of the patients with asthma are prone to future poor outcomes. We then enroll these patients in care management, ensuring care managers call them periodically to help them schedule health and related services. As used by health care systems like University of Washington Medicine, Intermountain Healthcare, and Kaiser Permanente Northern California [3] and by health plans in 9 of 12 metropolitan communities [4], this method, if implemented properly, can cut up to 40% of patients' future hospital visits [5-8].

Having a limited capacity, a care management program can enroll only a small portion of patients [9], and its effectiveness is upper bounded by the accuracy of the predictive model. Because they are missing certain important features, the existing models for forecasting asthma-related hospital visits among patients with asthma [3,10-22] are inaccurate, with each model missing over half of patients who will undergo future asthma-related hospital visits and mislabeling many others as making such visits. As a result, care management programs continue to be used inefficiently because they are unable to focus on the highest-risk patients. In addition, patient outcomes deteriorate, whereas health care costs increase. To address this problem, we recently considered many candidate features and developed two extreme gradient boosting (XGBoost) [23] machine learning models, one on Intermountain Healthcare data [24] and the other on Kaiser Permanente Southern California (KPSC) data [25], to forecast asthma-related hospital visits in the succeeding 12 months among patients with asthma with a higher accuracy. As is typical for machine learning approaches, these two models do not explain their forecasting results. Clinicians would know that a patient is considered to be at a high risk by the model, but the model offers no reason why this is the case. This makes it difficult for clinicians to understand and trust the model's prediction result, determine whether the patient should be put into care management, and pinpoint interventions suitable for the patient. To address the interpretability issue of black-box models, we designed an automatic method to offer rule format explanations for any machine learning model's forecasting results on imbalanced tabular data and to suggest customized interventions with no accuracy loss [26]. Our method worked well for explaining our Intermountain Healthcare model's forecasting results [26], but its generalizability to other health care systems remains unknown.

Objectives

The objective of this study is to evaluate the generalizability of our automatic explanation method to KPSC in forecasting asthma-related hospital visits. In the following sections, we describe our evaluation approach and results.

Methods

Ethics Approval and Study Design

After receiving approval from the institutional review boards of KPSC and University of Washington Medicine, in this study, we conducted a secondary analysis of retrospective data.

Patient Population

We adopted the same patient cohort from our previous KPSC predictive model paper [25]: a random sample of 70% of patients with asthma who had a KPSC health plan for any period between 2015 and 2018. This sample size is the largest one permitted by KPSC for sharing its data with another non-Kaiser Permanente institution for research. KPSC has 227 clinics and 15 hospitals. It is the largest integrated health care system in Southern California, offering care to approximately 19% of residents there [27]. A patient was deemed asthmatic in a specific year if during that year, at least one asthma diagnosis code (*International Classification of Diseases, Tenth Revision* [*ICD-10*]: J45.x; *International Classification of Diseases, Ninth Revision* [*ICD-9*]: 493.0x, 493.1x, 493.8x, and 493.9x) was recorded on the patient in the encounter billing database [11,28,29]. Patient death during that year served as the exclusion criterion.

Prediction Target (Dependent Variable)

We adopted the same prediction target as our prior KPSC predictive model paper [25]. For every patient deemed to have asthma in a specific year, the indicator of any asthma-related hospital visit in the succeeding year is the outcome. An asthma-related hospital visit is a hospitalization or ED visit with asthma as its principal diagnosis (*ICD-10*: J45.x; *ICD-9*: 493.0x, 493.1x, 493.8x, and 493.9x). When training and testing our automatic explanation method and our KPSC XGBoost model, for each patient who had a KPSC health plan on a year's last day and was also deemed asthmatic in the year, we used the patient's data up to the year's last day to forecast the patient's outcome in the succeeding year.

Data Set

We adopted the same administrative and clinical data set from our prior KPSC predictive model paper [25]. Obtained from KPSC's research data warehouse, this structured data set covered our patient cohort's visits at KPSC between 2010 and 2018.

Features (Independent Variables), Predictive Models, and Data Preprocessing

Our KPSC model [25] uses the XGBoost classification algorithm [23] and 221 features to forecast asthma-related hospital visits in the succeeding year in patients with asthma. These features are listed in our previous KPSC predictive model paper [25], were computed on the structured attributes in our data set, and cover various characteristics such as patient demographics, medications, visits, diagnoses, vital signs, procedures, and laboratory tests. An example feature is the total number of asthma relievers that the patient filled in the previous 12 months. Every input data instance to our KPSC model aims at a (patient, index year) pair, includes these 221 features, and is used to forecast the succeeding year's outcome of the patient. As in our

prior KPSC predictive model paper [25], the top 10% of patients with asthma projected at the highest risk were used as the cutoff point for binary classification. We used the same data preprocessing approach adopted in our prior KPSC predictive model paper [25] to clean, normalize, and prepare the data.

Review of Our Automatic Explanation Method

Previously, we designed an automatic method to offer rule format explanations for the forecasting results of any machine learning model on tabular data and to suggest customized interventions with no accuracy loss. The original method [30] was designed for relatively balanced data. Recently, we extended the method to handle imbalanced data [26], where one value of the outcome variable has a much lower prevalence rate than another. This fits the case of forecasting asthma-related hospital visits in patients with asthma. At KPSC, the prevalence rate of having asthma-related hospital visits in the succeeding year was approximately 2%. In the remainder of this paper, we focus on the extended automatic explanation method.

Main Idea

The central idea of our automatic explanation method is to use two models side by side to separate the forecasting and offering explanations. Each model serves a different purpose. The first model was used for the forecasting. Typically chosen to be the most accurate one, this model can be any model built on continuous and categorical features. The second model contains class-based association rules [31,32] mined from past data. It is used not to forecast but to explain the forecasting results of the first model. After using an automatic discretization method [31,33] to convert continuous features to categorical features, we use a standard approach such as Apriori to mine the association rules [32]. Each rule presents a feature pattern linking to a value u of the outcome variable and has the form:

r_1 AND r_2 AND ... AND $r_s \rightarrow u$.

The values of s and u can differ across the rules. For the binary classification of poor versus good outcomes, u is typically a poor outcome value. Each item r_i ($1 \leq i \leq s$) is a feature-value pair (g, v). When v is a value, r_i shows that feature g has a value v. When v is a range, r_i indicates that the value of g is within v. The rule signifies that a patient's outcome is apt to be u if r_1, r_2, ..., and r_s are all satisfied by the patient. An exemplar rule is:

> The patient had 8 or 9 primary or principal asthma diagnoses in the previous 12 months
>
> AND the patient had ≥6 no shows in the prior 12 months
>
> → The patient will undergo ≥1 asthma-related hospital visit in the subsequent 12 months.

The Rule Mining and Pruning Process

Our automatic explanation method uses 5 parameters: the minimum commonality threshold, the minimum confidence threshold, the largest number of items permitted on an association rule's left-hand side, the confidence difference threshold, and the number of top features used to construct rules. For a given rule

r_1 AND r_2 AND ... AND $r_s \rightarrow u$,

its commonality reflects its coverage in the context of u and refers to the fraction of data instances fulfilling r_1, r_2, ..., and r_s among all the data instances connected to u. Its confidence reflects its precision and refers to the fraction of data instances connecting to u among all the data instances fulfilling r_1, r_2, ..., and r_s. Our method uses those rules whose commonality is no less than the minimum commonality threshold, whose confidence is no less than the minimum confidence threshold, and each containing no more than the maximum permitted number of items on its left-hand side.

We use 3 techniques to reduce the number of association rules and prevent it from being excessively large. First, we remove every more specific rule q_1 in the presence of a more general rule q_2 satisfying q_2's confidence ≥ q_1's confidence−the confidence difference threshold. Second, certain machine learning algorithms, such as XGBoost [23], can automatically compute every feature's importance value. When handling a large data set with many features, only the top few features having the largest importance values and used in the first model are adopted to construct rules. Third, a clinician in the design team of the automatic explanation function examines all possible values and value ranges of the features adopted to construct rules and labels those values and value ranges that could have a positive correlation with the poor outcome value. Only the labeled feature values and value ranges are adopted to form rules.

For each feature-value pair item that is adopted to construct association rules, a clinician in the design team of the automatic explanation function compiles zero or more interventions. We tag an item actionable if it links to at least one intervention. Each rule passing the rule pruning process is automatically linked to the interventions related to the actionable items on the left-hand side of the rule. We tag a rule actionable if it contains at least one actionable item on its left-hand side; that is, it links to at least one intervention.

The Explanation Approach

For every patient whom the first model forecasts to take a poor outcome value, we explain the forecasting result by showing the association rules in the second model having this value on their right-hand sides and whose left-hand sides are satisfied by the patient. Each rule provides a reason why the patient is forecasted to take this value. For each actionable rule that is shown, the interventions connected to it are listed next to it. The automatic explanation function's user can find customized interventions that fit the patient from the listed interventions. Usually, the rules in the second model present common reasons for having poor outcomes. Some patients will experience poor outcomes for other reasons. Thus, the second model can explain most, but not all, of the poor outcomes correctly forecasted by the first model.

Parameter Setting

In our experiments, we used the same parameter setting approach used in our previous automatic explanation paper [26]. Each association rule had no more than 5 items on its left-hand

side. Our KPSC XGBoost model [25] used 221 features to forecast asthma-related hospital visits. We used the top 50 features that our KPSC model ranked with the largest importance values to construct association rules. Our KPSC model gained an area under the receiver operating characteristic curve (AUC) of 0.820 using all 221 features and an AUC of 0.815 using the top 50 features.

For forecasting asthma-related hospital visits, our KPSC model [25] obtained a lower AUC for KPSC data than our Intermountain Healthcare model on Intermountain Healthcare data [24]. As mentioned in our previous automatic explanation paper [26], the harder it is to forecast the outcome, the smaller the minimum commonality and confidence thresholds need to be to ensure that our automatic explanation method can provide explanations for a large percentage of the patients whom the first model correctly forecasts to take a poor outcome value. Following this guideline on KPSC data, we set the minimum commonality threshold to 0.08%, which is lower than the corresponding value of 0.2% we used on Intermountain Healthcare data [26]. We set the minimum confidence threshold to 25%, which is lower than the corresponding value of 50% used for the Intermountain Healthcare data [26]. Despite not looking large, 25% is much greater than 2%, which is the percentage of KPSC data instances associated with asthma-related hospital visits in the succeeding year, as well as our KPSC model's positive predictive value of 11.03% [25].

To set the value of the confidence difference threshold τ, we calculated the number of association rules passing the rule pruning process versus τ. Our previous paper [26] shows that this number of rules first drops quickly as τ rises and then drops slowly when τ becomes sufficiently large. The value of τ was set at the transition point.

Data Analysis

Partitioning of the Training and Test Sets

We used the same method adopted in our prior KPSC predictive model paper [25] to divide the entire data set into training and test sets. As several features were computed on the data from up to 2 years before the index year and the outcomes came from the succeeding year, our data set included 6 years of effective data (2012-2017) over the 9-year period of 2010-2018. To match the use of our KPSC model and our automatic explanation method in clinical practice, we used the 2012-2016 data as the training set to train our KPSC model and mine the association rules adopted by our automatic explanation method. We used the 2017 data as the test set to gauge the performance of our KPSC model and the automatic explanation method.

Performance Metrics

We used the same performance metrics from our previous automatic explanation paper [26] to assess the performance of our automatic explanation method. A performance metric on our method's explanation power is the fraction of patients with asthma whom our method could offer explanations for among the patients whom our KPSC model correctly forecasted to undergo asthma-related hospital visits in the succeeding year. We computed the average number of rules and the average number of actionable rules that suit such a patient. A rule suits a patient if all items on its left-hand side are satisfied for the patient.

As our previous automatic explanation paper [26] showed, several rules suiting a patient often differ by a single item on their left-hand sides. When multiple rules suit a patient, the amount of nonredundant information included in them is usually much less than the number of rules in them. To plot a full picture of the amount of information included in the automatic explanations given to the patients, we computed 3 distributions of the patients with asthma whom our KPSC model correctly forecasted to undergo asthma-related hospital visits in the succeeding year: (1) by the number of actionable rules suiting a patient, (2) by the number of different actionable items included in all the rules suiting a patient, and (3) by the number of rules suiting a patient.

Results

Demographic and Clinical Characteristics of Our Patient Cohort

Remember that each data instance aims at a different (patient, index year) pair. Tables 1 and 2 present the demographic and clinical characteristics of our KPSC patient cohort during 2012-2016 and 2017, respectively. The two sets of characteristics are sufficiently similar to each other. During 2012-2016, 2.42% (18,925/782,762) of data instances were linked to asthma-related hospital visits in the succeeding year. During 2017, this fraction was 2.13% (4353/204,744). Our previous KPSC predictive model paper [25] provides a detailed comparison of the two sets of characteristics.

Table 1. Demographic and clinical characteristics of our Kaiser Permanente Southern California patient cohort during 2012-2016.

Characteristics	Data instances associated with no asthma-related hospital visit in the succeeding year (n=763,837), n (%)	Data instances associated with asthma-related hospital visits in the succeeding year (n=18,925), n (%)	Data instances (n=782,762), n (%)
Age (years)			
≥65	108,662 (14.23)	2288 (12.09)	110,950 (14.17)
18-65	415,889 (54.45)	8557 (45.22)	424,446 (54.22)
6 to <18	188,583 (24.69)	5039 (26.63)	193,622 (24.74)
<6	50,703 (6.64)	3041 (16.07)	53,744 (6.87)
Gender			
Female	443,410 (58.05)	10,590 (55.96)	454,000 (58.00)
Male	320,427 (41.95)	8335 (44.04)	328,762 (42.00)
Race			
White	477,542 (62.52)	10,040 (53.05)	487,582 (62.29)
Native Hawaiian or other Pacific Islander	7692 (1.01)	230 (1.22)	7922 (1.01)
Black or African American	110,869 (14.51)	4982 (26.33)	115,851 (14.80)
Asian	68,781 (9.00)	1282 (6.77)	70,063 (8.95)
American Indian or Alaska native	3745 (0.49)	86 (0.45)	3831 (0.49)
Unknown or unreported	95,208 (12.46)	2305 (12.18)	97,513 (12.46)
Ethnicity			
Non-Hispanic	449,795 (58.89)	10,577 (55.89)	460,372 (58.81)
Hispanic	299,240 (39.18)	8131 (42.96)	307,371 (39.27)
Unknown or unreported	14,802 (1.94)	217 (1.15)	15,019 (1.92)
Insurance			
Self-paid plan	104,479 (13.68)	2224 (11.75)	106,703 (13.63)
Public	216,320 (28.32)	7469 (39.47)	223,789 (28.59)
High deductible plan	80,393 (10.52)	1426 (7.54)	81,819 (10.45)
Exchange (also known as market-place)	39,050 (5.11)	735 (3.88)	39,785 (5.08)
Commercial (employer-paid)	521,101 (68.22)	11,311 (59.77)	532,412 (68.02)
Other	265,264 (34.73)	6064 (32.04)	271,328 (34.66)
Number of years from the first visit related to asthma in the data set			
>3	439,930 (57.59)	10,919 (57.70)	450,849 (57.60)
≤3	323,907 (42.41)	8006 (42.30)	331,913 (42.40)
Asthma medication fill			
Systemic corticosteroid	236,246 (30.93)	10,837 (57.26)	247,083 (31.57)
Short-acting, inhaled beta-2 agonist	537,442 (70.36)	16,242 (85.82)	553,684 (70.73)
Mast cell stabilizer	20 (0.00)	0 (0.00)	20 (0.00)
Long-acting beta-2 agonist	33,576 (4.40)	1694 (8.95)	35,270 (4.51)
Leukotriene modifier	85,299 (11.17)	4125 (21.80)	89,424 (11.42)
Combination of long-acting beta-2 agonist and inhaled corticosteroid	88,847 (11.63)	3975 (21.00)	92,822 (11.86)
Inhaled corticosteroid	325,156 (42.57)	11,841 (62.57)	336,997 (43.05)
Comorbidity			
Sleep apnea	20,465 (2.68)	575 (3.04)	21,040 (2.69)

Characteristics	Data instances associated with no asthma-related hospital visit in the succeeding year (n=763,837), n (%)	Data instances associated with asthma-related hospital visits in the succeeding year (n=18,925), n (%)	Data instances (n=782,762), n (%)
Sinusitis	112,341 (14.71)	2832 (14.96)	115,173 (14.71)
Premature birth	16,607 (2.17)	690 (3.65)	17,297 (2.21)
Obesity	171,666 (22.47)	4776 (25.24)	176,442 (22.54)
Gastroesophageal reflux	101,180 (13.25)	2778 (14.68)	103,958 (13.28)
Eczema	82,425 (10.79)	2944 (15.56)	85,369 (10.91)
Cystic fibrosis	135 (0.02)	3 (0.02)	138 (0.02)
Chronic obstructive pulmonary disease	27,388 (3.59)	999 (5.28)	28,387 (3.63)
Bronchopulmonary dysplasia	241 (0.03)	22 (0.12)	263 (0.03)
Anxiety or depression	160,719 (21.04)	4231 (22.36)	164,950 (21.07)
Allergic rhinitis	164,036 (21.48)	4673 (24.69)	168,709 (21.55)
Smoking status			
Never smoker or unknown	477,263 (62.48)	11,885 (62.80)	489,148 (62.49)
Former smoker	133,456 (17.47)	2870 (15.17)	136,326 (17.42)
Current smoker	153,118 (20.05)	4170 (22.03)	157,288 (20.09)

Table 2. Demographic and clinical characteristics of our Kaiser Permanente Southern California patient cohort in 2017.

Characteristics	Data instances associated with no asthma-related hospital visit in the succeeding year (n=200,391), n (%)	Data instances associated with asthma-related hospital visits in the succeeding year (n=4353), n (%)	Data instances (n=204,744), n (%)
Age (years)			
≥65	35,342 (17.64)	679 (15.60)	36,021 (17.59)
18-65	109,969 (54.88)	2052 (47.14)	112,021 (54.71)
6 to <18	43,856 (21.89)	1012 (23.25)	44,868 (21.91)
<6	11,224 (5.60)	610 (14.01)	11,834 (5.78)
Gender			
Female	118,013 (58.89)	2482 (57.02)	120,495 (58.85)
Male	82,378 (41.11)	1871 (42.98)	84,249 (41.15)
Race			
White	124,514 (62.14)	2302 (52.88)	126,816 (61.94)
Native Hawaiian or other Pacific Islander	1910 (0.95)	42 (0.96)	1952 (0.95)
Black or African American	26,864 (13.41)	1075 (24.70)	27,939 (13.65)
Asian	18,555 (9.26)	319 (7.33)	18,874 (9.22)
American Indian or Alaska native	987 (0.49)	31 (0.71)	1018 (0.50)
Unknown or unreported	27,561 (13.75)	584 (13.42)	28,145 (13.75)
Ethnicity			
Non-Hispanic	116,801 (58.29)	2410 (55.36)	119,211 (58.22)
Hispanic	78,153 (39.00)	1868 (42.91)	80,021 (39.08)
Unknown or unreported	5437 (2.71)	75 (1.72)	5512 (2.69)
Insurance			
Self-paid plan	33,758 (16.85)	647 (14.86)	34,405 (16.80)
Public	64,727 (32.30)	1904 (43.74)	66,631 (32.54)
High deductible plan	24,647 (12.30)	356 (8.18)	25,003 (12.21)
Exchange (also known as market-place)	17,677 (8.82)	269 (6.18)	17,946 (8.77)
Commercial (employer-paid)	127,724 (63.74)	2420 (55.59)	130,144 (63.56)
Other	83,108 (41.47)	1675 (38.48)	84,783 (41.41)
Number of years from the first visit related to asthma in the data set			
>3	116,285 (58.03)	2616 (60.10)	118,901 (58.07)
≤3	84,106 (41.97)	1737 (39.90)	85,843 (41.93)
Asthma medication fill			
Systemic corticosteroid	64,878 (32.38)	2597 (59.66)	67,475 (32.96)
Short-acting, inhaled beta-2 agonist	137,077 (68.40)	3742 (85.96)	140,819 (68.78)
Mast cell stabilizer	0 (0.00)	0 (0.00)	0 (0.00)
Long-acting beta-2 agonist	11,343 (5.66)	467 (10.73)	11,810 (5.77)
Leukotriene modifier	26,996 (13.47)	1099 (25.25)	28,095 (13.72)
Combination of long-acting beta-2 agonist and inhaled corticosteroid	28,580 (14.26)	1151 (26.44)	29,731 (14.52)
Inhaled corticosteroid	78,220 (39.03)	2586 (59.41)	80,806 (39.47)
Comorbidity			
Sleep apnea	12,811 (6.39)	333 (7.65)	13,144 (6.42)

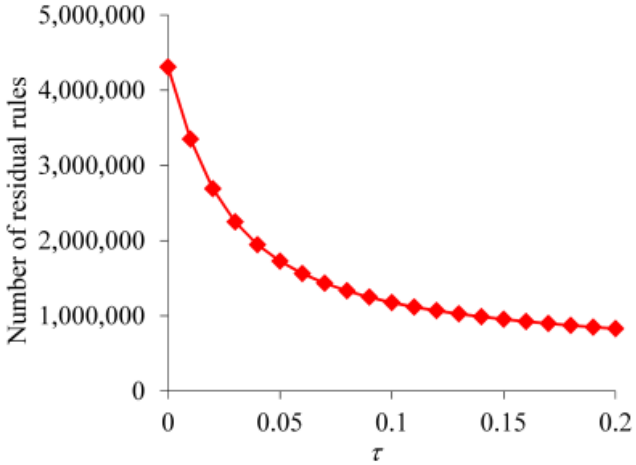
Characteristics	Data instances associated with no asthma-related hospital visit in the succeeding year (n=200,391), n (%)	Data instances associated with asthma-related hospital visits in the succeeding year (n=4353), n (%)	Data instances (n=204,744), n (%)
Sinusitis	29,202 (14.57)	680 (15.62)	29,882 (14.59)
Premature birth	4381 (2.19)	132 (3.03)	4513 (2.20)
Obesity	48,548 (24.23)	1190 (27.34)	49,738 (24.29)
Gastroesophageal reflux	32,462 (16.20)	797 (18.31)	33,259 (16.24)
Eczema	20,521 (10.24)	638 (14.66)	21,159 (10.33)
Cystic fibrosis	40 (0.02)	2 (0.05)	42 (0.02)
Chronic obstructive pulmonary disease	7306 (3.65)	285 (6.55)	7591 (3.71)
Bronchopulmonary dysplasia	29 (0.01)	1 (0.02)	30 (0.01)
Anxiety or depression	46,176 (23.04)	1124 (25.82)	47,300 (23.10)
Allergic rhinitis	39,849 (19.89)	1084 (24.90)	40,933 (19.99)
Smoking status			
Never smoker or unknown	125,245 (62.50)	2663 (61.18)	127,908 (62.47)
Former smoker	36,026 (17.98)	717 (16.47)	36,743 (17.95)
Current smoker	39,120 (19.52)	973 (22.35)	40,093 (19.58)

The Number of Residual Association Rules

Taking the top 50 features that our KPSC model ranked with the largest importance values, we mined 11,628,850 association rules from the training set. Figure 1 displays the number of residual rules versus the confidence difference threshold τ. This number first drops quickly as τ rises and then drops slowly when τ becomes ≥0.15. Accordingly, the value of τ was set to 0.15, resulting in 954,493 residual rules.

An asthma clinical expert in our team labeled the values and value ranges of the top 50 features that could have a positive correlation with asthma-related hospital visits in the succeeding year. After we removed those rules involving any other value or value range, 725,632 association rules remained. Each rule provides a reason why a patient is forecasted to undergo future asthma-related hospital visits. Almost all (725,623) of these rules were actionable. Thus, our automatic explanation method's performance numbers are almost the same regardless of whether all these rules or only the actionable rules were used. In the remainder of this section, we present only the performance numbers when only the actionable rules were used.

Figure 1. The number of residual rules versus the confidence difference threshold τ.



Example Association Rules Adopted by the Second Model

To allow the reader to gain a sense of the association rules the second model adopted, we present 5 example rules:

- Rule 1: The patient filled ≥89 asthma relievers in total in the previous 12 months

  → The patient will undergo ≥1 asthma-related hospital visits in the subsequent 12 months.

The use of many asthma relievers indicates poor asthma control. An intervention tied to the item *the patient filled ≥89 asthma*

relievers in total in the prior 12 months is to tailor prescribed medications and to suggest the patient to maximize adherence to asthma control medications or to improve avoidance of asthma triggers.

- Rule 2: The patient had ≥25 nebulizer medication orders in the previous 12 months

  AND the patient incurred ≥16 major visits for asthma in the previous 12 months

  → The patient will undergo ≥1 asthma-related hospital visits in the subsequent 12 months.

The use of many nebulizer medications indicates a poor asthma control. An intervention tied to the item *the patient had ≥25 nebulizer medication orders in the prior 12 months* is to tailor prescribed medications and to suggest the patient to maximize adherence to asthma control medications or to improve the avoidance of asthma triggers.

As defined in our previous paper [24], major visits for asthma cover outpatient visits linked to a primary diagnosis of asthma, and ED visits and hospitalizations linked to an asthma diagnosis code. Outpatient visits linked to a secondary, but not a primary, diagnosis of asthma are deemed minor visits for asthma. Having many major visits for asthma indicates a poor asthma control. An intervention tied to the item *the patient incurred ≥16 major visits for asthma in the prior 12 months* is to adopt control strategies for the patient to avoid needing emergency care.

- Rule 3: The patient had 8 or 9 primary or principal asthma diagnoses in the previous 12 months

  AND the patient had ≥6 no shows in the previous 12 months

  → The patient will undergo ≥1 asthma-related hospital visits in the subsequent 12 months.

Having many primary or principal asthma diagnoses indicates a poor asthma control. An intervention tied to the item *the patient had 8 or 9 primary or principal asthma diagnoses in the prior 12 months* is to offer the patient suggestions on how to improve asthma control.

Having many no shows correlates with poor outcomes. An intervention tied to the item *the patient had ≥6 no shows in the prior 12 months* is to give the patient social resources to handle socioeconomic challenges to keep appointments.

- Rule 4: The patient incurred ≥8 ED visits in the previous 12 months

  AND the patient was prescribed ≥28 short-acting beta-2 agonist medications in total in the previous 12 months

  AND the patient is Black or African American

  → The patient will undergo ≥1 asthma-related hospital visits in the subsequent 12 months.

In the United States, Black and African American people tend to have poorer asthma outcomes than others. Frequent ED visits indicated a poor asthma control. An intervention tied to the item *the patient incurred ≥8 ED visits in the prior 12 months* is to adopt control strategies for the patient to avoid needing emergency care.

Short-acting beta-2 agonists are rescue medications for the quick relief of asthma symptoms. The use of many short-acting beta-2 agonists indicates a poor asthma control. An intervention tied to the item *the patient was ordered ≥28 short-acting beta-2 agonist medications in total in the prior 12 months* is to tailor prescribed medications and to suggest the patient to maximize adherence to asthma control medications or to improve the avoidance of asthma triggers.

- Rule 5: The highest exacerbation severity of all asthma diagnoses recorded on the patient in the previous 12 months is status asthmaticus

  AND the patient incurred ≥11 and ≤17 visits with same-day appointments in the previous 12 months

  AND the admission type of the patient's last visit in the previous 12 months is nonelective

  → The patient will undergo ≥1 asthma-related hospital visits in the subsequent 12 months.

Status asthmaticus is the most severe form of asthma exacerbation. An intervention tied to the item *the highest exacerbation severity of all of the asthma diagnoses recorded on the patient in the prior 12 months is status asthmaticus* is to offer the patient suggestions on how to improve asthma control.

Having many visits with same-day appointments indicates a poor asthma control. An intervention tied to the item *the patient incurred ≥11 and ≤17 visits with same day appointments in the prior 12 months* is to improve support offered to the patient between visits to enhance medication adherence, address asthma triggers, and maximize the value of each visit.

A patient incurs a nonelective visit when the patient's condition requires an immediate medical attention, for example, when the patient experiences severe asthma exacerbation. An intervention tied to the item *the admission type of the patient's last visit in the prior 12 months is nonelective* is to adopt control strategies for the patient to avoid needing emergency care.

The Performance of Our Automatic Explanation Method

We evaluated our automatic explanation method on the test set. Our method explained the forecasting results for 97.9% (599/612) of the children (age <18 years) with asthma and 97.45% (1605/1647) of the adults (age ≥18 years) with asthma our KPSC model correctly forecasted to undergo asthma-related hospital visits in the succeeding year. Put together, our method explained the forecasting results for 97.57% (2204/2259) of the patients with asthma who were correctly forecasted to undergo asthma-related hospital visits in the succeeding year. For every such patient, on average, our method provided 1516.25 (SD 2161.30) explanations, each from one rule, and found 24.04 (SD 8.68) actionable items.

For the patients with asthma whom our KPSC model correctly forecasted to undergo asthma-related hospital visits in the succeeding year, Figures 2 and 3 display the patient distribution according to the number of actionable rules suiting a patient.

Having a long tail, this distribution is significantly skewed toward the left. As the number of rules suiting a patient increases, the number of patients each covered by this number of rules tends to decline nonmonotonically. The biggest number of rules suiting a patient is fairly large (15,252). However, only one patient matched this number of rules.

For the patients with asthma whom our KPSC model correctly forecasted to undergo asthma-related hospital visits in the succeeding year, Figure 4 displays the patient distribution according to the number of different actionable items included in all the rules suiting a patient. The largest number of different actionable items included in all the rules suiting a patient is 42, much less than the largest number of actionable rules suiting a patient. As noted in our previous automatic explanation paper [26], 2 or more actionable items included in the rules that suit a patient often connect to the same intervention.

Our automatic explanation method provided explanations for 67.61% (2943/4353) of patients with asthma who would undergo asthma-related hospital visits in the succeeding year.

Figure 2. The patient distribution by the number of actionable rules suiting a patient for the patients with asthma whom our Kaiser Permanente Southern California model correctly forecasted to undergo asthma-related hospital visits in the succeeding year.

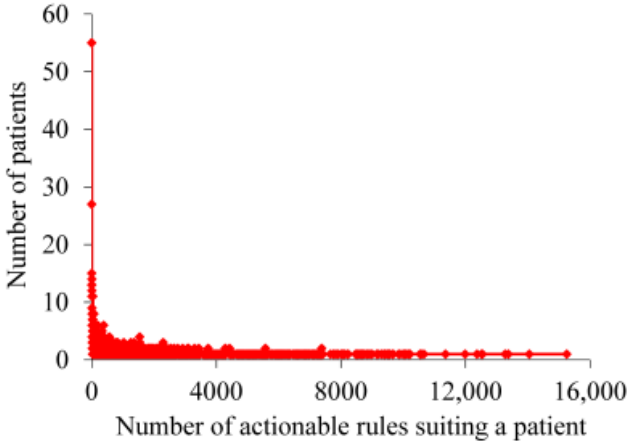


Figure 3. The patient distribution by the number of actionable rules suiting a patient when this number is ≤250 for the patients with asthma whom our Kaiser Permanente Southern California model correctly forecasted to undergo asthma-related hospital visits in the succeeding year.

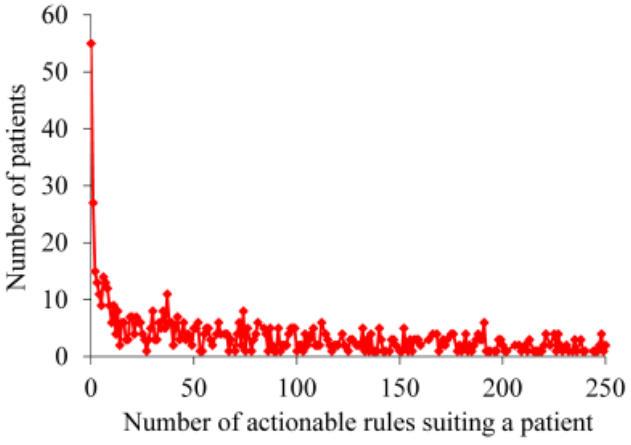
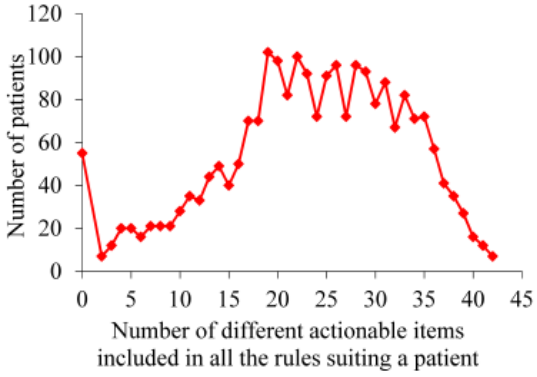


Figure 4. The patient distribution by the number of different actionable items included in all the rules suiting a patient for the patients with asthma whom our Kaiser Permanente Southern California model correctly forecasted to undergo asthma-related hospital visits in the succeeding year.



Discussion

Principal Findings

The results presented in this paper are similar to those presented in our previous automatic explanation paper [26]. For forecasting asthma-related hospital visits, our automatic explanation method exhibited an acceptable generalizability to KPSC. In particular, our method explained the forecasting results for 97.57% (2204/2259) of the patients with asthma who were correctly forecasted to undergo asthma-related hospital visits in the succeeding year. This fraction is comparable with that (89.68%) on Intermountain Healthcare data in our previous automatic explanation paper [26] and is large enough to put our automatic explanation method into daily clinical use. After further development to boost its accuracy, our KPSC model combined with our automatic explanation method could be used to guide asthma care management's use to help enhance patient outcomes and reduce health care costs.

Our automatic explanation method provided explanations for 67.61% (2943/4353) of patients with asthma who would undergo asthma-related hospital visits in the succeeding year. This fraction is less than the 97.57% (2204/2259) success rate, at which our method explained the forecasting results for the patients with asthma our KPSC model correctly forecasted to undergo asthma-related hospital visits in the succeeding year. This is possibly due to the correlation between the association rules' and our KPSC model's forecasting results. Among the patients with asthma whom our KPSC model correctly forecasted to undergo asthma-related hospital visits in the succeeding year, many are easy cases for us to explain their outcomes using association rules. Among the patients with asthma who would undergo asthma-related hospital visits in the succeeding year and whose outcomes were incorrectly forecasted by our KPSC model, many are difficult cases for any model to correctly explain or forecast their outcomes.

Displaying the Automatic Explanations

Many rules could suit a patient. In this case, it is undesirable to list all of them simultaneously and overwhelm the user of the automatic explanation function. Instead, we should rank these rules and display the top few (eg, 3) of them by default. If desired, the user can ask an automatic explanation function to show more rules. In ranking the rules suiting a patient and the items on the left-hand side of a rule, we consider the following factors and strike a balance among them:

1. All else being equal, rules with fewer items on their left-hand sides are easier to understand and should be ranked higher.
2. All else being equal, rules with higher confidence are more precise and should be ranked higher.
3. All else being equal, rules with a higher commonality cover more patients with poor outcomes and should be ranked higher.
4. The automatic explanation function's user tends to read the rules one by one in the display order. All else being equal, the more items on the left-hand side of a rule appear in higher-ranked rules, the less new information that the user has not seen so far is contained in the rule and the lower the rule should be ranked.
5. Consider the items on the left-hand side of a rule. The automatic explanation function's user tends to read the items one by one in the display order. All else being equal, the items that have appeared in higher-ranked rules contain repeated information and should be put after the other items that have not appeared in any of the higher-ranked rules.
6. The automatic explanation function's user cares about finding suitable interventions for the patient. Consider the items on the left-hand side of a rule. All else being equal, the actionable items should be placed before the nonactionable items.
7. Actionable rules should be ranked higher than nonactionable rules.

We are in the process of preparing a paper describing our rule-ranking method in detail.

Related Work

As described in the book [34] and the survey paper [35], many other researchers have proposed miscellaneous methods for automatically offering explanations for the forecasting results of machine learning models. Such explanations are typically not in a rule format. Many such methods sacrifice a part of the forecasting accuracy and/or are designed for a particular machine learning algorithm. In addition, none of these methods can automatically suggest customized interventions. In comparison, our automatic explanation method supplies rule format explanations for any machine learning model's forecasting results on tabular data and suggests customized interventions with no accuracy loss. Rule format explanations are easier to comprehend and can suggest customized interventions more directly than other forms of explanations.

To the best of our knowledge, we were the first to use association rules to automatically offer rule format explanations for any machine learning model's forecasting results on tabular data and to suggest customized interventions with no accuracy loss [30]. Our original method [30] was designed for relatively balanced data and was initially tested in the case of forecasting type 2 diabetes diagnoses. Subsequently, Alaa et al [36,37] applied the original method to multiple medical prediction tasks. So far, no researcher outside of our group has applied our extended automatic explanation method [26], which can handle imbalanced data, to any prediction task. Rudin et al [38] and Ribeiro et al [39] used rules to automatically offer explanations for the forecasting results of any machine learning model. These rules are not association rules and are unknown before the prediction time; hence, they cannot be used to automatically suggest customized interventions at the prediction time. In comparison, the association rules used in our automatic explanation method are mined before the prediction time and used to automatically suggest customized interventions at the prediction time.

Limitations

This study has three limitations, all of which can be fine areas for future work:

1. For forecasting asthma-related hospital visits, our study evaluated the generalizability of our automatic explanation method to a single health care system. It would be nice to assess our automatic explanation method's generalizability to other health care systems, such as academic ones, which have different properties from Intermountain Healthcare and KPSC. In comparison with nonacademic systems, academic health care systems tend to handle more complex and sicker patients [40]. To prepare for such an evaluation, we are currently retrieving a data set of patients with asthma from the enterprise data warehouse of the University of Washington Medicine [41].
2. Our study evaluated the generalizability of our automatic explanation method only for forecasting asthma-related hospital visits. It would be nice to assess the generalizability of our automatic explanation method for other diseases and outcomes [41].
3. Our current automatic explanation method is designed for structured data and traditional machine learning algorithms that are not deep learning algorithms. It would be nice to extend our method so it can also handle deep learning models built directly on longitudinal data [41,42].

Conclusions

In its first generalizability assessment, our automatic explanation method for imbalanced tabular data exhibited a decent generalizability to KPSC for forecasting asthma-related hospital visits. After further development to boost its accuracy, our KPSC model combined with our automatic explanation method could be used to guide asthma care management's use to help enhance patient outcomes and reduce health care costs.

Acknowledgments

The authors thank Lee J Barton, Don McCarthy, Emily Rozema, Amanda I Messinger, and Michael D Johnson for useful discussions and helping retrieve the KPSC data set. GL, CLN, MS, RSZ, and CK were partially supported by the National Heart, Lung, and Blood Institute of the National Institutes of Health under award number R01HL142503. The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Authors' Contributions

GL was mainly responsible for this study. He conceptualized and designed the study, performed a literature review and data analysis, and wrote the paper. CK, CLN, WWC, MS, and RSZ provided feedback on various medical issues, contributed to conceptualizing the presentation, and revised the paper. CK and CLN took part in retrieving the KPSC data set and interpreting its detected peculiarities.

Conflicts of Interest

None declared.

References

1. Moorman JE, Akinbami LJ, Bailey CM, Zahran HS, King ME, Johnson CA, et al. National surveillance of asthma: United States, 2001-2010. Vital Health Stat 3 2012 Nov(35):1-58 [FREE Full text] [Medline: 24252609]
2. Nurmagambetov T, Kuwahara R, Garbe P. The economic burden of asthma in the United States, 2008-2013. Ann Am Thorac Soc 2018 Mar;15(3):348-356. [doi: 10.1513/AnnalsATS.201703-259OC] [Medline: 29323930]
3. Lieu TA, Quesenberry CP, Sorel ME, Mendoza GR, Leong AB. Computer-based models to identify high-risk children with asthma. Am J Respir Crit Care Med 1998 Apr;157(4 Pt 1):1173-1180. [doi: 10.1164/ajrccm.157.4.9708124] [Medline: 9563736]
4. Mays GP, Claxton G, White J. Managed care rebound? Recent changes in health plans' cost containment strategies. Health Aff (Millwood) 2004;Suppl Web Exclusives:W4-427-W4-436 [FREE Full text] [doi: 10.1377/hlthaff.w4.427] [Medline: 15451964]
5. Caloyeras JP, Liu H, Exum E, Broderick M, Mattke S. Managing manifest diseases, but not health risks, saved PepsiCo money over seven years. Health Aff (Millwood) 2014 Jan;33(1):124-131. [doi: 10.1377/hlthaff.2013.0625] [Medline: 24395944]
6. Greineder DK, Loane KC, Parks P. A randomized controlled trial of a pediatric asthma outreach program. J Allergy Clin Immunol 1999 Mar;103(3 Pt 1):436-440. [Medline: 10069877]
7. Kelly CS, Morrow AL, Shults J, Nakas N, Strope GL, Adelman RD. Outcomes evaluation of a comprehensive intervention program for asthmatic children enrolled in Medicaid. Pediatrics 2000 May;105(5):1029-1035. [Medline: 10790458]
8. Axelrod RC, Zimbro KS, Chetney RR, Sabol J, Ainsworth VJ. A disease management program utilizing life coaches for children with asthma. J Clin Outcomes Manag 2001;8(6):38-42.
9. Axelrod RC, Vogel D. Predictive modeling in health plans. Dis Manag Health Outcomes 2003;11(12):779-787. [doi: 10.2165/00115677-200311120-00003]
10. Loymans RJB, Honkoop PJ, Termeer EH, Snoeck-Stroband JB, Assendelft WJJ, Schermer TRJ, et al. Identifying patients at risk for severe exacerbations of asthma: development and external validation of a multivariable prediction model. Thorax 2016 Sep;71(9):838-846. [doi: 10.1136/thoraxjnl-2015-208138] [Medline: 27044486]

11. Schatz M, Cook EF, Joshua A, Petitti D. Risk factors for asthma hospitalizations in a managed care organization: development of a clinical prediction rule. Am J Manag Care 2003 Aug;9(8):538-547 [FREE Full text] [Medline: 12921231]
12. Eisner MD, Yegin A, Trzaskoma B. Severity of asthma score predicts clinical outcomes in patients with moderate to severe persistent asthma. Chest 2012 Jan;141(1):58-65. [doi: 10.1378/chest.11-0020] [Medline: 21885725]
13. Sato R, Tomita K, Sano H, Ichihashi H, Yamagata S, Sano A, et al. The strategy for predicting future exacerbation of asthma using a combination of the Asthma Control Test and lung function test. J Asthma 2009 Sep;46(7):677-682. [doi: 10.1080/02770900902972160] [Medline: 19728204]
14. Osborne ML, Pedula KL, O'Hollaren M, Ettinger KM, Stibolt T, Buist AS, et al. Assessing future need for acute care in adult asthmatics: the Profile of Asthma Risk Study: a prospective health maintenance organization-based study. Chest 2007 Oct;132(4):1151-1161. [doi: 10.1378/chest.05-3084] [Medline: 17573515]
15. Miller MK, Lee JH, Blanc PD, Pasta DJ, Gujrathi S, Barron H, TENOR Study Group. TENOR risk score predicts healthcare in adults with severe or difficult-to-treat asthma. Eur Respir J 2006 Dec;28(6):1145-1155 [FREE Full text] [doi: 10.1183/09031936.06.00145105] [Medline: 16870656]
16. Peters D, Chen C, Markson LE, Allen-Ramey FC, Vollmer WM. Using an asthma control questionnaire and administrative data to predict health-care utilization. Chest 2006 Apr;129(4):918-924. [doi: 10.1378/chest.129.4.918] [Medline: 16608939]
17. Yurk RA, Diette GB, Skinner EA, Dominici F, Clark RD, Steinwachs DM, et al. Predicting patient-reported asthma outcomes for adults in managed care. Am J Manag Care 2004 May;10(5):321-328 [FREE Full text] [Medline: 15152702]
18. Loymans RJ, Debray TP, Honkoop PJ, Termeer EH, Snoeck-Stroband JB, Schermer TR, et al. Exacerbations in adults with asthma: a systematic review and external validation of prediction models. J Allergy Clin Immunol Pract 2018;6(6):1942-1952.e15. [doi: 10.1016/j.jaip.2018.02.004] [Medline: 29454163]
19. Lieu TA, Capra AM, Quesenberry CP, Mendoza GR, Mazar M. Computer-based models to identify high-risk adults with asthma: is the glass half empty of half full? J Asthma 1999 Jun;36(4):359-370. [Medline: 10386500]
20. Schatz M, Nakahiro R, Jones CH, Roth RM, Joshua A, Petitti D. Asthma population management: development and validation of a practical 3-level risk stratification scheme. Am J Manag Care 2004 Jan;10(1):25-32 [FREE Full text] [Medline: 14738184]
21. Grana J, Preston S, McDermott PD, Hanchak NA. The use of administrative data to risk-stratify asthmatic patients. Am J Med Qual 1997;12(2):113-119. [doi: 10.1177/0885713X9701200205] [Medline: 9161058]
22. Forno E, Fuhlbrigge A, Soto-Quirós ME, Avila L, Raby BA, Brehm J, et al. Risk factors and predictive clinical scores for asthma exacerbations in childhood. Chest 2010 Nov;138(5):1156-1165 [FREE Full text] [doi: 10.1378/chest.09-2426] [Medline: 20472862]
23. Chen T, Guestrin C. XGBoost: A scalable tree boosting system. In: Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2016 Presented at: KDD'16; August 13-17, 2016; San Francisco, CA p. 785-794. [doi: 10.1145/2939672.2939785]
24. Luo G, He S, Stone BL, Nkoy FL, Johnson MD. Developing a model to predict hospital encounters for asthma in asthmatic patients: secondary analysis. JMIR Med Inform 2020 Jan 21;8(1):e16080 [FREE Full text] [doi: 10.2196/16080] [Medline: 31961332]
25. Luo G, Nau CL, Crawford WW, Schatz M, Zeiger RS, Rozema E, et al. Developing a predictive model for asthma-related hospital encounters in patients with asthma in a large integrated healthcare system: secondary analysis. JMIR Med Inform 2020 Nov 09;8(11):e22689 [FREE Full text] [doi: 10.2196/22689] [Medline: 33164906]
26. Luo G, Johnson MD, Nkoy FL, He S, Stone BL. Automatically explaining machine learning prediction results on asthma hospital visits in patients with asthma: secondary analysis. JMIR Med Inform 2020 Dec 31;8(12):e21965 [FREE Full text] [doi: 10.2196/21965] [Medline: 33382379]
27. Koebnick C, Langer-Gould AM, Gould MK, Chao CR, Iyer RL, Smith N, et al. Sociodemographic characteristics of members of a large, integrated health care system: comparison with US Census Bureau data. Perm J 2012;16(3):37-41 [FREE Full text] [Medline: 23012597]
28. Desai JR, Wu P, Nichols GA, Lieu TA, O'Connor PJ. Diabetes and asthma case identification, validation, and representativeness when using electronic health data to construct registries for comparative effectiveness and epidemiologic research. Med Care 2012 Jul;50 Suppl:S30-S35. [doi: 10.1097/MLR.0b013e318259c011] [Medline: 22692256]
29. Wakefield DB, Cloutier MM. Modifications to HEDIS and CSTE algorithms improve case recognition of pediatric asthma. Pediatr Pulmonol 2006 Oct;41(10):962-971. [doi: 10.1002/ppul.20476] [Medline: 16871628]
30. Luo G. Automatically explaining machine learning prediction results: a demonstration on type 2 diabetes risk prediction. Health Inf Sci Syst 2016;4:2 [FREE Full text] [doi: 10.1186/s13755-016-0015-4] [Medline: 26958341]
31. Liu B, Hsu W, Ma Y. Integrating classification and association rule mining. In: Proceedings of the 4th International Conference on Knowledge Discovery and Data Mining. 1998 Presented at: KDD'98; August 27-31, 1998; New York City, NY p. 80-86.
32. Thabtah FA. A review of associative classification mining. Knowledge Eng Review 2007 Mar 01;22(1):37-65. [doi: 10.1017/s0269888907001026]

33. Fayyad UM, Irani KB. Multi-interval discretization of continuous-valued attributes for classification learning. In: Proceedings of the 13th International Joint Conference on Artificial Intelligence. 1993 Presented at: IJCAI'93; August 28-September 3, 1993; Chambéry, France p. 1022-1029.
34. Molnar C. Interpretable Machine Learning. Morrisville, NC: lulu.com; 2020.
35. Guidotti R, Monreale A, Ruggieri S, Turini F, Giannotti F, Pedreschi D. A survey of methods for explaining black box models. ACM Comput Surv 2019 Jan 23;51(5):93. [doi: 10.1145/3236009]
36. Alaa AM, van der Schaar M. Prognostication and risk factors for cystic fibrosis via automated machine learning. Sci Rep 2018 Jul 26;8(1):11242 [FREE Full text] [doi: 10.1038/s41598-018-29523-2] [Medline: 30050169]
37. Alaa AM, van der Schaar M. AutoPrognosis: automated clinical prognostic modeling via Bayesian optimization with structured kernel learning. In: Proceedings of the 35th International Conference on Machine Learning. 2018 Presented at: ICML'18; July 10-15, 2018; Stockholm, Sweden p. 139-148.
38. Rudin C, Shaposhnik Y. Globally-consistent rule-based summary-explanations for machine learning models: application to credit-risk evaluation. In: Proceedings of INFORMS 11th Conference on Information Systems and Technology. 2019 Presented at: CIST'19; October 19-20, 2019; Seattle, WA p. 1-19. [doi: 10.2139/ssrn.3395422]
39. Ribeiro MT, Singh S, Guestrin C. Anchors: high-precision model-agnostic explanations. In: Proceedings of the 32nd AAAI Conference on Artificial Intelligence. 2018 Presented at: AAAI'18; February 2-7, 2018; New Orleans, LA p. 1527-1535.
40. Liu LL, Forgione DA, Younis MZ. A comparative analysis of the CVP structure of nonprofit teaching and for-profit non-teaching hospitals. J Health Care Finance 2012;39(1):12-38. [Medline: 23155742]
41. Luo G, Stone BL, Koebnick C, He S, Au DH, Sheng X, et al. Using temporal features to provide data-driven clinical early warnings for chronic obstructive pulmonary disease and asthma care management: protocol for a secondary analysis. JMIR Res Protoc 2019 Jun 06;8(6):e13783 [FREE Full text] [doi: 10.2196/13783] [Medline: 31199308]
42. Luo G. A roadmap for semi-automatically extracting predictive and clinically meaningful temporal features from medical data for predictive modeling. Glob Transit 2019;1:61-82 [FREE Full text] [doi: 10.1016/j.glt.2018.11.001] [Medline: 31032483]

Abbreviations

AUC: area under the receiver operating characteristic curve
ED: emergency department
ICD-9: International Classification of Diseases, Ninth Revision
ICD-10: International Classification of Diseases, Tenth Revision
KPSC: Kaiser Permanente Southern California
XGBoost: extreme gradient boosting

Edited by R Kukafka; submitted 17.10.20; peer-reviewed by A Ramjewan, Q Zhang; comments to author 06.12.20; revised version received 07.12.20; accepted 22.03.21; published 15.04.21.

Please cite as:
Luo G, Nau CL, Crawford WW, Schatz M, Zeiger RS, Koebnick C
Generalizability of an Automatic Explanation Method for Machine Learning Prediction Results on Asthma-Related Hospital Visits in Patients With Asthma: Quantitative Analysis
J Med Internet Res 2021;23(4):e24153
URL: https://www.jmir.org/2021/4/e24153
doi:10.2196/24153
PMID:33856359

Original Paper

Usability of Electronic Health Record–Generated Discharge Summaries: Heuristic Evaluation

Patrice D Tremoulet[1,2], PhD; Priyanka D Shah[2], BTech, MS; Alisha A Acosta[3]; Christian W Grant[1]; Jon T Kurtz[4]; Peter Mounas[5]; Michael Kirchhoff[6], MD; Elizabeth Wade[7], PharmD

[1]Department of Psychology, Rowan University, Glassboro, NJ, United States
[2]Device Evaluation, ECRI, Plymouth Meeting, PA, United States
[3]Department of Biochemistry, Rowan University, Glassboro, NJ, United States
[4]Department of Computer Science, Rowan University, Glassboro, NJ, United States
[5]Department of Biological Sciences, Rowan University, Glassboro, NJ, United States
[6]Department of Emergency Medicine, Cooper Medical School, Rowan University, Camden, NJ, United States
[7]Concord Hospital, Concord, NH, United States

Corresponding Author:
Patrice D Tremoulet, PhD
Department of Psychology
Rowan University
201 Mullica Hill Rd
Robinson Hall Room 115K
Glassboro, NJ, 08028
United States
Phone: 1 8562564500 ext 53777
Email: tremoulet@rowan.edu

Abstract

Background: Obtaining accurate clinical information about recent acute care visits is extremely important for outpatient providers. However, documents used to communicate this information are often difficult to use. This puts patients at risk of adverse events. Elderly patients who are seen by more providers and have more care transitions are especially vulnerable.

Objective: This study aimed to (1) identify the information about elderly patients' recent acute care visits needed to coordinate their care, (2) use this information to assess discharge summaries, and (3) provide recommendations to help improve the quality of electronic health record (EHR)–generated discharge summaries, thereby increasing patient safety.

Methods: A literature review, clinician interviews, and a survey of outpatient providers were used to identify and categorize data needed to coordinate care for recently discharged elderly patients. Based upon those data, 2 guidelines for creating useful discharge summaries were created. The new guidelines, along with 17 previously developed medical documentation usability heuristics, were applied to assess 4 simulated elderly patient discharge summaries.

Results: The initial research effort yielded a list of 29 items that should always be included in elderly patient discharge summaries and a list of 7 "helpful, but not always necessary" items. Evaluation of 4 deidentified elderly patient discharge summaries revealed that none of the documents contained all 36 necessary items; between 14 and 18 were missing. The documents each had several other issues, and they differed significantly in organization, layout, and formatting.

Conclusions: Variations in content and structure of discharge summaries in the United States make them unnecessarily difficult to use. Standardization would benefit both patients, by lowering the risk of care transition–related adverse events, and outpatient providers, by helping reduce frustration that can contribute to burnout. In the short term, acute care providers can help improve the quality of their discharge summaries by working with EHR vendors to follow recommendations based upon this study. Meanwhile, additional human factors work should determine the most effective way to organize and present information in discharge summaries, to facilitate effective standardization.

(J Med Internet Res 2021;23(4):e25657) doi:10.2196/25657

KEYWORDS

discharge summary; usability; electronic health record (EHR); care coordination; elderly patients; patient safety; heuristic evaluation; human factors

Introduction

Rising rates of burnout among outpatient clinicians have been linked to the adoption of electronic health records (EHRs) [1,2], which are known to have poor usability [3,4]. Other work has shown that updating EHRs is a burden for many outpatient providers [5,6]. However, few researchers have explored how inpatient providers' adoption of EHRs impacted outpatient clinicians' ability to assimilate information about recently discharged patients [7,8].

Most acute care facilities in the United States currently use EHRs to generate documents intended to communicate important information about patients and their recent acute care visits to outpatient providers, including skilled nursing facilities (SNFs), and long-term care facilities. These may be called transition of care documents, clinical handover documents, end-of-visit summaries, or, most commonly, discharge summaries (DSs). Since each acute care organization's EHR system is "customized" during implementation, outpatient providers who treat patients that utilize different acute care facilities must extract information from DSs that can vary greatly in content, format, and organization. Moreover, EHR customization is performed by information technology professionals who work for EHR vendors or inpatient providers. These professionals do not possess a complete understanding of the data needs, priorities, and information presentation preferences of outpatient providers. Not surprisingly, acute care and outpatient providers have different impressions about how useful EHR-generated DSs are [9].

DSs that either make it difficult for outpatient providers to find, or do not include, all of the information needed to coordinate care put patients at risk of adverse events [10]. Elderly patients (65 years old or older) are especially vulnerable: They tend to have more comorbidities and thus are often followed by multiple outpatient specialists, and they are more likely to have multiple postacute care transitions than younger, less complex patients. One study found discrepancies between medication lists from a referring hospital and a home health care agency for *all* 770 elderly patient participants [11], putting them at risk of medication errors. Other care transition–related adverse events include treatment delays and unnecessary tests [12-14]. However, a literature search yielded only 1 study, conducted in Canada nearly a decade ago, that suggests creating specialized DSs for elderly patients [15].

A standard method of presenting information in US discharge summaries could boost patient safety by (1) facilitating information transfer, by making it easier for outpatient clinicians to find the data needed to coordinate care for their patients and (2) reducing risk of burnout-related adverse events, by decreasing frustration associated with trying to extract information from poorly organized documents [16]. A first step towards standardizing all US DSs is identifying the data that US outpatient providers need to coordinate care for recently discharged patients who are transitioning to SNFs. We focused on this scenario because we judged that medically complex elderly patients would be more likely to transition from acute care to a SNF setting than to other outpatient care settings. Moreover, the data needed to coordinate care for adults transitioning to SNFs should be a superset of the data needed to coordinate care for other elderly patients and younger adults.

This study addresses 2 gaps in knowledge: (1) Through qualitative methods, it *identifies the information that clinicians in US SNFs need to coordinate care for elderly patients* who were recently discharged from acute care and (2) through a well-established human factors technique, called heuristic evaluation, it *provides insight into how well US acute care providers' DSs currently support outpatient providers* who coordinate care for elderly patients.

Methods

The study included 2 phases. First, the data requirements of outpatient providers who coordinate care for elderly patients were identified through an exploratory, descriptive effort, and the knowledge gleaned from that effort was applied to specify 2 new guidelines for creating useful DSs. Second, the new guidelines were combined with 17 previously developed medical documentation usability heuristics [7] (see Multimedia Appendix 1) and applied to assess how well 4 examples of elderly patient discharge summaries support care coordination. The second part not only shed light on how difficult it can be for outpatient providers to use the discharge summaries currently being produced by inpatient providers' EHRs but also yielded several recommendations for improving the quality of EHR-generated discharge summaries. All research processes and procedures for both parts of this study were approved by Rowan University's Institutional Review Board, and exemptions were granted by the institutional review boards of the 2 hospitals that provided deidentified examples of discharge summaries.

Development of DS Content Guidelines

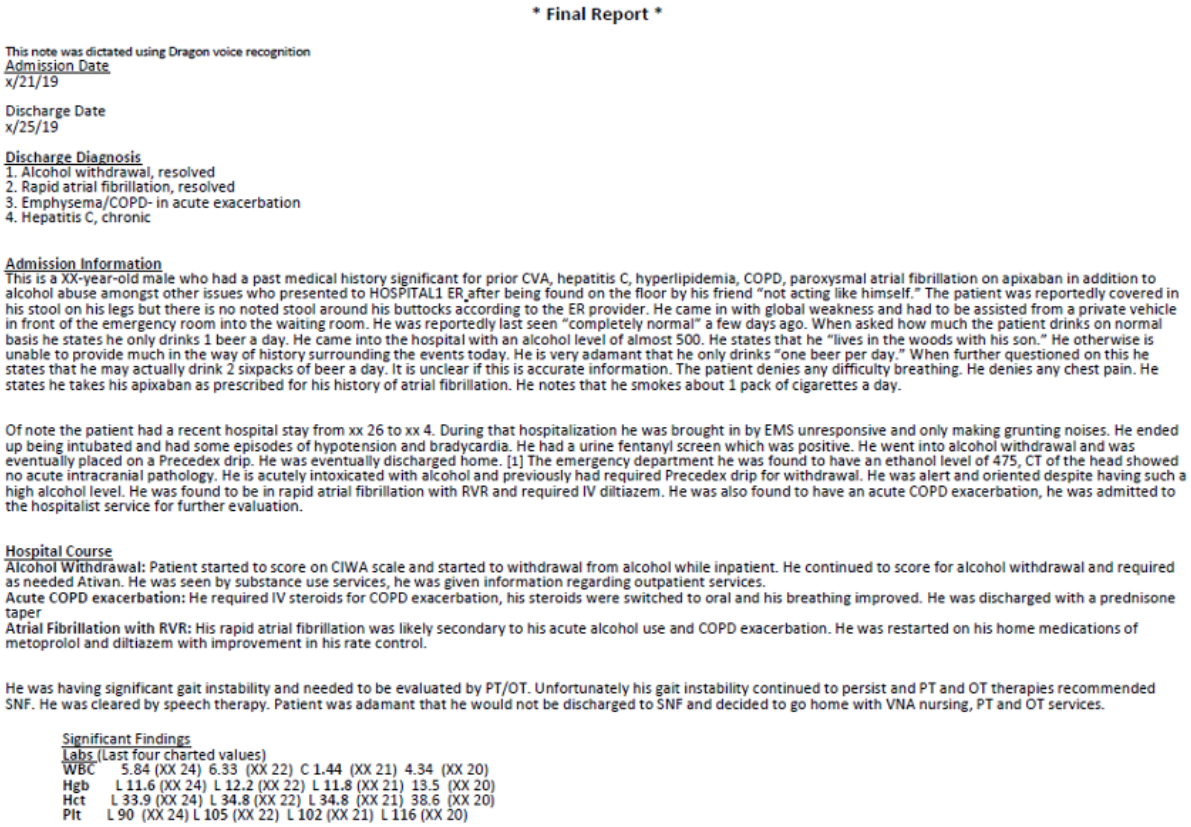
A literature review served as the first step towards identifying the data that are necessary or helpful when coordinating care for a recently discharged elderly patient. Aggregating the items identified at the Transitions of Care Consensus Conference [17], which includes 6 items that the Joint Commission mandates be included in all DSs [18], with the items in the standardized DSs used in Australia [19] and the items recommended in a Canadian study that focused specifically upon creating DSs for elderly patients [15], yielded a list of 26 items. See Multimedia Appendix 2 for items recommended by different sources. Next, 15 outpatient care providers who frequently care for elderly patients were interviewed. These care providers included primary care physicians, nurse practitioners, directors of nursing, social workers, transition-of-care nurses, and medical directors. They were asked to categorize each of those 26 items as "always necessary," "helpful, but not required," or "not relevant/distracting" and then invited to name any additional

data that they recommend be included in elderly patient DSs. After the 15 interviews, the list had grown to 36 items. See Multimedia Appendix 3 to view the structured interview questions. Those items were included in an online survey sent to 2500 members of a Continuing Care Risk Management community that asked participants to categorize the 36 items using the same 3 categories. See Multimedia Appendix 4 to view the survey questions. A majority of the 58 respondents indicated that all 36 items were helpful and that 29 of the 36 should always be included in elderly patient DSs. Thus, 2 content guidelines were established: One states that each of 29 items should be included in DSs, and the second recommends that the remaining 7 items be considered for inclusion (see Results section).

Creation of Simulated DSs

Two hospitals, which use systems from different EHR vendors, each provided 10 deidentified elderly patient DSs that were produced by their EHR system. Two of the DSs from each hospital were randomly selected. Simulations of those 4 DSs, which appeared the same as the original documents (eg, same font size and style, layout, headings), were created to keep heuristic evaluation participants blind to the hospitals. The safe harbor method was applied in accordance with the Health Insurance Portability and Accountability Act of 1996 Privacy Rule. This entailed replacing not only all protected health information about the patient (which had already been deidentified) but also all doctors' names and all health care organization names with fictitious data. The 4 simulated DSs, which are referred to as H1P1 (Hospital 1, Patient 1), H1P2 (Hospital 1, Patient 2), H2P1 (Hospital 2, Patient 1), and H2P2 (Hospital 2, Patient 2), were then reviewed for validity, including faithful replication of formatting, layout, and organization. Figures 1 and 2 show portions of 2 of the simulated DSs. See Multimedia Appendices 5-8 to view the full simulated DSs.

Figure 1. The top portion of a simulated discharge summary from hospital 1.



*** Final Report ***

This note was dictated using Dragon voice recognition

Admission Date
x/21/19

Discharge Date
x/25/19

Discharge Diagnosis
1. Alcohol withdrawal, resolved
2. Rapid atrial fibrillation, resolved
3. Emphysema/COPD- in acute exacerbation
4. Hepatitis C, chronic

Admission Information
This is a XX-year-old male who had a past medical history significant for prior CVA, hepatitis C, hyperlipidemia, COPD, paroxysmal atrial fibrillation on apixaban in addition to alcohol abuse amongst other issues who presented to HOSPITAL1 ER after being found on the floor by his friend "not acting like himself." The patient was reportedly covered in his stool on his legs but there is no noted stool around his buttocks according to the ER provider. He came in with global weakness and had to be assisted from a private vehicle in front of the emergency room into the waiting room. He was reportedly last seen "completely normal" a few days ago. When asked how much the patient drinks on normal basis he states he only drinks 1 beer a day. He came into the hospital with an alcohol level of almost 500. He states that he "lives in the woods with his son." He otherwise is unable to provide much in the way of history surrounding the events today. He is very adamant that he only drinks "one beer per day." When further questioned on this he states that he may actually drink 2 sixpacks of beer a day. It is unclear if this is accurate information. The patient denies any difficulty breathing. He denies any chest pain. He states he takes his apixaban as prescribed for his history of atrial fibrillation. He notes that he smokes about 1 pack of cigarettes a day.

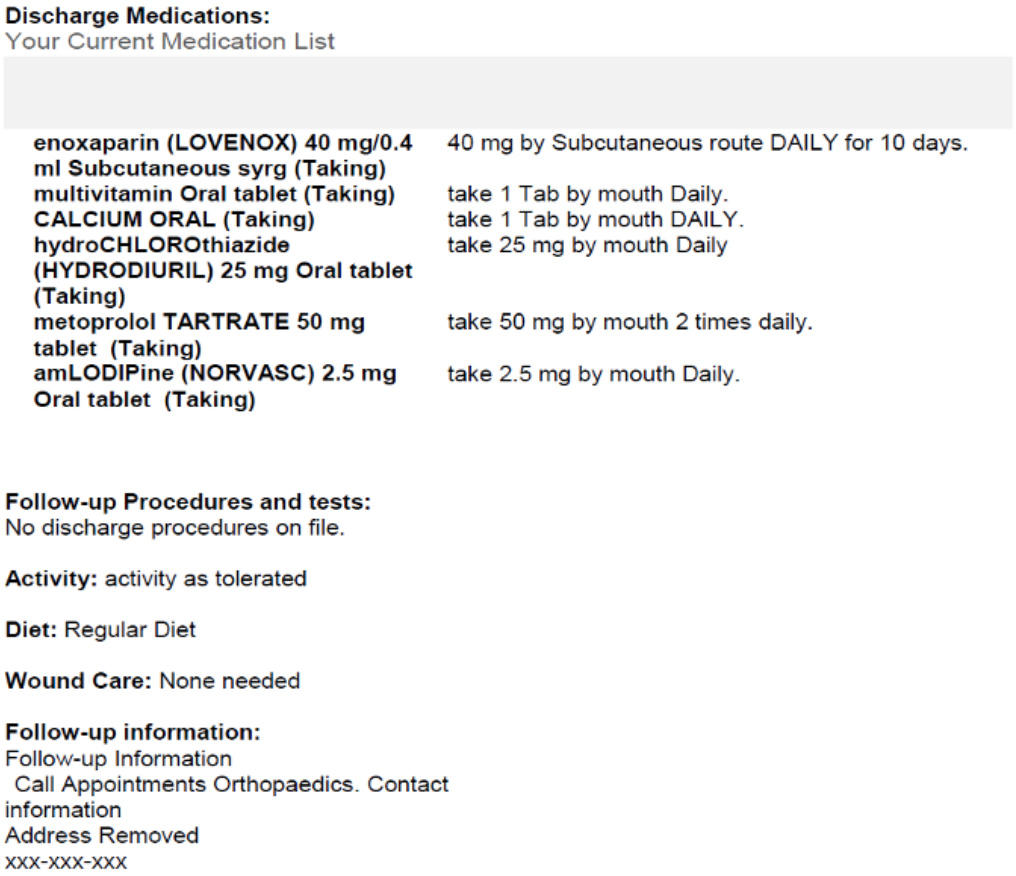
Of note the patient had a recent hospital stay from xx 26 to xx 4. During that hospitalization he was brought in by EMS unresponsive and only making grunting noises. He ended up being intubated and had some episodes of hypotension and bradycardia. He had a urine fentanyl screen which was positive. He went into alcohol withdrawal and was eventually placed on a Precedex drip. He was eventually discharged home. [1] The emergency department he was found to have an ethanol level of 475, CT of the head showed no acute intracranial pathology. He is acutely intoxicated with alcohol and previously had required Precedex drip for withdrawal. He was alert and oriented despite having such a high alcohol level. He was found to be in rapid atrial fibrillation with RVR and required IV diltiazem. He was also found to have an acute COPD exacerbation, he was admitted to the hospitalist service for further evaluation.

Hospital Course
Alcohol Withdrawal: Patient started to score on CIWA scale and started to withdrawal from alcohol while inpatient. He continued to score for alcohol withdrawal and required as needed Ativan. He was seen by substance use services, he was given information regarding outpatient services.
Acute COPD exacerbation: He required IV steroids for COPD exacerbation, his steroids were switched to oral and his breathing improved. He was discharged with a prednisone taper
Atrial Fibrillation with RVR: His rapid atrial fibrillation was likely secondary to his acute alcohol use and COPD exacerbation. He was restarted on his home medications of metoprolol and diltiazem with improvement in his rate control.

He was having significant gait instability and needed to be evaluated by PT/OT. Unfortunately his gait instability continued to persist and PT and OT therapies recommended SNF. He was cleared by speech therapy. Patient was adamant that he would not be discharged to SNF and decided to go home with VNA nursing, PT and OT services.

Significant Findings
Labs (Last four charted values)
WBC 5.84 (XX 24) 6.33 (XX 22) C 1.44 (XX 21) 4.34 (XX 20)
Hgb L 11.6 (XX 24) L 12.2 (XX 22) L 11.8 (XX 21) 13.5 (XX 20)
Hct L 33.9 (XX 24) L 34.8 (XX 22) L 34.8 (XX 21) 38.6 (XX 20)
Plt L 90 (XX 24) L 105 (XX 22) L 102 (XX 21) L 116 (XX 20)

Figure 2. The middle portion of a simulated discharge summary from hospital 2.



Discharge Medications:
Your Current Medication List

enoxaparin (LOVENOX) 40 mg/0.4 ml Subcutaneous syrg (Taking)	40 mg by Subcutaneous route DAILY for 10 days.
multivitamin Oral tablet (Taking)	take 1 Tab by mouth Daily.
CALCIUM ORAL (Taking)	take 1 Tab by mouth DAILY.
hydroCHLOROthiazide (HYDRODIURIL) 25 mg Oral tablet (Taking)	take 25 mg by mouth Daily
metoprolol TARTRATE 50 mg tablet (Taking)	take 50 mg by mouth 2 times daily.
amLODIPine (NORVASC) 2.5 mg Oral tablet (Taking)	take 2.5 mg by mouth Daily.

Follow-up Procedures and tests:
No discharge procedures on file.

Activity: activity as tolerated

Diet: Regular Diet

Wound Care: None needed

Follow-up information:
Follow-up Information
Call Appointments Orthopaedics. Contact information
Address Removed
xxx-xxx-xxx

Heuristic Evaluation

Heuristic evaluation is a usability assessment technique in which 3-5 trained experts independently apply a set of design best practices, called heuristics, to identify potential usability problems [20]. The participants also rate the severity of the issues, and then all issues and ratings are analyzed collectively [21]. Participants are not expected to discover the same issues, but together they generally identify the most important usability problems [22]. Proponents of heuristic evaluation recommend developing a set of relevant heuristics for the particular type of item being assessed, which can serve as guidance for developing useable products as well as tools for assessing usability [20].

In this study, 5 human factors experts independently assessed each of the 4 simulated discharge summaries. Next, 5 clinical experts experienced in providing outpatient care each independently rated the severity of all of the potential issues identified by the human factors team. Clinical experts were also given the opportunity to report any additional issues they found in the simulated discharge summaries. This iterative approach to heuristic evaluation, which reduces the time required by clinical experts, has been used successfully to evaluate medical technology [7,23]. This approach is also consistent with Nielsen's recommendation that experts first be asked to use heuristics to identify issues and later be asked to rate the severity of all issues [24].

The human factors team was given a set of 19 assessment tools: 17 previously developed medical document usability heuristics [7] (see Multimedia Appendix 1) and the 2 new discharge summary content guidelines (see Results section) and instructions on how to apply the heuristics to assess discharge summaries. The 5 human factors team members independently identified potential usability issues by looking for violations of the heuristics or the guidelines. Once each team member had evaluated each document, the issues were aggregated into 4 lists (one per simulated discharge document). Then, the items in the lists were paraphrased, and duplicates were removed.

Each of the 4 issue lists was ordered and grouped based upon the way information was presented in the simulated discharge summaries and then described in a set of slides (see Figure 3). Next, 5 clinical experts independently reviewed the simulated discharge documents, provided severity ratings for each issue, and reported and rated any additional usability issues they identified. The clinical experts were provided with a 4-step severity scale. We adapted Sauro's 3-point scale [25] by adding a level 0 for "not an issue" and then replacing the term "critical" with the word "severe" for level 3. The latter change was intended to prevent participants from avoiding use of that rating; in prior work, some clinicians were reluctant to use "critical," reserving that for issues that indisputably cause harm [7]. Table 1 describes the levels and provides an example of an issue at each level.

Figure 3. Screenshot of a slide used to summarize potential issues found by the human factors team.

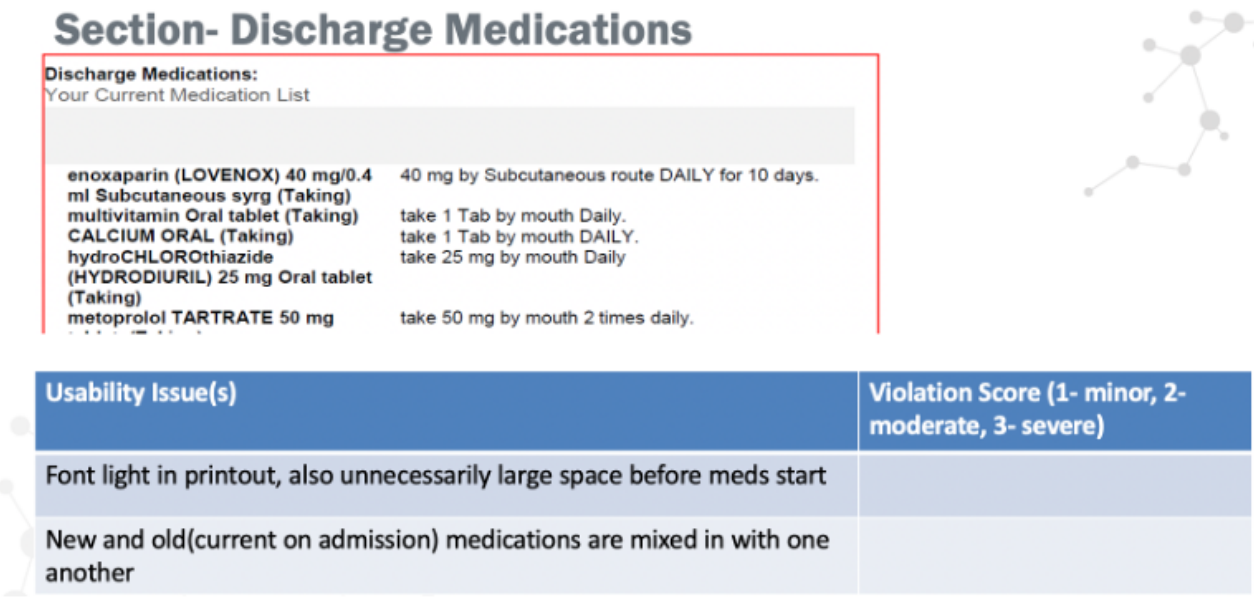


Table 1. Descriptions of the 4 levels used to rate the severity of issues found in simulated discharge summaries and examples of issues at each level.

Rating	Description	Example issue
0	Not a problem	Note stating that contents were produced using Dragon Dictate not needed
1	Minor: hesitation or slight irritation	Redundant information found in hospital course
2	Moderate: causes delays and moderate irritation	Nutritional status on discharge and Do not resuscitate orders “hidden” in relatively lengthy hospital course section
3	Severe: causes frustration or potential for error; must be fixed	No indication of whether medication is temporary or permanent

In some cases, the evaluators suggested specific solutions to issues (eg, restrict use of all CAPS, never abbreviate medication names, include page numbers in the format “page X of Y”). All solutions were summarized and grouped so that common themes could be extracted. These themes were transformed into recommendations (see Discussion section).

Results

The knowledge assembled during the literature review, interviews, and survey was applied to develop 2 guidelines for creating useful, easy-to-use discharge summaries. These guidelines are shown in Tables 2 and 3.

Table 2. First new guideline, with the list of items that should be included in elderly patient discharge summaries and how many of the 4 simulated discharge summaries were missing each item.

Guideline 1. Include each of the following in elderly patient discharge summaries	Number of documents missing this item (n=4)
Patient identifiers (eg, LAST name, first name, middle name, date of birth, age, gender, medical record number)	2
Date of admission/discharge	0
Hospital admission diagnosis	0
Principal/primary diagnosis (responsible for the largest portion of the patient's stay)	3
List of discharge diagnosis	0
Discharge medications (when patient is discharged from acute care)	0
History of present illness for hospitalization	2
Hospital course (events occurring during patient's hospital stay)	0
Procedures performed in hospital	0
Laboratory tests and investigation results (including pending results & tests due)	3
Patient physical and cognitive functional ability at discharge	3
Discharge status/Patient's discharge condition (how the patient is doing, relevant physical findings, patient's health status)	4
Medication on hospital admission	4
(Reasons for) Changes in medication during patient's stay in hospital	4
Adverse reactions during stay (including allergies to medications and other allergies)	4
Discharge instructions	2
Appointments after discharge	2
Life-sustaining treatments preferences (DNR[a], lifesaving instructions, POLST[b])	3
Nutritional status at discharge from hospital	3
Immunization history	4
Patient demographics (eg, address, phone number)	4
Follow-up issues	1
Emergency contact information	4
Nutritional status at hospital admission	4
Patient's physical and cognitive functional ability at admission	3
Goals of care and treatment plan during hospital stay	4
Discharging physician contact information	4
Contact information of clinician(s) who consulted patient in hospital	4
Patient weight	3

[a]DNR: do not resuscitate.

[b]POLST: portable medical orders.

Table 3. Second new guideline, with the list of items that should be considered to be included in elderly patient discharge summaries and how many of the 4 simulated discharge summaries were missing each item.

Guideline 2. Consider including the following items in elderly patient discharge summaries	Number of simulated documents missing this item (n=4)
Family history	4
Social and lifestyle history	4
Free-text comments (a field for clinicians to share miscellaneous notes about the patient)	3
Type of medical devices or equipment (eg, bariatric beds) that are needed for the patient in the SNF[a]	4
Goals of care and treatment plan post hospital discharge	4
Activities of daily living (ADL) status	4
Wound, skin, fall assessment	1

[a]SNF: skilled nursing facility.

When these 2 guidelines were combined with 17 previously developed medical document usability items (shown in Multimedia Appendix 1) and applied to assess the 4 simulated discharge documents, the human factors team identified 98 issues. See Multimedia Appendices 9-12 for the Powerpoint slides used to present these 98 issues to clinical experts. Clinical experts identified 19 additional issues. The average severity rating across all issues and all documents was 1.69, between minor and moderate. The total number of issues and average severity ratings for each of the simulated DSs were as follows: H1P1: 40, 1.91; H1P2: 30, 1.67; H2P1: 21, 1.73; H2P2: 26, 1.36. The entire set of issues was combined and then grouped based upon how they impact the 5 categories of document usability identified in a previous study [7]: readability, comprehensibility, minimalism, content, and organization. The counts and average severity ratings of the issues associated with each of these categories were as follows: comprehensibility: 8, 2.00; content: 42, 1.99; organization: 31, 1.77; readability: 36, 1.27; minimalism: 0. Finally, Table 4 lists all issues with average severity ratings greater than or equal to 2.

Table 4. Usability issues with average severity ratings ≥2.

Issue	Average severity rating
Postoperative patient with no wound care instructions	3
Treatable condition listed in discharge diagnoses, but patient status, plan, and medication missing	3
No indication whether medications are temporary or permanent[a]	3
Formatting makes medication section difficult to read or understand[a]	3
Medications need to be prioritized[a]	3
Medication frequency missing[a]	3
Length of time on IV[b] meds is missing[a]	3
No diagnosis linked to the medications[a]	3
Medication is missing information on “PRN” (when is it needed?) [a]	3
New and “old” medications are mixed in with one another[a]	2.67
Electronic signature gives name but no phone or email	2.67
Indenting gives impression that patient discharge condition and time spent on discharge are subordinate to physical exam	2.5
Nutritional status and Do Not Resuscitate orders “hidden” in Hospital Course section	2.33
No page numbers	2.13
Follow-up appointments listed in Hospital Course section	2.08
Heading with no content	2
Formatting made significant findings section hard to read	2

[a]Related to the section(s) listing medications.

[b]IV: intravenous.

Each of the clinical experts indicated that the simulated discharge summaries were representative of discharge summaries that they had seen before. They further characterized the examples as "fairly good" in quality when asked to compare them to those that they typically encounter. However, they agreed that each had significant room for improvement. They also agreed that it would be beneficial if DSs from all acute care providers could be standardized.

Discussion

Principal Findings

The knowledge developed through this effort may be broadly applied by acute care organizations seeking to assess or improve the usability of their discharge documents. More specifically, acute care providers can work with EHR vendors to apply the 2 guidelines introduced here to help ensure that they produce discharge summaries that contain the information that outpatient providers need to coordinate care for elderly patients. In addition, acute care providers should consider implementing the recommendations listed in the following sections, which are based upon the results of assessing 4 simulated discharge summaries that were based on real documents. These recommendations have been divided into 2 groups: The first contains those that can be implemented in software, and the second contains those that must be fulfilled by humans — though EHR software could definitely provide prompts or help verify these recommendations have been followed. See Multimedia Appendix 13 to review how several of the issues identified by clinical experts are associated with particular recommendations.

Recommendations

Textbox 1 and Textbox 2 include the recommendations for adapting EHRs so they produce more useable discharge summaries.

Textbox 1. Recommendations for adapting electronic health records (EHRs).

1. Require users to make sure that medication information is complete.
   a. For each medication, require and display the following: medication name; medication strength, dose, dosage unit, route of administration, and frequency; indication for the medication (which diagnosis or complaint is targeted by this medication; if PRN, ensure that the "as needed" criteria are defined within the medication section of the discharge summary); start date; end or refill date; indicator of whether the medication is temporary (end at completion of course) or permanent or chronic (will need refills).
   b. If any information (dose, frequency, end or refill date) is missing, prompt users to insert missing information.
2. Present all medication information clearly and consistently.
   a. Apply formatting and layout to draw attention to needed information (eg, bold the medication name, separate information with spaces or dashes).
   b. Do not use abbreviations.
   c. List medications by generic name only, not a mix of generic and brand names.
   d. Differentiate new start versus continued medications versus medications that should be stopped.
3. Provide a section for durable medical equipment, so it can be separated from medications.
4. Require and display contact information so outpatient providers can follow-up with an inpatient provider.
   a. Prompt for an email address or phone number if only a name is given.
5. Display start and end dates for all procedures.
   a. Prompt for dates if any are missing.
6. Print page numbers, in the form "xx of yy pages" when documents are printed, and display page numbers when viewed online to facilitate conversations where one person is viewing a print out and one is viewing online.
7. Use colors or shading that provide sufficient contrast if documents are printed in black and white.
8. Show all headings, even if there is no content in a section.
   a. When no accompanying text is provided, prompt for input or obtain user's approval to populate it with "—" so readers can verify that the section has been intentionally left blank.
9. Don't allow page breaks between section headings and section content.
10. Apply consistent font style, font sizing, spacing, layout, indentation, and heading style.
    a. Font size of the text must be at least 12 points to be easily readable. In some cases (eg, older audiences), selective use of a larger font (14 points) may be advisable, since it can help readers more easily see and focus attention on the most important information. Since the font size depends on the font type selected, maintain a size of 16 pixels at minimum [26].
    b. Maintain a line height that is 130% to 150% larger than the font size [27].
    c. Ensure that the section headings and subheadings stand out. Consider increasing the font size, or bold the heading text.

Textbox 2. Recommendations that require user action (but that software can prompt or try to verify).

1. Provide an explanation of patient's condition that includes at least a grade (eg, poor, good) and add a justification for any deviation from good.
2. Do not abbreviate medication or procedure names (eg, AMOX for amoxicillin) to ensure absolute clarity of the conveyed information.
3. Avoid using "all caps," which makes the content hard to read.
4. Emphasize important information. Ensure that the date of exam(s) or lab test(s) is prominently mentioned.
5. Ensure that all procedures undergone during an acute visit are listed. Examples include the use of feeding tubes, dietary restrictions, total parenteral nutrition (TPN), or urinary catheter.
6. Ensure that the content matches the headings and subheadings within each section.

Acute care providers may also consider working with EHR vendors to redesign after visit summaries (AVSs) based upon the these recommendations, since prior research has revealed that AVSs are frequently used to develop care plans [7]. Using an AVS, which is the document that is presented to a patient upon discharge from acute care, as a surrogate for a DS is not optimal practice. However, given the current practical limitations in EHR system provider-to-provider communication, it is in patients' best interests for acute care providers to adapt their AVSs so they contain the information that outpatient physicians need to coordinate care for recently discharged patients.

Limitations

While this work represents a necessary starting point for eventually standardizing adult patient discharge summaries, it has several limitations. Since the only incentive offered to survey recipients was access to survey results, only 58 out of 2500 (2.32%) of them responded, even after sending multiple reminders. Accordingly, the lists of important "discharge summary components" identified in this study need to be verified by a larger number of outpatient providers. Furthermore, a larger sample of discharge summaries or other usability testing methods may have revealed more issues, leading to additional recommendations. Finally, given the complexity and high level of customization of hospital EHR systems, it is currently unclear how difficult it would be for hospitals to configure their systems to output documents that follow our recommendations or to support users in following them. For example, some current systems are designed to group medication list items into Stop, Start, and Continue categories, which could make it hard to separate durable medical equipment from medications. On the other hand, it would make sense for inpatient providers to request that EHR vendors use the recommendations from this study to make changes to their products that could then be pushed out as software upgrades to *all* inpatient providers.

Comparison With Prior Work

A heuristic evaluation of 4 AVSs generated by acute care providers' EHR systems revealed formatting and organizational issues that made those documents very difficult to use [7]. The DSs evaluated in this study had fewer formatting and organizational issues, and those issues were rated lower in terms of severity than the ones found in the AVSs. The DSs in this study also had much less irrelevant or unhelpful text than the AVSs in the prior study. On the other hand, there were many more issues related to missing or unclear content in the DSs, and those issues were generally considered to be fairly severe (rated 2-3). This is not surprising because the content guidelines developed for this effort had not been established when the AVSs were evaluated, so their content was only assessed superficially, using the generic heuristics found in Multimedia Appendix 1.

Australia and the United Kingdom have developed national standards for electronic DSs [18,28], and Canadian researchers have explored standardizing DSs in Nova Scotia [29]. In contrast, most US research aimed at increasing patient safety during care transitions focuses upon improving discharge planning processes (eg, training clinicians to generate more useful discharge documentation [30,31] or involving patients and caregivers in discharge planning [31,32]).

A few US organizations have developed templates, outlines, or checklists to standardize their own DSs [33-35], but each of them organizes patient information differently. Moreover, even though a 2009 Transitions of Care Consensus Conference produced a list of items that participants recommended be included in transition records [17], recent research indicates that EHR-generated discharge summaries in the United States continue to omit data that outpatient providers need to effectively coordinate care [27,36-39].

Conclusions

In summary, this project addresses 2 current gaps in knowledge among inpatient providers and EHR vendors: (1) What data do outpatient providers need to coordinate care for elderly patients recently discharged from acute care facilities to an outpatient facility? and (2) How well do the EHR-generated DSs currently being produced by acute care providers meet outpatient providers' needs? This work revealed not only that documents currently being produced do not fully meet outpatient provider needs but also that 2 DSs produced by the same acute care provider organization can vary significantly in layout, organization, structure, and content. The current heterogeneity among DSs makes it unnecessarily difficult for outpatient providers to coordinate care for recently discharged patients. This puts patients at risk of adverse events and may contribute to outpatient provider burnout. This work is especially timely given that the Office of the National Coordinator for Health Information Technology and the Centers for Medicare and Medicaid Services recently released proposed rules to facilitate "seamless and secure" electronic transfer of patient data [40]. Seamlessly and securely sharing patient information is not sufficient to ensure high-quality care coordination. Patient data must be delivered in a form that enables clinicians to quickly and easily locate and understand the information most relevant to them. In short, there is an urgent need for additional applied

human factors research focused upon improving the quality of the clinical documentation produced by EHR systems. This research should be conducted in parallel with ongoing interoperability efforts, so that once it is possible for EHRs to seamlessly transfer patient data, they will be shared in a concise, well-organized, easy-to-understand form.

Acknowledgments

This work was funded by ECRI. A guidance article based upon the recommendations derived from this work was privately published, with access limited to ECRI's hospital members, in December of 2019. The authors gratefully acknowledge the valuable contributions from the following individuals: Elizabeth Anderson, Paul Anderson, Ellen Bernier, Joseph Conti, Ann Dimapalis, Kelly DiPaolo, Yvonne Fast, Dave Jamison, Susan Jilinski, Heather Kern, Nami Kim, Joseph Libby, Amy Moeller, Ankur Patel, Emily Patterson, Lorraine Possanza, Vic Rose, Firas Saidi, Karen Schoelles, Erin Sparnon, Richard Stefanacci, Ilene Warner-Maron. The authors also wish to thank an anonymous reviewer for suggesting that we add the recommendation that durable medical equipment be separated from medications.

Authors' Contributions

PT and PS designed the study, analyzed the data, and generated the recommendations. The human factors evaluation team consisted of PT, AA, CG, JK, and PM. Deidentification of actual elderly patient discharge summaries was performed by MK and EW. PS created the simulated discharge summaries from the deidentified examples. The initial draft of the paper was produced by PT.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Medical Document Usability Heuristics adapted from Tremoulet et al, 2018.
[DOCX File , 15 KB - jmir_v23i4e25657_app1.docx]

Multimedia Appendix 2
Discharge Summary components recommended by different sources.
[DOCX File , 15 KB - jmir_v23i4e25657_app2.docx]

Multimedia Appendix 3
Outpatient Provider Structured Interview Questions.
[DOCX File , 34 KB - jmir_v23i4e25657_app3.docx]

Multimedia Appendix 4
Survey used to help develop two new discharge summary heuristics.
[DOCX File , 25 KB - jmir_v23i4e25657_app4.docx]

Multimedia Appendix 5
Simulated Discharge Summary for patient 1 from hospital 1.
[PDF File (Adobe PDF File), 130 KB - jmir_v23i4e25657_app5.pdf]

Multimedia Appendix 6
Simulated Discharge Summary from patient 2 from hospital 1.
[PDF File (Adobe PDF File), 103 KB - jmir_v23i4e25657_app6.pdf]

Multimedia Appendix 7
Simulated Discharge Summary for patient 1 from hospital 2.
[PDF File (Adobe PDF File), 73 KB - jmir_v23i4e25657_app7.pdf]

Multimedia Appendix 8
Simulated Discharge Summary for patient 2 from hospital 2.
[PDF File (Adobe PDF File), 82 KB - jmir_v23i4e25657_app8.pdf]

Multimedia Appendix 9

Potential usability issues in simulated discharge summary for patient 1 from hospital 1.
[PPTX File , 6653 KB - jmir_v23i4e25657_app9.pptx]

Multimedia Appendix 10
Potential usability issues in simulated discharge summary for patient 2 from hospital 1.
[PPTX File , 6340 KB - jmir_v23i4e25657_app10.pptx]

Multimedia Appendix 11
Potential usability issues in simulated discharge summary for patient 1 from hospital 2.
[PPTX File , 6370 KB - jmir_v23i4e25657_app11.pptx]

Multimedia Appendix 12
Potential usability issues in simulated discharge summary for patient 2 from hospital 2.
[PPTX File , 6287 KB - jmir_v23i4e25657_app12.pptx]

Multimedia Appendix 13
Locations of several DS issues identified by clinical experts and the recommendations that are related to those issues.
[DOCX File , 17 KB - jmir_v23i4e25657_app13.docx]

References

1. Downing NL, Bates DW, Longhurst CA. Physician Burnout in the Electronic Health Record Era: Are We Ignoring the Real Cause? Ann Intern Med 2018 Jul 03;169(1):50-51. [doi: 10.7326/M18-0139] [Medline: 29801050]
2. Ehrenfeld JM, Wanderer JP. Technology as friend or foe? Do electronic health records increase burnout? Curr Opin Anaesthesiol 2018 Jun;31(3):357-360. [doi: 10.1097/ACO.0000000000000588] [Medline: 29474217]
3. Ellsworth M, Dziadzko M, O'Horo JC, Farrell A, Zhang J, Herasevich V. An appraisal of published usability evaluations of electronic health records via systematic review. J Am Med Inform Assoc 2017 Jan;24(1):218-226 [FREE Full text] [doi: 10.1093/jamia/ocw046] [Medline: 27107451]
4. Howe JL, Adams KT, Hettinger AZ, Ratwani RM. Electronic Health Record Usability Issues and Potential Contribution to Patient Harm. JAMA 2018 Mar 27;319(12):1276-1278 [FREE Full text] [doi: 10.1001/jama.2018.1171] [Medline: 29584833]
5. Reith TP. Burnout in United States Healthcare Professionals: A Narrative Review. Cureus 2018 Dec 04;10(12):e3681 [FREE Full text] [doi: 10.7759/cureus.3681] [Medline: 30761233]
6. Ommaya A, Cipriano P, Hoyt D, Horvath KA, Tang P, Paz HL, et al. Care-Centered Clinical Documentation in the Digital Environment: Solutions to Alleviate Burnout. NAM Perspectives 2018 Jan 29;8(1):1 [FREE Full text] [doi: 10.31478/201801c]
7. Tremoulet P, Krishnan R, Karavite D, Muthu N, Regli S, Will A, et al. A Heuristic Evaluation to Assess Use of After Visit Summaries for Supporting Continuity of Care. Appl Clin Inform 2018 Jul 12;9(3):714-724 [FREE Full text] [doi: 10.1055/s-0038-1668093] [Medline: 30208496]
8. National Academies of Sciences, Engineering, and Medicine. Taking Action Against Clinician Burnout: A Systems Approach to Professional Well-Being. Washington, DC: The National Academies Press; 2019.
9. Solan LG, Sherman SN, DeBlasio D, Simmons JM. Communication Challenges: A Qualitative Look at the Relationship Between Pediatric Hospitalists and Primary Care Providers. Acad Pediatr 2016 Jul;16(5):453-459. [doi: 10.1016/j.acap.2016.03.003] [Medline: 26979780]
10. Jones CD, Cumbler E, Honigman B, Burke RE, Boxer RS, Levy C, et al. Hospital to Post-Acute Care Facility Transfers: Identifying Targets for Information Exchange Quality Improvement. J Am Med Dir Assoc 2017 Jan;18(1):70-73 [FREE Full text] [doi: 10.1016/j.jamda.2016.09.009] [Medline: 27815110]
11. Brody AA, Gibson B, Tresner-Kirsch D, Kramer H, Thraen I, Coarr ME, et al. High Prevalence of Medication Discrepancies Between Home Health Referrals and Centers for Medicare and Medicaid Services Home Health Certification and Plan of Care and Their Potential to Affect Safety of Vulnerable Elderly Adults. J Am Geriatr Soc 2016 Nov 27;64(11):e166-e170 [FREE Full text] [doi: 10.1111/jgs.14457] [Medline: 27673753]
12. Jones CD, Jones J, Bowles KH, Flynn L, Masoudi FA, Coleman EA, et al. Quality of Hospital Communication and Patient Preparation for Home Health Care: Results From a Statewide Survey of Home Health Care Nurses and Staff. J Am Med Dir Assoc 2019 Apr;20(4):487-491 [FREE Full text] [doi: 10.1016/j.jamda.2019.01.004] [Medline: 30799224]
13. Gadbois EA, Tyler DA, Shield R, McHugh J, Winblad U, Teno JM, et al. Lost in Transition: a Qualitative Study of Patients Discharged from Hospital to Skilled Nursing Facility. J Gen Intern Med 2019 Jan 18;34(1):102-109 [FREE Full text] [doi: 10.1007/s11606-018-4695-0] [Medline: 30338471]

14. Moore AB, Krupp JE, Dufour AB, Sircar M, Travison TG, Abrams A, et al. Improving Transitions to Postacute Care for Elderly Patients Using a Novel Video-Conferencing Program: ECHO-Care Transitions. Am J Med 2017 Oct;130(10):1199-1204. [doi: 10.1016/j.amjmed.2017.04.041] [Medline: 28551043]
15. Kergoat M, Latour J, Julien I, Plante M, Lebel P, Mainville D, et al. A discharge summary adapted to the frail elderly to ensure transfer of relevant information from the hospital to community settings: a model. BMC Geriatr 2010 Sep 23;10(1):69 [FREE Full text] [doi: 10.1186/1471-2318-10-69] [Medline: 20863405]
16. Tawfik DS, Profit J, Morgenthaler TI, Satele DV, Sinsky CA, Dyrbye LN, et al. Physician Burnout, Well-being, and Work Unit Safety Grades in Relationship to Reported Medical Errors. Mayo Clin Proc 2018 Nov;93(11):1571-1580 [FREE Full text] [doi: 10.1016/j.mayocp.2018.05.014] [Medline: 30001832]
17. Snow V, Beck D, Budnitz T, Miller DC, Potter J, Wears RL, American College of Physicians, Society of General Internal Medicine, Society of Hospital Medicine, American Geriatrics Society, American College of Emergency Physicians, Society of Academic Emergency Medicine. Transitions of Care Consensus Policy Statement American College of Physicians-Society of General Internal Medicine-Society of Hospital Medicine-American Geriatrics Society-American College of Emergency Physicians-Society of Academic Emergency Medicine. J Gen Intern Med 2009 Aug 3;24(8):971-976 [FREE Full text] [doi: 10.1007/s11606-009-0969-x] [Medline: 19343456]
18. Kind AJ, Smith MA. Documentation of Mandated Discharge Summary Components in Transitions from Acute to Subacute Care. In: Henriksen K, Battles JB, Keyes MA, Grady ML, editors. Advances in Patient Safety: New Directions and Alternative Approaches (Vol. 2: Culture and Redesign). Rockville, MD: Agency for Healthcare Research and Quality; 2008:3715 URL: https://www.ncbi.nlm.nih.gov/books/NBK43715/#advances-kind_31.s1
19. National Guidelines for On-Screen Presentation of Discharge Summaries. Australian Commission on Safety and Quality in Health Care. 2017. URL: https://www.safetyandquality.gov.au/publications-and-resources/resource-library/national-guidelines-screen-presentation-discharge-summaries [accessed 2021-03-26]
20. Nielsen J, Molich R. Heuristic evaluation of user interfaces. 1990 Presented at: CHI90: Conference on Human Factors in Computing; April 1990; Seattle, WA. [doi: 10.1145/97243.97281]
21. Nielsen J. Reliability of severity estimates for usability problems found by heuristic evaluation. 1992 Presented at: CHI92: Conference on Human Factors in Computing; 1992; Monterey, CA p. 129-130. [doi: 10.1145/1125021.1125117]
22. Nielsen J. 10 usability heuristics for user interface design. Nielson Norman Group. 1994 Apr 24. URL: https://www.nngroup.com/articles/ten-usability-heuristics/ [accessed 2021-03-26]
23. Tremoulet PD, McManus M, Baranov D. Rendering ICU Data Useful via Formative Evaluations of Trajectory, Tracking, and Trigger (T3) Software. In: Proceedings of the International Symposium on Human Factors and Ergonomics in Health Care. 2017 May 15 Presented at: 2017 International Symposium on Human Factors and Ergonomics in Healthcare; March 2017; New Orleans, LA p. 50-56. [doi: 10.1177/2327857917061012]
24. Nielsen J. Severity ratings for usability problems. Nielson Norman Group. 1994 Nov 01. URL: https://www.nngroup.com/articles/how-to-rate-the-severity-of-usability-problems/ [accessed 2021-03-26]
25. Sauro J. Rating the Severity of Usability Problems. MeasuringU. 2013 Jul 30. URL: https://measuringu.com/rating-severity/ [accessed 2021-03-26]
26. Section 3: Display content clearly on the page. Health Literacy Online. 2016. URL: https://health.gov/healthliteracyonline/display/section-3-3/ [accessed 2021-03-26]
27. Everdell I. Web content. In: Bergstrom JR, Schall AJ, editors. Eye tracking in user experience design. Waltham, MA: Morgan Kaufmann Publishers; 2014:163-186.
28. Implementation guidance report: e-Discharge standard. Professional Record Standards Body. 2019. URL: https://theprsb.org/wp-content/uploads/2019/02/eDischarge-Summary-Maintenance-Release-Implementation-Guidance-Report-v2.1-23.1.19.pdf [accessed 2021-03-26]
29. Savvopoulos S, Sampalli T, Harding R, Blackmore G, Janes S, Kumanan K, et al. Development of a quality scoring tool to assess quality of discharge summaries. J Family Med Prim Care 2018;7(2):394-400 [FREE Full text] [doi: 10.4103/jfmpc.jfmpc_407_16] [Medline: 30090783]
30. Ming D, Zietlow K, Song Y, Lee H, Clay A. Discharge summary training curriculum: a novel approach to training medical students how to write effective discharge summaries. Clin Teach 2019 Oct 30;16(5):507-512. [doi: 10.1111/tct.12960] [Medline: 30378265]
31. Backman C, Cho-Young D. Engaging patients and informal caregivers to improve safety and facilitate person- and family-centered care during transitions from hospital to home - a qualitative descriptive study. Patient Prefer Adherence 2019 Apr;13:617-626 [FREE Full text] [doi: 10.2147/PPA.S201054] [Medline: 31114175]
32. Medicare and Medicaid Programs; Revisions to Requirements for Discharge Planning for Hospitals, Critical Access Hospitals, and Home Health Agencies, and Hospital and Critical Access Hospital Changes to Promote Innovation, Flexibility, and Improvement in Patient Care. Federal Register: The Daily Journal of the United States Government. 2019 Sep 30. URL: https://www.federalregister.gov/documents/2019/09/30/2019-20732/medicare-and-medicaid-programs-revisions-to-requirements-for-discharge-planning-for-hospitals [accessed 2021-03-26]
33. Discharge Summary Template. University of Tennessee Health Science Center. URL: https://www.uthsc.edu/pediatrics/clerkship/documents/discharge-summary-template.pdf [accessed 2021-03-26]

34. Sample Discharge Summary. Creighton University. URL: https://medschool.creighton.edu/fileadmin/user/medicine/Departments/ob-gyn/docs/Sample_Dictation_-_OB_Discharge_Summary.pdf [accessed 2021-03-26]
35. Discharge Summary Outline. WVU Medicine. URL: https://wvumedicine.org/ruby-memorial-hospital/wp-content/uploads/sites/3/2017/04/Discharge-Summary-Outline.pdf [accessed 2021-03-26]
36. Dean SM, Gilmore-Bykovskyi A, Buchanan J, Ehlenfeldt B, Kind AJ. Design and Hospitalwide Implementation of a Standardized Discharge Summary in an Electronic Health Record. The Joint Commission Journal on Quality and Patient Safety 2016 Dec;42(12):555-AP11. [doi: 10.1016/S1553-7250(16)30107-6]
37. Moy NY, Lee SJ, Chan T, Grovey B, Boscardin WJ, Gonzales R, et al. Development and Sustainability of an Inpatient-to-Outpatient Discharge Handoff Tool: A Quality Improvement Project. The Joint Commission Journal on Quality and Patient Safety 2014 May;40(5):219-227. [doi: 10.1016/s1553-7250(14)40029-1]
38. Wimsett J, Harper A, Jones P. Review article: Components of a good quality discharge summary: a systematic review. Emerg Med Australas 2014 Oct 03;26(5):430-438. [doi: 10.1111/1742-6723.12285] [Medline: 25186466]
39. Callen JL, Alderton M, McIntosh J. Evaluation of electronic discharge summaries: a comparison of documentation in electronic and handwritten discharge summaries. Int J Med Inform 2008 Sep;77(9):613-620. [doi: 10.1016/j.ijmedinf.2007.12.002] [Medline: 18294904]
40. Butler M. CMS and ONC Release Proposed Rules on Data Exchange and Interoperability. Journal of AHIMA. 2019 Feb 12. URL: https://journal.ahima.org/cms-and-onc-release-proposed-rules-on-data-exchange-and-interoperability/ [accessed 2021-03-26]

Abbreviations

AVS: after visit summary
DS: discharge summary
EHR: electronic health record
SNF: skilled nursing facility

Edited by R Kukafka; submitted 10.11.20; peer-reviewed by D Scherer, C Johnson, D Newman; comments to author 25.12.20; revised version received 11.02.21; accepted 16.03.21; published 15.04.21.

Please cite as:
Tremoulet PD, Shah PD, Acosta AA, Grant CW, Kurtz JT, Mounas P, Kirchhoff M, Wade E
Usability of Electronic Health Record–Generated Discharge Summaries: Heuristic Evaluation
J Med Internet Res 2021;23(4):e25657
URL: https://www.jmir.org/2021/4/e25657
doi:10.2196/25657
PMID:33856353

Original Paper

Real-Time Clinical Decision Support Based on Recurrent Neural Networks for In-Hospital Acute Kidney Injury: External Validation and Model Interpretation

Kipyo Kim[1*], MD, PhD; Hyeonsik Yang[2*], BSc; Jinyeong Yi[3], MA; Hyung-Eun Son[4], MD; Ji-Young Ryu[4], MD; Yong Chul Kim[5], MD, PhD; Jong Cheol Jeong[4], MD, PhD; Ho Jun Chin[4], MD, PhD; Ki Young Na[4], MD, PhD; Dong-Wan Chae[4], MD, PhD; Seung Seok Han[5], MD, PhD; Sejoong Kim[4,6], MD, PhD

[1]Division of Nephrology and Hypertension, Department of Internal Medicine, Inha University College of Medicine, Incheon, Republic of Korea

[2]Department of Computer Science and Engineering, Seoul National University, Seoul, Republic of Korea

[3]Department of Health Science and Technology, Graduate School of Convergence Science and Technology, Seoul National University, Seoul, Republic of Korea

[4]Department of Internal Medicine, Seoul National University Bundang Hospital, Seongnam, Republic of Korea

[5]Department of Internal Medicine, Seoul National University Hospital, Seoul, Republic of Korea

[6]Center for Artificial Intelligence in Healthcare, Seoul National University Bundang Hospital, Seongnam, Republic of Korea

[*]these authors contributed equally

Corresponding Author:
Sejoong Kim, MD, PhD
Department of Internal Medicine
Seoul National University Bundang Hospital
82 Gumi-ro 173-beon-gil Bundang-gu
Seongnam, 13620
Republic of Korea
Phone: 82 31 787 7051
Fax: 82 31 787 4052
Email: sejoong@snubh.org

Abstract

Background: Acute kidney injury (AKI) is commonly encountered in clinical practice and is associated with poor patient outcomes and increased health care costs. Despite it posing significant challenges for clinicians, effective measures for AKI prediction and prevention are lacking. Previously published AKI prediction models mostly have a simple design without external validation. Furthermore, little is known about the process of linking model output and clinical decisions due to the black-box nature of neural network models.

Objective: We aimed to present an externally validated recurrent neural network (RNN)–based continuous prediction model for in-hospital AKI and show applicable model interpretations in relation to clinical decision support.

Methods: Study populations were all patients aged 18 years or older who were hospitalized for more than 48 hours between 2013 and 2017 in 2 tertiary hospitals in Korea (Seoul National University Bundang Hospital and Seoul National University Hospital). All demographic data, laboratory values, vital signs, and clinical conditions of patients were obtained from electronic health records of each hospital. We developed 2-stage hierarchical prediction models (model 1 and model 2) using RNN algorithms. The outcome variable for model 1 was the occurrence of AKI within 7 days from the present. Model 2 predicted the future trajectory of creatinine values up to 72 hours. The performance of each developed model was evaluated using the internal and external validation data sets. For the explainability of our models, different model-agnostic interpretation methods were used, including Shapley Additive Explanations, partial dependence plots, individual conditional expectation, and accumulated local effects plots.

Results: We included 69,081 patients in the training, 7675 in the internal validation, and 72,352 in the external validation cohorts for model development after excluding cases with missing data and those with an estimated glomerular filtration rate less than 15 mL/min/1.73 m2 or end-stage kidney disease. Model 1 predicted any AKI development with an area under the receiver operating characteristic curve (AUC) of 0.88 (internal validation) and 0.84 (external validation), and stage 2 or higher AKI development

with an AUC of 0.93 (internal validation) and 0.90 (external validation). Model 2 predicted the future creatinine values within 3 days with mean-squared errors of 0.04-0.09 for patients with higher risks of AKI and 0.03-0.08 for those with lower risks. Based on the developed models, we showed AKI probability according to feature values in total patients and each individual with partial dependence, accumulated local effects, and individual conditional expectation plots. We also estimated the effects of feature modifications such as nephrotoxic drug discontinuation on future creatinine levels.

Conclusions: We developed and externally validated a continuous AKI prediction model using RNN algorithms. Our model could provide real-time assessment of future AKI occurrences and individualized risk factors for AKI in general inpatient cohorts; thus, we suggest approaches to support clinical decisions based on prediction models for in-hospital AKI.

(J Med Internet Res 2021;23(4):e24120) doi:10.2196/24120

KEYWORDS

acute kidney injury; recurrent neural network; prediction model; external validation; internal validation; kidney; neural networks

Introduction

Acute kidney injury (AKI) is a common clinical condition that can be attributed to multiple causes in various clinical settings. AKI increases mortality, morbidity, length of hospital stay, and health care costs [1-4]. According to the National Confidential Enquiry on Patient Outcomes and Death reports, approximately 17% of AKI is estimated to be avoidable and preventable [5]. Therefore, many research efforts have been expended to detect AKI early on and to manage patients with high risk [6]. Nevertheless, the reported incidence of AKI is 17%-25% in the hospital setting [7,8] and it has continued to rise globally during the recent decades [9]. Markedly increasing amounts of electronic health record (EHR) data and recent developments in machine learning techniques offer greater possibilities for the improvement of quality of care and medical research [10]. Various clinical decision support systems that use EHR and machine learning have been increasingly reported for various diseases [11-14]. Machine learning methods can incorporate tremendously large number of features in the model compared with conventional regression models and thus enable the use of nonlinear algorithms. As a result, in AKI research, several studies have adopted machine learning methods such as random forest and neural network models and reported improved model performance [15-17]. Major risk factors associated with in-hospital AKI include the use of various nephrotoxins, repeatedly measured laboratory findings, and vital signs that are dynamic rather than static variables [18-20]. The recurrent neural network (RNN) is a powerful tool used to handle such sequential data [21]; various RNN models have shown excellent performance in the field of natural language processing and time-series forecasting models [22,23]. Although the RNN model is a promising approach, time-updated predictive models for AKI using the RNN algorithm are still in their infancy [24]. Only a few studies have investigated these, and they have not been externally validated [20,25]. Moreover, despite their enhanced performance, these neural network models cannot provide insights into how to link clinical decision supports; therefore, interpreting the output using these models can be difficult. In this respect, neural network models have been criticized as being black-box models [26]. Previous studies on clinical decision support for AKI have not implemented predictions for AKI development but have only served as alarm systems for the timely diagnosis of AKI according to diagnostic criteria [27,28]. Therefore, in this study, we propose an externally validated RNN-based prediction model for in-hospital AKI and aimed to provide a framework to link the developed model with clinical decision support.

Methods

Study Population

This study was performed in accordance with the recommendations laid out in the World Medical Association Declaration of Helsinki. The study protocol was approved by the Institutional Review Boards (IRBs) of the Seoul National University Bundang Hospital (SNUBH; IRB No. B-1912/583-406) and Seoul National University Hospital (SNUH; IRB No. H-1911-043-1076). Written consent was waived by the IRB because of the retrospective nature of the study, and all data were completely anonymized. Study populations comprised all patients aged 18 years or older who were hospitalized for more than 48 hours at the SNUBH from 2013 to 2017 (training and internal validation cohorts) and at the SNUH from 2013 to 2017 (external validation cohort). These 2 tertiary hospitals are affiliated with each other. However, they are located in different regions of Korea and have different patient populations and EHR systems. The exclusion criteria were as follows: (1) no baseline or follow-up creatinine (Cr) measurements, (2) baseline estimated glomerular filtration rate (eGFR) less than 15 mL/min/1.73 m^2 or Cr greater than 4.0 mg/dL or end-stage kidney disease at admission, (3) no other laboratory test results used in the model, (4) no BMI or vital sign measurements, and (5) an AKI diagnosis at admission (day 1).

Data Collection

All demographics, laboratory values, vital signs, and clinical conditions were obtained from the EHR of each hospital. The features that are considered as risk factors of AKI from the related literature or are correlated with AKI development were selected for model development [18]. A total of 107 variables were included in the model. These are summarized in Multimedia Appendix 1. Each variable was classified as either static or dynamic. The static variables were assigned to time-invariant values during hospitalization, and the dynamic variables were assigned to values that were updated on a daily basis. Demographics and comorbidities were static variables, while laboratory tests, vital signs, and clinical conditions were dynamic variables. The use of medications during admission

was incorporated in the model as a dynamic categorical variable, and medication history before admission was treated as a static variable. Patient comorbidities were classified into 17 categories according to the International Statistical Classification of Disease and Related Health Problems (10th revision) codes from which the Charlson Comorbidity Index was assessed [29]. BMI was calculated from the height and weight measured at admission. Laboratory values included Cr, white blood cells, hemoglobin, platelets, albumin, sodium, potassium, chloride, aspartate aminotransferase, alanine aminotransferase, blood urea nitrogen, total CO_2, bilirubin, calcium, glucose, creatine kinase, lipase, and troponin I. The means of laboratory values that were measured more than once each day were used in our model. Medications included well-known nephrotoxic agents, such as nonsteroidal anti-inflammatory drugs, aminoglycoside, vancomycin, and colistin. Vital signs included systolic, diastolic, and mean arterial blood pressures; pulse; and body temperature, which were usually measured 3 times a day in general wards. Therefore, the mean, maximum, and minimum values of the vital signs during a day were used as different variables.

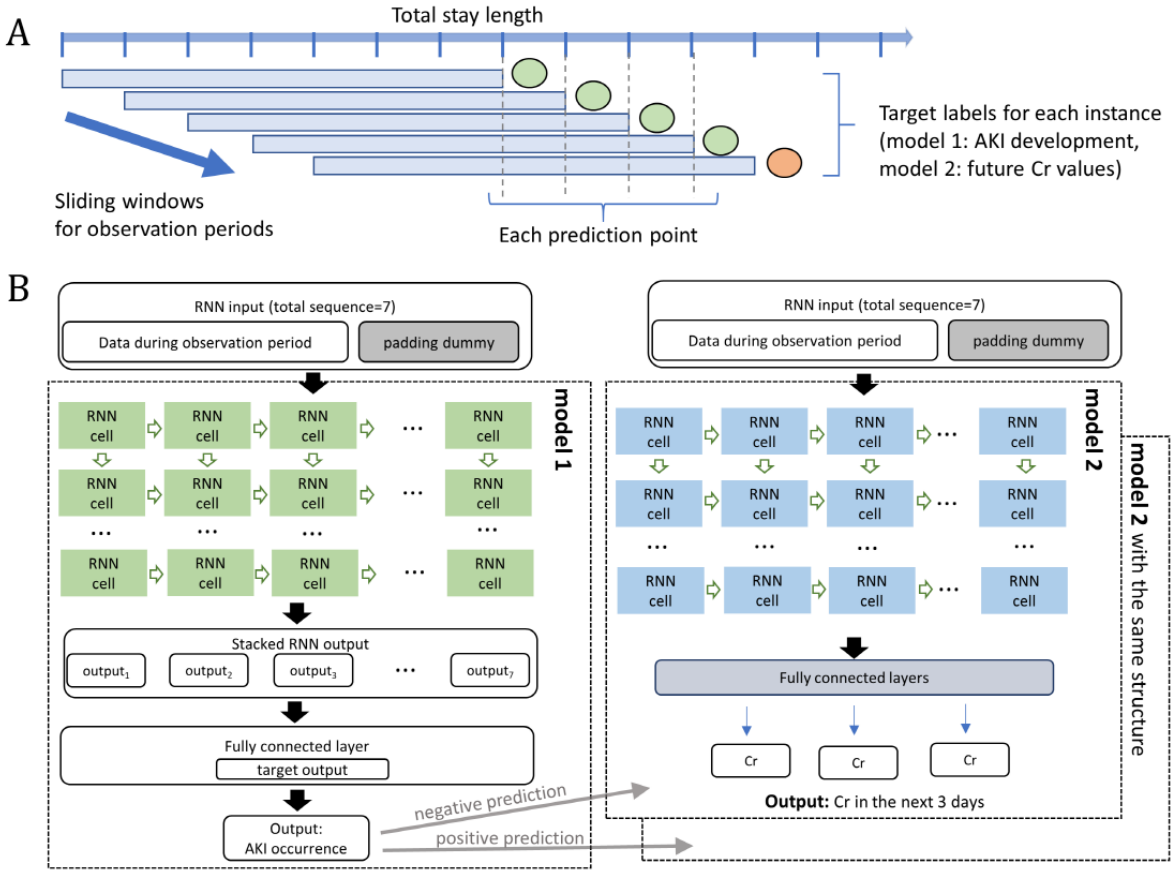
AKI and Baseline Creatinine Definitions

AKI was defined according to the Kidney Disease: Improving Global Outcomes (KDIGO) Clinical Practice Guideline for AKI [30]. Because urine volume data were not available, AKI stage was defined based on serum Cr levels. Baseline Cr levels were determined by searching the minimum serum Cr level within a period of 2 weeks before admission. If there were no serum Cr measurements during this period, the minimum value of Cr measured within 90 or 180 days before admission was used as the baseline Cr. In the absence of Cr measurements up to 180 days before admission, the serum Cr value measured on the first day of hospitalization was defined as the baseline Cr.

Data Preprocessing and Statistical Analysis

During data collection, error values were treated as missing, and outliers of all variables were examined and removed after review by domain experts. Patients with missing variables at baseline were excluded according to the exclusion criteria. The last observation carried forward method was used for missing values after baseline. Variables were scaled using the min–max normalization before training the neural network. For continuous prediction of AKI, the training and validation data sets were organized as multiple sliding windows of features and target labels fed to the input layer of the RNN model (Figure 1A). The length of the sliding window was selected as 7 days and features up to 2 weeks after admission were utilized. Therefore, all time points were considered for both patients with AKI and non-AKI during the length of stay before AKI occurrence or discharge. The SNUBH data set was divided into a training set (69,081/76,756, 90%) and an internal validation set (7675/76,756, 10%) using the stratified random split. Categorical variables are expressed as numbers and percentages, and continuous variables as means (SD). The chi-square test and *t* test were used to compare differences in baseline characteristics between the training and validation cohorts. *P* values <.05 were considered statistically significant.

Figure 1. Architectural overview of the developed model using the recurrent neural network (RNN). (A) The sliding window approach for handling sequential data and (B) the 2-stage hierarchical model comprising models 1 and 2. Model 1 could predict the occurrence of acute kidney injury (AKI) in the next 7 days from the present, and model 2 could provide the predicted value of serum creatinine (Cr) for the next 3 days.



Model Development and Assessment

We developed 2 different models (models 1 and 2) with stacked RNN algorithms. Model 1 had a many-to-one architecture with 7 sequential inputs and 1 prediction output. The outcome variable for model 1 was the occurrence of AKI in the next 7 days. Padding and masking techniques were used for instances with sequences shorter than the length of the sliding window. The padded sequences were not used in training and inference processes. To compare the performance with model 1, we developed an additional gradient boosting model based on the same training data set. The gradient boosting model was trained using the XGBoost algorithm. In model 2, we constructed a prediction model of the trajectory of Cr values after 24, 48, and 72 hours with available Cr values during the observation window. Model 2 had a many-to-many structure with an output length of 3 on 7 input sequences. To improve the predictive accuracy of model 2, we developed a 2-stage hierarchical RNN prediction model. That is, based on the results of model 1, different types of model 2 were applied. These are the models for each patient group that are or are not predicted to have AKI. The architectures of models 1 and 2 are illustrated in Figure 1. The optimal hyperparameters of each model were determined using a fivefold cross validation. The tuned hyperparameters include number of hidden RNN neurons, number of hidden layers, dropout, activation functions, and batch size in the stacked RNN model, and learning rate, depth of a tree, and minimum child weight in the XGBoost model. Cross entropy loss function and the AdamW optimizer were used to train the RNN models. The entire data set was imbalanced due to the relatively low incidence of in-hospital AKI. Therefore, we applied a class weight parameter to the loss function to handle class imbalances. Alpha dropout, L2 regularization, and early stopping approaches were implemented to prevent overfitting. Batch normalization was applied to each RNN layer for efficient and effective learning. The learning rate was set to 10^{-4} for pretraining, then to 10^{-6} for early stop learning. In the process of model development, different sample sizes were tested to evaluate the sensitivity of the developed models. As shown in Table 1 in the *Results* section, the overall distribution of features is quite different between the training and external validation data sets. Therefore, we additionally fine-tuned our models to overcome the heterogeneity of independent data sets. Specifically, we refitted our model using a small proportion of data (7234/72,352, 10%) from the external data set. The refitted model was validated with the rest of the external data set (65,118/72,352, 90%). Model performances were assessed based on the area under the receiver operating characteristic curve (AUC), accuracy, sensitivity, specificity, positive predictive value, negative predictive value, and F1 score for model 1 and based on the mean-squared error (MSE) for model 2.

Model Explainability for Clinical Decision Support

The neural network algorithm cannot directly offer any explanations regarding the clinical meaning of features.

Therefore, to identify the association between multiple features and response (the occurrence of AKI), we examined the following approaches using model-agnostic methods in model 1: (1) global interpretation with Shapley Additive Explanations (SHAP), partial dependence plots (PDPs), and accumulated local effects plots; and (2) instance-wise interpretation with individual conditional expectation (ICE) plots. The SHAP method is based on Shapley values from game theory [31]. Shapley values indicate marginal contribution of a feature to the difference between the actual prediction and the mean prediction for all possible coalitions of features [32]. The global feature importance was presented by SHAP feature importance plots. PDP provides average marginal predictions across all instances when the feature of interest is forced to be a certain value [33]. Similarly, ICE plots indicate an average marginal effect of a feature for individual instances. PDP and ICE plots can intuitively show the relationship between specific features and the outcome variable [34]. In model 2, we applied ICE to predict the future Cr values, whereby the model output was estimated from new instances with modified feature values as well. Given that PDP is not reliable when the features are highly correlated, we also presented accumulated local effects plots. In an accumulated local effects plot, feature values are divided by set intervals and the differences in the prediction between the upper and lower bounds of the interval are calculated [33]. The estimated differences are accumulated, and the mean prediction is centered at 0.

Results

Study Population

A total of 482,467 patients (182,976 in the SNUBH and 299,491 in the SNUH) were initially screened from the EHR data obtained from each participating hospital. After considering the exclusion criteria, 69,081 patients were finally included in the SNUBH training data set, 7675 in the SNUBH internal validation data set, and 72,352 in the SNUH external validation data set (Multimedia Appendix 2). The characteristics of the study population are listed in Table 1. Patients in the training data set were older than those in the external validation data set (59.8 versus 57.1 years). The mean baseline eGFR was higher in the training data set than in the external validation data set (94.9 versus 89.4 mL/min/1.73 m^2). Although more patients had hypertension or diabetes in the training data set, the mean Charlson Comorbidity Index was higher in the external validation data set. There was no difference in baseline characteristics between the training and internal validation data sets. During the 2-week period after admission, the cumulative incidence of AKI (any stage) was 5.91% in the training data set and 3.63% in the external validation data set (Multimedia Appendix 3). The cumulative incidence of severe AKI (stages 2 or 3) was 1.58% and 1.11% in the training and external validation data sets, respectively.

Table 1. Baseline characteristics of the training and validation data sets.

Variables	SNUBH[a] training set (n=69,081)	SNUBH validation set (n=7675)	SNUH[b] validation set (n=72,352)	*P* value[c]
Age (years), mean (SD)	59.8 (16.5)	59.6 (16.6)	57.1 (16.0)	<.001
Male sex, n (%)	36,732 (53.2)	4114 (53.6)	37,405 (51.7)	<.001
Body mass index (kg/m^2), mean (SD)	23.8 (3.5)	23.8 (3.5)	23.2 (3.5)	<.001
Stay length (days), mean (SD)	8.7 (14.2)	8.5 (10.4)	6.9 (11.7)	<.001
ICU[d] admission, n (%)	4478 (6.5)	475 (6.2)	2340 (3.2)	<.001
Charlson Comorbidity Index, mean (SD)	1.0 (1.4)	1.1 (1.4)	1.1 (1.4)	<.001
Specific preexisting comorbidities				
Hypertension, n (%)	9455 (13.7)	995 (13.0)	6624 (9.2)	<.001
Diabetes mellitus, n (%)	7187 (10.4)	769 (10.0)	6513 (9.0)	<.001
Ischemic heart disease, n (%)	7452 (10.8)	790 (10.3)	6289 (8.7)	<.001
Heart failure, n (%)	1500 (2.2)	169 (2.2)	793 (1.1)	<.001
Baseline eGFR[e] (mL/min/1.73 m^2), mean (SD)	94.9 (38.9)	94.6 (37.4)	89.4 (21.9)	<.001
No CKD[f] or CKD 1 or 2 (≥60 mL/min/1.73 m^2), n (%)	60,201 (87.1)	6683 (87.1)	65,310 (90.3)	<.001
CKD G3a (45-59 mL/min/1.73 m^2), n (%)	4839 (7.0)	550 (7.2)	4188 (5.8)	
CKD G3b (30-44 mL/min/1.73 m^2), n (%)	2629 (3.8)	279 (3.6)	1940 (2.7)	
CKD G4 (15-29 mL/min/1.73 m^2), n (%)	1412 (2.0)	163 (2.1)	914 (1.3)	
Hemoglobin (g/dL), mean (SD)	13.0 (2.1)	13.0 (2.1)	13.0 (2.0)	.54
Albumin (g/dL), mean (SD)	4.0 (0.6)	4.0 (0.6)	4.1 (0.5)	<.001
Bilirubin (mg/dL), mean (SD)	0.9 (1.5)	0.9 (1.6)	0.9 (1.5)	.03
Calcium (mg/dL), mean (SD)	8.8 (0.6)	8.8 (0.6)	9.1 (0.6)	<.001
Glucose (mg/dL), mean (SD)	126.4 (55.5)	125.8 (54.1)	121.1 (48.9)	<.001
Sodium (mEq/L), mean (SD)	138.9 (3.8)	138.8 (3.8)	139.8 (3.3)	<.001
Potassium (mEq/L), mean (SD)	4.1 (0.4)	4.2 (0.5)	4.2 (0.4)	<.001
Chloride (mEq/L), mean (SD)	102.5 (4.0)	102.5 (4.0)	103.5 (3.5)	<.001
AST[g] (IU/L), mean (SD)	47.3 (245.3)	46.7 (213.2)	32.4 (108.7)	<.001
ALT[h] (IU/L), mean (SD)	42.6 (192.1)	42.0 (154.1)	30.5 (100.4)	<.001
Total CO_2 (mEq/L), mean (SD)	24.7 (3.3)	24.7 (3.3)	25.6 (3.6)	<.001
Platelet (10^3/μL), mean (SD)	232.9 (85.0)	232.3 (86.7)	235.8 (479.1)	<.001
White blood cell count (cells/mm^2), mean (SD)	8.2 (7.1)	8.3 (8.3)	7.3 (5.5)	.10
Medication use				
RAS[i] blockers, n (%)	5494 (8.0)	610 (7.9)	6039 (8.3)	.007
Diuretics, n (%)	3437 (5.0)	381 (5.0)	3723 (5.1)	.14
NSAIDs[j], n (%)	16,974 (24.6)	1874 (24.4)	13,290 (18.4)	<.001

Variables	SNUBH[a] training set (n=69,081)	SNUBH validation set (n=7675)	SNUH[b] validation set (n=72,352)	*P* value[c]
Systolic blood pressure (mmHg), mean (SD)	128.7 (16.5)	128.6 (16.5)	124.8 (15.8)	<.001
Diastolic blood pressure (mmHg), mean (SD)	74.1 (10.8)	74.0 (10.7)	76.2 (10.7)	<.001
Heart rate (beats/minute), mean (SD)	78.6 (14.4)	78.7 (14.3)	77.7 (14.2)	<.001
Body temperature (°C), mean (SD)	36.7 (0.5)	36.7 (0.5)	36.5 (0.4)	<.001

[a]SNUBH: Seoul National University Bundang Hospital.

[b]SNUH: Seoul National University Hospital.

[c]*P* value between the SNUBH training set and the SNUH validation set (variables are not statistically different between the SNUBH training set and the SNUBH validation set).

[d]ICU: intensive care unit.

[e]eGFR: estimated glomerular filtration rate.

[f]CKD: chronic kidney disease.

[g]AST: aspartate aminotransferase.

[h]ALT: alanine aminotransferase.

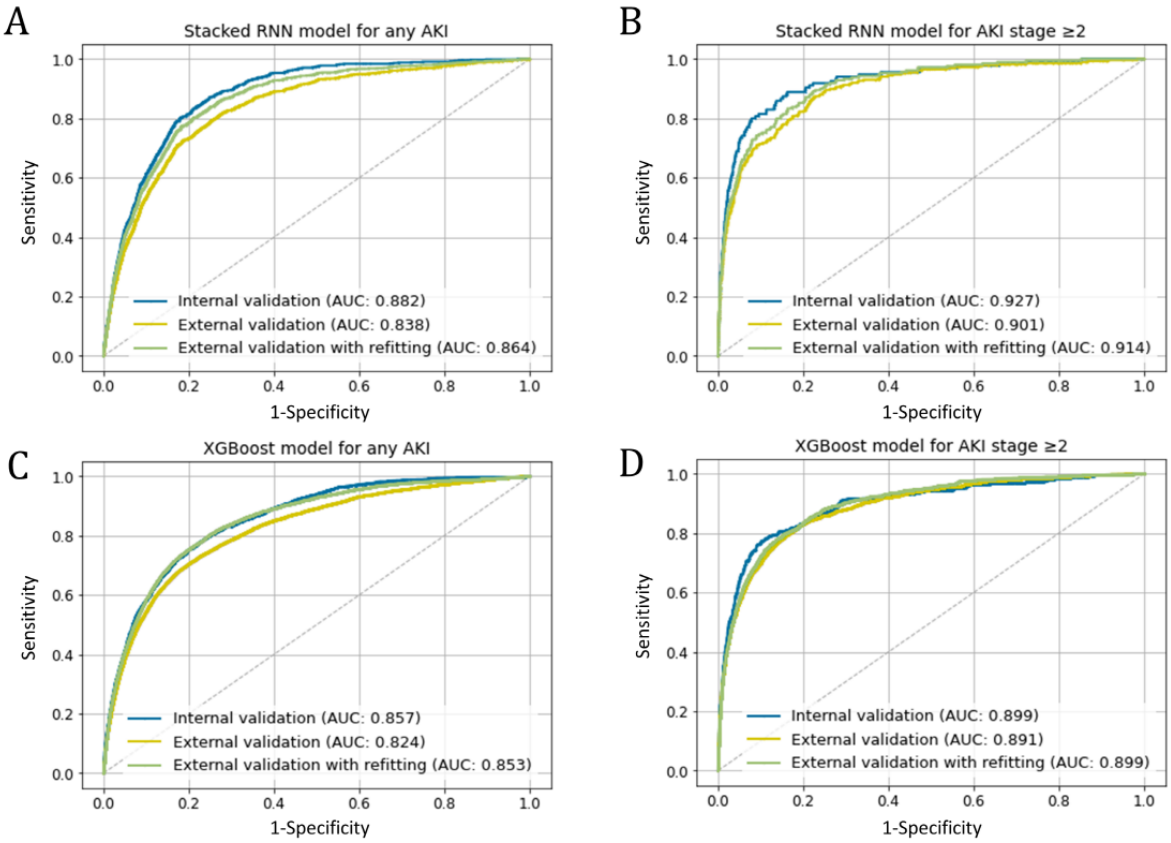
[i]RAS: renin–angiotensin system.

[j]NSAIDs: nonsteroidal anti-inflammatory drugs.

Prediction of AKI Development

First, we assessed the performance of model 1 to predict AKI development (any stage and stage 2 or higher) within the next 7 days based on different algorithms. Overall, the AUC was higher in the stacked RNN model than in the XGBoost model for any AKI and stage 2 or higher AKI. The AUC of the RNN model was 0.88/0.84 (internal/external validation) for any AKI and 0.93/0.90 (internal/external validation) for stage 2 or higher AKI, while the AUC of the XGBoost model was 0.86/0.82 (internal/external validation) for any AKI and 0.90/0.89 (internal/external validation) for stage 2 or higher AKI (Figure 2). Overall, better performance was found with the larger training sample size, which might be due to the class imbalance of the data set, suggesting that 80%-90% of the split ratio was sufficient in training (Multimedia Appendix 4). The model performance on the external validation set was slightly lower than that on the internal validation set. However, the performance of the updated model was improved, reducing the difference in the AUC between the internal and external validation sets. The RNN model to predict stage 2 or higher AKI revealed the highest performance. The evaluation metrics other than AUC are shown for different probability cutoff values in Multimedia Appendices 5 and 6. We evaluated model 2 based on model 1 for any AKI, considering the AKI incidence rate.

Figure 2. The receiver operating characteristic curve for model 1. (A) The stacked RNN model for any AKI, (B) the stacked RNN model for AKI stage 2 or higher, (C) the XGBoost model for any AKI, and (D) the XGBoost model for AKI stage 2 or higher. AKI: acute kidney injury; RNN: recurrent neural network.



Prediction of Creatinine Trajectory

We constructed and validated the prediction model (model 2) for Cr trajectory separately in the positive and negative prediction groups based on the prediction results of model 1 for any AKI. The MSE values are presented in Table 2, and ranged from 0.03 to 0.06 in the internal validation case, and from 0.06 to 0.09 in the external validation case. The refitted model in the external validation data set showed the improved MSE of 0.03-0.08. Overall, the MSE values at different prediction points (24, 48, and 72 hours) were comparable to each other.

Table 2. Predictive performance of model 2 for the future value of serum creatinine (Cr).

Subgroup and validation method	MSE at different time points		
	24 hours	48 hours	72 hours
Positive prediction			
Internal validation	0.04	0.04	0.06
External validation	0.06	0.06	0.09
External validation (refitting)	0.05	0.06	0.08
Negative prediction			
Internal validation	0.03	0.04	0.04
External validation	0.06	0.06	0.08
External validation (refitting)	0.03	0.05	0.05

Application of Interpretability Techniques on the Developed Models

Application of model 1 allows the identification of patients with high AKI risk. Moreover, the future values of Cr can be predicted using model 2. However, risk assessment and predicted laboratory values are insufficient for the management of patients with high risk. The prediction model itself could not provide measures on how to manage or prevent AKI. First, we examined feature importance obtained by mean SHA*P* values in model 1 (Multimedia Appendix 7). Baseline eGFR and serum Cr were the top 2 features in both RNN and XGBoost models.

The remaining features showed relatively less impact on prediction. To determine the feature importance in the RNN model, the SHA*P* values were also averaged over the sequence length (7 days); thus, many static variables were of high rank. However, the importance ranking of some dynamic features, such as pulse rate, diuretic use, and white blood cell count, was found to increase over time (Multimedia Appendix 8). To illustrate the relationships between the AKI risk and each feature, we have presented PDP and ICE plots in Multimedia Appendix 9, where yellow lines denote the average probability of AKI versus the selected feature values in the studied patients (PDP), and each black line represents an individual's probability (ICE plot). These plots provide global and instance-level model explanations. They primarily offer visual representations of a selected feature's effect on AKI probability. Overall, substantial changes in AKI probability with the level of features were observed, especially for aspartate aminotransferase, platelet count, white blood cell count, and vital signs (eg, blood pressure and pulse rate). The patterns of ICE plots were quite different in individual patients. However, PDP could not fully explain the feature–response relationship in each patient. We did not find a clear effect of some clinically meaningful features such as hemoglobin and albumin on AKI prediction. However, accumulated local effects plots better showed the association between these features and AKI prediction (Figure 3). Therefore, we performed an additional analysis using accumulated local effects plots to evaluate the feature effects at different time gaps from the prediction point (Figure 3 and Multimedia Appendix 10). Vital signs, white blood cell counts, and Cr levels had the greatest effect the day before the prediction point, while albumin and hemoglobin showed the greatest effect 5-6 days before the prediction point. Some dynamic features, such as the use of medication, could be considered as correctable features in actual, clinical settings. In this regard, we estimated the predicted trajectories of serum Cr values according to the use of nephrotoxic drugs. Predicted values of Cr with and without nephrotoxic drug administration are shown in Figure 4. The presented cases represent the past Cr values (green lines), predicted values (blue lines), and predicted values after the discontinuation of nephrotoxic drugs (orange lines). These approaches could help clinicians make decisions regarding the prevention and intervention for future AKI occurrence.

Figure 3. Accumulated local effects plots for the selected variables at the different time gaps from the prediction point. The same y-scale was used for all plots. AST: aspartate aminotransferase; ALT: alanine aminotransferase; BUN: blood urea nitrogen; Hb: hemoglobin; WBC: white blood cell count.

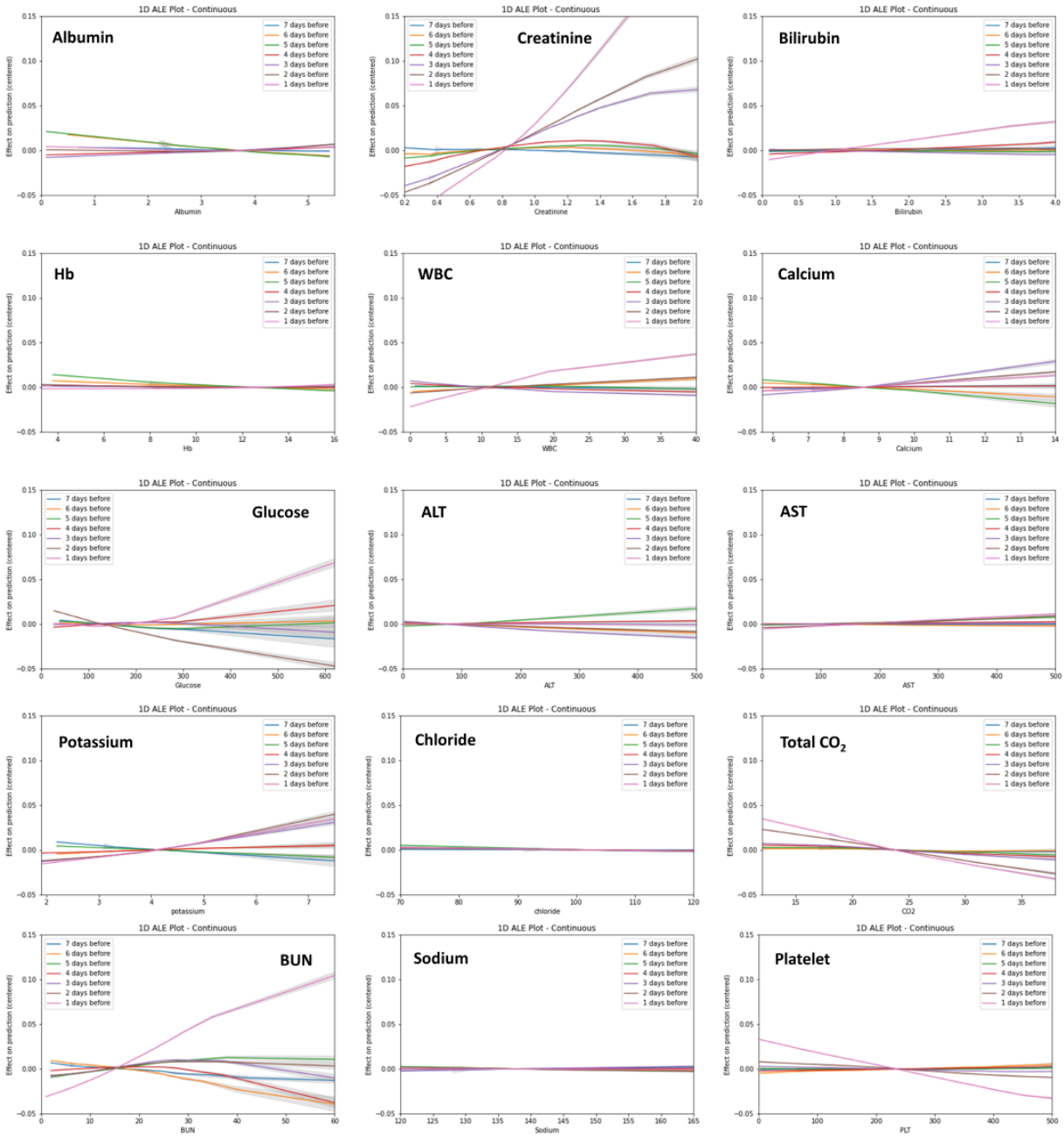
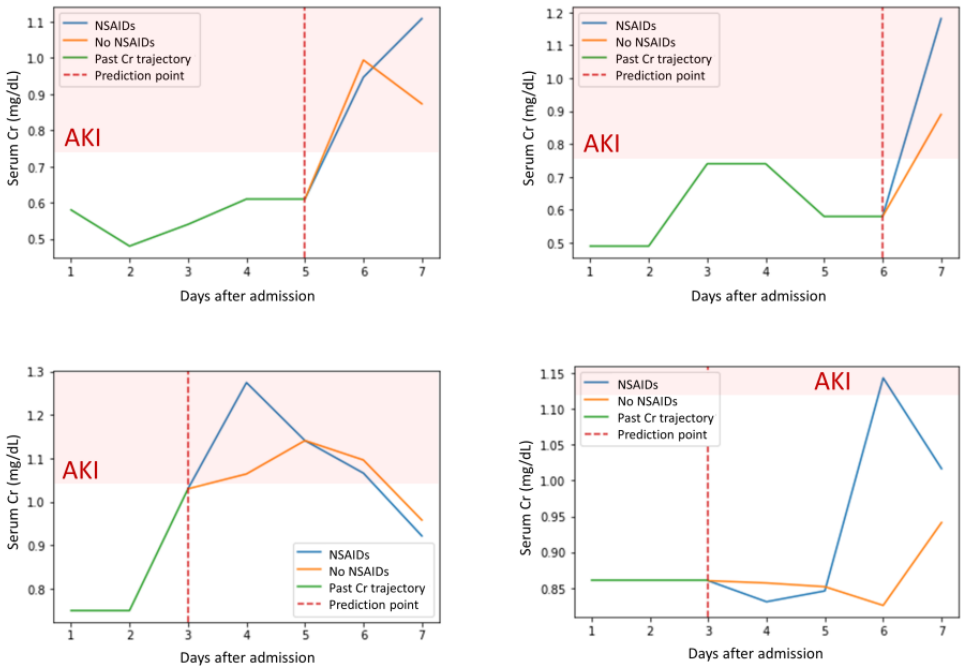


Figure 4. Individual interpretation with individual conditional expectations (green lines: the past Cr values; blue lines: predicted Cr values on certain nephrotoxic drugs; orange lines: predicted Cr value when nephrotoxic drugs are discontinued; and dotted red lines: prediction point). Modified drugs are NSAIDs in these 4 cases. AKI: acute kidney injury; Cr: creatinine; NSAIDs: nonsteroidal anti-inflammatory drugs.



Discussion

Principal Findings

In this study, we developed a continuous prediction model for in-hospital AKI using the RNN algorithm with external validation and demonstrated its applicability for the support of clinical decision making. The developed model was constructed for all general inpatients based on the use of various dynamic and static clinical features, showing relatively good performance. External validation was performed with data from an independent center. Furthermore, we showed examples relevant to the presentation of feature information to help actual clinical decisions at the global and individual patient level using several model-agnostic interpretation methods.

To prevent and manage AKI more effectively, timely diagnosis and intervention are emphasized. Thus, AKI is an area of interest for the application of predictive clinical models. To date, numerous studies have been published on predictive models related to AKI [18]. The identification and appropriate management of patients with high risk could improve patient outcomes and reduce economic burden in health care facilities [35]. Nevertheless, previous studies of the AKI prediction model have mainly focused on specific population groups, such as patients who are critically ill or those who underwent cardiovascular surgery, thus making generalization difficult [36-38]. Moreover, most models have not been externally validated, and there are only few models that are capable of real-time assessments. In a recent report of artificial intelligence research, only 6% of conducted studies performed external validation [39]. Many predictive models used clinical or laboratory data collected at the time of admission or at specific time points, such as during the preoperative period or admission [40-42]. However, the definition of AKI is dynamic and includes the concept of time duration (Cr increase by at least 0.3 mg/dL in 48 hours or 1.5 times over 1 week) as shown in the KDIGO definition [30]. Not only the changes in Cr levels, but also the changed rates or slopes were important in clinical decisions [43]. In this regard, the RNN algorithms could have a considerable strength and could fully utilize the sequential information of feature values. The RNN models were found to be effective in learning sequential data and demonstrated superior model performance over conventional models developed using data at one time point [44-46]. These reflect the trends of clinical variables and are expected to be advantageous for real-time risk assessment.

In this study, we developed 2-stage hierarchical prediction models (model 1 and model 2) using the stacked RNN structure. Model 1 focused on the distinction of patients with high risk of AKI from those without risk. Model 1 using stacked RNN layers predicted any AKI with an AUC of 0.84-0.88 and stage 2 or higher AKI with an AUC of 0.90-0.93. In particular, the RNN model outperformed the XGBoost-based model for both any AKI and stage 2 or higher AKI. Gradient boosting algorithm has been frequently utilized for the model of AKI prediction in recent related studies. The overall performance of the gradient boosting models was reported to have an AUC of 0.67-0.77 for AKI stage 1 or higher and an AUC of 0.84-0.87 for AKI stage

2 or higher in the next 48 hours [47-49]. Model 2 intended to predict the future values of serum Cr that could help physicians understand and interpret the models in a more intuitive manner. We adapted a study design that is updated on a daily basis because most laboratory tests or medications usually do not change more than once a day in the general ward of institutions. Similarly, to minimize the burden of real-time data processing, a daily AKI alert system was implemented into the EHR, which demonstrated its effectiveness for the diagnosis of unlooked AKI [28]. In this study, the overall model performance was somewhat different between the internal and external validation cohorts. It may be attributed to different patient characteristics, such as age and comorbidities, and to the patterns of clinical practice in the 2 centers. Specifically, the incidence of AKI stage 1 was quite different between the 2 centers. In this regard, some authors suggest multiple external validations for a generalized prediction model [50]. The construction of a generalizable prediction model in superpopulations is burdensome in terms of time and cost. Therefore, we additionally fine-tuned our model to overcome the heterogeneity of independent data sets. This approach has been employed in recent related studies [32,49]. Although only a small amount of data was used for refitting, the model performance was greatly improved. The results of the refitted models were comparable to those of the internal validation data set.

However, machine learning–based models do not explain why their decisions are right, and it is uncertain how they relate the results to clinical decision making. To apply the prediction results in clinical practice, clinicians must understand how the risk assessment is derived. Unlike conventional regression models, neural network models process enormous amounts of input data through complex, multiple hidden layers and weight parameters that make it difficult for clinicians to understand and interpret the network structure. In this context, machine learning models are usually referred to as black boxes. Therefore, we presented some model-agnostic interpretation approaches, such as the SHAP feature importance, PDP with ICE plots, and accumulated local effects plots. These methods are practical and helpful in utilizing neural network–based models and could help us identify correctable risk factors at the individual patient level. ICE plots could offer insights into how selected features affect prediction results in the black-box model. As shown in ICE plots, quite different relationships between feature values and prediction outcomes were observed in each patient. This suggests interindividual heterogeneity of feature contributions. PDP does not reflect the heterogeneous effects between individuals and requires an assumption of independence [33]. In this regard, the accumulated local effects plot provides a better interpretation of the association between features and predictions. Accumulated local effects plots are unbiased and superior to the PDP when evaluating feature effects in data sets with correlated features. Nevertheless, the interpretation methods themselves do not indicate causal inferences between features and the model output. Currently, various feature engineering techniques are actively being studied and constitute promising fields of artificial intelligence. Therefore, human-friendly interpretable models that reflect possible causal inferences will be developed in the future.

Limitations

There are some limitations associated with this study. First, we used retrospective data in model training and validations. Thus, our results do not indicate the performance of the model in actual clinical practice. Second, because the model is updated daily, its application is more appropriate for patients in general wards than for those who are critically ill. Nevertheless, we developed an AKI prediction model with a relatively high performance and validated it externally. We presented the model interpretation methods for the RNN model based on sequential data and showed an example of the effective utilization of an AKI prediction model.

Conclusion

Our study demonstrated how to support clinical decisions based on RNN-based prediction models in the clinical setting. Our model can provide real-time assessment of future AKI occurrences and individualized risk factors for AKI in general inpatient cohorts.

Acknowledgments

This work was supported by a grant from the Korean Society of Nephrology (FMC 2019), a grant from the National Research Foundation (No. 2019R1A2C1085411), and a grant from the Seoul National University Bundang Hospital research fund (No. 16-2021-004).

Authors' Contributions

KK wrote the manuscript and performed data analyses; HY performed model development and conducted data analyses; JY, H-ES, J-YR, and YCK participated in data collection and preprocessing; JCJ, HJC, KYN, D-WC, and SSH contributed to data acquisition and interpretation; SK designed and supervised the study. All authors approved the final version of the manuscript to be published.

Conflicts of Interest

None declared.

Multimedia Appendix 1
List of variables used in model development.

[DOCX File , 15 KB - jmir_v23i4e24120_app1.docx]

Multimedia Appendix 2
Flow diagram for study participants.
[DOCX File , 63 KB - jmir_v23i4e24120_app2.docx]

Multimedia Appendix 3
Cumulative incidence of AKI.
[DOCX File , 268 KB - jmir_v23i4e24120_app3.docx]

Multimedia Appendix 4
The AUC of the developed model according to different training sample sizes.
[DOCX File , 111 KB - jmir_v23i4e24120_app4.docx]

Multimedia Appendix 5
Evaluation metrics of model 1 for different probability cutoffs (any stage AKI).
[DOCX File , 21 KB - jmir_v23i4e24120_app5.docx]

Multimedia Appendix 6
Evaluation metrics of model 1 for different probability cutoffs (AKI stage ≥2).
[DOCX File , 21 KB - jmir_v23i4e24120_app6.docx]

Multimedia Appendix 7
SHAP feature importance plot for model 1.
[DOCX File , 323 KB - jmir_v23i4e24120_app7.docx]

Multimedia Appendix 8
SHAP feature importance plots according to different time points.
[DOCX File , 352 KB - jmir_v23i4e24120_app8.docx]

Multimedia Appendix 9
Partial dependence plots and individual conditional expectation plots.
[DOCX File , 1274 KB - jmir_v23i4e24120_app9.docx]

Multimedia Appendix 10
Accumulated local effects plots for the vital sign variables at different time gaps from the prediction point.
[DOCX File , 610 KB - jmir_v23i4e24120_app10.docx]

Multimedia Appendix 11
CONSORT-EHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 532 KB - jmir_v23i4e24120_app11.pdf]

References

1. Chertow GM, Burdick E, Honour M, Bonventre JV, Bates DW. Acute kidney injury, mortality, length of stay, and costs in hospitalized patients. J Am Soc Nephrol 2005 Nov;16(11):3365-3370 [FREE Full text] [doi: 10.1681/ASN.2004090740] [Medline: 16177006]
2. Liangos O, Wald R, O'Bell JW, Price L, Pereira BJ, Jaber BL. Epidemiology and outcomes of acute renal failure in hospitalized patients: a national survey. Clin J Am Soc Nephrol 2006 Jan;1(1):43-51 [FREE Full text] [doi: 10.2215/CJN.00220605] [Medline: 17699189]
3. Wang HE, Muntner P, Chertow GM, Warnock DG. Acute kidney injury and mortality in hospitalized patients. Am J Nephrol 2012;35(4):349-355 [FREE Full text] [doi: 10.1159/000337487] [Medline: 22473149]
4. Coca SG, Peixoto AJ, Garg AX, Krumholz HM, Parikh CR. The prognostic importance of a small acute decrement in kidney function in hospitalized patients: a systematic review and meta-analysis. Am J Kidney Dis 2007 Nov;50(5):712-720. [doi: 10.1053/j.ajkd.2007.07.018] [Medline: 17954284]

5. Stewart J, National Confidential Enquiry into Patient Outcome and Death. Adding insult to injury: A review of the care of patients who died in hospital with a primary diagnosis of acute kidney injury (acute renal failure). London, UK: NCEPOD; 2009.
6. Vanmassenhove J, Kielstein J, Jörres A, Biesen WV. Management of patients at risk of acute kidney injury. Lancet 2017 May 27;389(10084):2139-2151. [doi: 10.1016/S0140-6736(17)31329-6] [Medline: 28561005]
7. Susantitaphong P, Cruz DN, Cerda J, Abulfaraj M, Alqahtani F, Koulouridis I, Acute Kidney Injury Advisory Group of the American Society of Nephrology. World incidence of AKI: a meta-analysis. Clin J Am Soc Nephrol 2013 Sep;8(9):1482-1493 [FREE Full text] [doi: 10.2215/CJN.00710113] [Medline: 23744003]
8. Zeng X, McMahon GM, Brunelli SM, Bates DW, Waikar SS. Incidence, outcomes, and comparisons across definitions of AKI in hospitalized individuals. Clin J Am Soc Nephrol 2014 Jan;9(1):12-20 [FREE Full text] [doi: 10.2215/CJN.02730313] [Medline: 24178971]
9. Lameire NH, Bagga A, Cruz D, De Maeseneer J, Endre Z, Kellum JA, et al. Acute kidney injury: an increasing global concern. Lancet 2013 Jul 13;382(9887):170-179. [doi: 10.1016/S0140-6736(13)60647-9] [Medline: 23727171]
10. Sutherland S, Goldstein S, Bagshaw S. Acute Kidney Injury and Big Data. Contrib Nephrol 2018;193:55-67. [doi: 10.1159/000484963] [Medline: 29393191]
11. Wang F, Preininger A. AI in Health: State of the Art, Challenges, and Future Directions. Yearb Med Inform 2019 Aug;28(1):16-26 [FREE Full text] [doi: 10.1055/s-0039-1677908] [Medline: 31419814]
12. Medic G, Kosaner Kließ M, Atallah L, Weichert J, Panda S, Postma M, et al. Evidence-based Clinical Decision Support Systems for the prediction and detection of three disease states in critical care: A systematic literature review. F1000Res 2019;8:1728 [FREE Full text] [doi: 10.12688/f1000research.20498.2] [Medline: 31824670]
13. Chen J, Chokshi S, Hegde R, Gonzalez J, Iturrate E, Aphinyanaphongs Y, et al. Development, Implementation, and Evaluation of a Personalized Machine Learning Algorithm for Clinical Decision Support: Case Study With Shingles Vaccination. J Med Internet Res 2020 Apr 29;22(4):e16848 [FREE Full text] [doi: 10.2196/16848] [Medline: 32347813]
14. Contreras I, Vehi J. Artificial Intelligence for Diabetes Management and Decision Support: Literature Review. J Med Internet Res 2018 May 30;20(5):e10775 [FREE Full text] [doi: 10.2196/10775] [Medline: 29848472]
15. Chiofolo C, Chbat N, Ghosh E, Eshelman L, Kashani K. Automated Continuous Acute Kidney Injury Prediction and Surveillance: A Random Forest Model. Mayo Clin Proc 2019 May;94(5):783-792. [doi: 10.1016/j.mayocp.2019.02.009] [Medline: 31054606]
16. Tomašev N, Glorot X, Rae JW, Zielinski M, Askham H, Saraiva A, et al. A clinically applicable approach to continuous prediction of future acute kidney injury. Nature 2019 Aug;572(7767):116-119. [doi: 10.1038/s41586-019-1390-1] [Medline: 31367026]
17. Simonov M, Ugwuowo U, Moreira E, Yamamoto Y, Biswas A, Martin M, et al. A simple real-time model for predicting acute kidney injury in hospitalized patients in the US: A descriptive modeling study. PLoS Med 2019 Jul;16(7):e1002861 [FREE Full text] [doi: 10.1371/journal.pmed.1002861] [Medline: 31306408]
18. Hodgson LE, Sarnowski A, Roderick PJ, Dimitrov BD, Venn RM, Forni LG. Systematic review of prognostic prediction models for acute kidney injury (AKI) in general hospital populations. BMJ Open 2017 Sep 27;7(9):e016591 [FREE Full text] [doi: 10.1136/bmjopen-2017-016591] [Medline: 28963291]
19. Wang Y, Fang Y, Teng J, Ding X. Acute Kidney Injury Epidemiology: From Recognition to Intervention. Contrib Nephrol 2016;187:1-8. [doi: 10.1159/000443008] [Medline: 26881802]
20. Meyer A, Zverinski D, Pfahringer B, Kempfert J, Kuehne T, Sündermann SH, et al. Machine learning for real-time prediction of complications in critical care: a retrospective study. Lancet Respir Med 2018 Dec;6(12):905-914. [doi: 10.1016/S2213-2600(18)30300-X] [Medline: 30274956]
21. Karako K, Chen Y, Tang W. On medical application of neural networks trained with various types of data. Biosci Trends 2019 Jan 22;12(6):553-559 [FREE Full text] [doi: 10.5582/bst.2018.01264] [Medline: 30555113]
22. Hewamalage H, Bergmeir C, Bandara K. Recurrent Neural Networks for Time Series Forecasting: Current status and future directions. International Journal of Forecasting 2021 Jan;37(1):388-427 [FREE Full text] [doi: 10.1016/j.ijforecast.2020.06.008]
23. Goldberg Y. Neural network methods for natural language processing. San Rafael, CA, USA: Morgan & Claypool Publishers; 2017:1-287.
24. Wilson FP. Machine Learning to Predict Acute Kidney Injury. Am J Kidney Dis 2020 Jun;75(6):965-967. [doi: 10.1053/j.ajkd.2019.08.010] [Medline: 31677894]
25. Tran N, Sen S, Palmieri T, Lima K, Falwell S, Wajda J, et al. Artificial intelligence and machine learning for predicting acute kidney injury in severely burned patients: A proof of concept. Burns 2019 Sep;45(6):1350-1358 [FREE Full text] [doi: 10.1016/j.burns.2019.03.021] [Medline: 31230801]
26. Castelvecchi D. Can we open the black box of AI? Nature 2016 Oct 06;538(7623):20-23. [doi: 10.1038/538020a] [Medline: 27708329]
27. Wilson FP, Greenberg JH. Acute Kidney Injury in Real Time: Prediction, Alerts, and Clinical Decision Support. Nephron 2018;140(2):116-119 [FREE Full text] [doi: 10.1159/000492064] [Medline: 30071528]

28. Park S, Baek SH, Ahn S, Lee K, Hwang H, Ryu J, et al. Impact of Electronic Acute Kidney Injury (AKI) Alerts With Automated Nephrologist Consultation on Detection and Severity of AKI: A Quality Improvement Study. Am J Kidney Dis 2018 Jan;71(1):9-19. [doi: 10.1053/j.ajkd.2017.06.008] [Medline: 28754457]
29. Sundararajan V, Henderson T, Perry C, Muggivan A, Quan H, Ghali WA. New ICD-10 version of the Charlson comorbidity index predicted in-hospital mortality. J Clin Epidemiol 2004 Dec;57(12):1288-1294. [doi: 10.1016/j.jclinepi.2004.03.012] [Medline: 15617955]
30. KDIGO Work Group. KDIGO clinical practice guideline for acute kidney injury. Kidney Int Suppl 2012 Mar;2:8-12 [FREE Full text] [doi: 10.1038/kisup.2012.7] [Medline: 25018916]
31. Lundberg S, Lee SI. A Unified Approach to Interpreting Model Predictions. 2017 Presented at: 31st Conference on Neural Information Processing Systems; December 4-9 2017; Long Beach, CA, USA p. 4768-4777 URL: https://arxiv.org/pdf/1705.07874.pdf
32. da Cruz HF, Pfahringer B, Martensen T, Schneider F, Meyer A, Böttinger E, et al. Using interpretability approaches to update "black-box" clinical prediction models: an external validation study in nephrology. Artif Intell Med 2021 Jan;111:101982. [doi: 10.1016/j.artmed.2020.101982] [Medline: 33461682]
33. Molnar C. Interpretable machine learning: a guide for making black box models explainable. Victoria, BC: Leanpub; 2020:1-314.
34. Friedman JH. Greedy function approximation: A gradient boosting machine. Ann. Statist 2001 Oct 1;29(5):1189-1232. [doi: 10.1214/aos/1013203451]
35. Kashani K, Rosner MH, Haase M, Lewington AJP, O'Donoghue DJ, Wilson FP, et al. Quality Improvement Goals for Acute Kidney Injury. Clin J Am Soc Nephrol 2019 Jun 07;14(6):941-953. [doi: 10.2215/CJN.01250119] [Medline: 31101671]
36. Wijeysundera DN, Karkouti K, Dupuis J, Rao V, Chan CT, Granton JT, et al. Derivation and validation of a simplified predictive index for renal replacement therapy after cardiac surgery. JAMA 2007 Apr 25;297(16):1801-1809. [doi: 10.1001/jama.297.16.1801] [Medline: 17456822]
37. Malhotra R, Kashani KB, Macedo E, Kim J, Bouchard J, Wynn S, et al. A risk prediction score for acute kidney injury in the intensive care unit. Nephrol Dial Transplant 2017 May 01;32(5):814-822. [doi: 10.1093/ndt/gfx026] [Medline: 28402551]
38. Basu RK, Zappitelli M, Brunner L, Wang Y, Wong HR, Chawla LS, et al. Derivation and validation of the renal angina index to improve the prediction of acute kidney injury in critically ill children. Kidney Int 2014 Mar;85(3):659-667 [FREE Full text] [doi: 10.1038/ki.2013.349] [Medline: 24048379]
39. Kim DW, Jang HY, Kim KW, Shin Y, Park SH. Design Characteristics of Studies Reporting the Performance of Artificial Intelligence Algorithms for Diagnostic Analysis of Medical Images: Results from Recently Published Papers. Korean J Radiol 2019 Mar;20(3):405-410 [FREE Full text] [doi: 10.3348/kjr.2019.0025] [Medline: 30799571]
40. Kate RJ, Perez RM, Mazumdar D, Pasupathy KS, Nilakantan V. Prediction and detection models for acute kidney injury in hospitalized older adults. BMC Med Inform Decis Mak 2016 Mar 29;16(1):39 [FREE Full text] [doi: 10.1186/s12911-016-0277-4] [Medline: 27025458]
41. Mohamadlou H, Lynn-Palevsky A, Barton C, Chettipally U, Shieh L, Calvert J, et al. Prediction of Acute Kidney Injury With a Machine Learning Algorithm Using Electronic Health Record Data. Can J Kidney Health Dis 2018;5:2054358118776326 [FREE Full text] [doi: 10.1177/2054358118776326] [Medline: 30094049]
42. He J, Hu Y, Zhang X, Wu L, Waitman LR, Liu M. Multi-perspective predictive modeling for acute kidney injury in general hospital populations using electronic medical records. JAMIA Open 2019 Apr;2(1):115-122 [FREE Full text] [doi: 10.1093/jamiaopen/ooy043] [Medline: 30976758]
43. Warnock DG. The pressing need for real-time risk assessment of hospital-acquired acute kidney injury. Nephrol Dial Transplant 2017 May 01;32(5):766-770. [doi: 10.1093/ndt/gfw282] [Medline: 27461745]
44. Williams RJ, Zipser D. A Learning Algorithm for Continually Running Fully Recurrent Neural Networks. Neural Computation 1989 Jun;1(2):270-280. [doi: 10.1162/neco.1989.1.2.270]
45. Mikolov T, Kombrink S, Burget L, Cernocký J, Khudanpur S. Extensions of recurrent neural network language model. 2011 Presented at: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP); 22-27 May 2011; Prague, Czech Republic. [doi: 10.1109/icassp.2011.5947611]
46. Choi E, Schuetz A, Stewart WF, Sun J. Using recurrent neural network models for early detection of heart failure onset. J Am Med Inform Assoc 2017 Mar;24(2):361-370 [FREE Full text] [doi: 10.1093/jamia/ocw112] [Medline: 27521897]
47. Koyner JL, Carey KA, Edelson DP, Churpek MM. The Development of a Machine Learning Inpatient Acute Kidney Injury Prediction Model. Crit Care Med 2018 Jul;46(7):1070-1077. [doi: 10.1097/CCM.0000000000003123] [Medline: 29596073]
48. Churpek MM, Carey KA, Edelson DP, Singh T, Astor BC, Gilbert ER, et al. Internal and External Validation of a Machine Learning Risk Score for Acute Kidney Injury. JAMA Netw Open 2020 Aug 03;3(8):e2012892 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.12892] [Medline: 32780123]
49. Song X, Yu ASL, Kellum JA, Waitman LR, Matheny ME, Simpson SQ, et al. Cross-site transportability of an explainable artificial intelligence model for acute kidney injury prediction. Nat Commun 2020 Nov 09;11(1):5668 [FREE Full text] [doi: 10.1038/s41467-020-19551-w] [Medline: 33168827]

50. Debray TPA, Vergouwe Y, Koffijberg H, Nieboer D, Steyerberg EW, Moons KGM. A new framework to enhance the interpretation of external validation studies of clinical prediction models. J Clin Epidemiol 2015 Mar;68(3):279-289 [FREE Full text] [doi: 10.1016/j.jclinepi.2014.06.018] [Medline: 25179855]

Abbreviations

AKI: acute kidney injury
AUC: area under the receiver operating characteristic curve
DBP: diastolic blood pressure
EHR: electronic health record
eGFR: estimated glomerular filtration rate
ICE: individual conditional expectation
ICU: intensive care unit
PDP: partial dependence plots
RAS: renin–angiotensin system
RNN: recurrent neural network
SBP: systolic blood pressure
SHAP: Shapley Additive Explanations
WBC: white blood cell count

Edited by G Eysenbach; submitted 05.09.20; peer-reviewed by KD Yoo, M Liu, X Zhang; comments to author 01.12.20; revised version received 26.02.21; accepted 15.03.21; published 16.04.21.

Please cite as:
Kim K, Yang H, Yi J, Son HE, Ryu JY, Kim YC, Jeong JC, Chin HJ, Na KY, Chae DW, Han SS, Kim S
Real-Time Clinical Decision Support Based on Recurrent Neural Networks for In-Hospital Acute Kidney Injury: External Validation and Model Interpretation
J Med Internet Res 2021;23(4):e24120
URL: https://www.jmir.org/2021/4/e24120
doi:10.2196/24120
PMID:33861200

Original Paper

Forecasting Future Asthma Hospital Encounters of Patients With Asthma in an Academic Health Care System: Predictive Model Development and Secondary Analysis Study

Yao Tong[1], MSc; Amanda I Messinger[2], MD; Adam B Wilcox[1], DPhil; Sean D Mooney[1], DPhil; Giana H Davidson[3,4], MPH, MD; Pradeep Suri[5,6,7], MD, MS; Gang Luo[1], DPhil

[1]Department of Biomedical Informatics and Medical Education, University of Washington, Seattle, WA, United States

[2]The Breathing Institute, Department of Pediatrics, University of Colorado School of Medicine, Children's Hospital Colorado, Aurora, CO, United States

[3]Department of Surgery, University of Washington, Seattle, WA, United States

[4]Department of Health Services, University of Washington, Seattle, WA, United States

[5]Seattle Epidemiologic Research and Information Center & Division of Rehabilitation Care Services, VA Puget Sound Health Care System, Seattle, WA, United States

[6]Clinical Learning, Evidence, and Research (CLEAR) Center, University of Washington, Seattle, WA, United States

[7]Department of Rehabilitation Medicine, University of Washington, Seattle, WA, United States

Corresponding Author:
Gang Luo, DPhil
Department of Biomedical Informatics and Medical Education
University of Washington
UW Medicine South Lake Union, 850 Republican Street, Building C, Box 358047
Seattle, WA, 98109
United States
Phone: 1 206 221 4596
Email: gangluo@cs.wisc.edu

Abstract

Background: Asthma affects a large proportion of the population and leads to many hospital encounters involving both hospitalizations and emergency department visits every year. To lower the number of such encounters, many health care systems and health plans deploy predictive models to prospectively identify patients at high risk and offer them care management services for preventive care. However, the previous models do not have sufficient accuracy for serving this purpose well. Embracing the modeling strategy of examining many candidate features, we built a new machine learning model to forecast future asthma hospital encounters of patients with asthma at Intermountain Healthcare, a nonacademic health care system. This model is more accurate than the previously published models. However, it is unclear how well our modeling strategy generalizes to academic health care systems, whose patient composition differs from that of Intermountain Healthcare.

Objective: This study aims to evaluate the generalizability of our modeling strategy to the University of Washington Medicine (UWM), an academic health care system.

Methods: All adult patients with asthma who visited UWM facilities between 2011 and 2018 served as the patient cohort. We considered 234 candidate features. Through a secondary analysis of 82,888 UWM data instances from 2011 to 2018, we built a machine learning model to forecast asthma hospital encounters of patients with asthma in the subsequent 12 months.

Results: Our UWM model yielded an area under the receiver operating characteristic curve (AUC) of 0.902. When placing the cutoff point for making binary classification at the top 10% (1464/14,644) of patients with asthma with the largest forecasted risk, our UWM model yielded an accuracy of 90.6% (13,268/14,644), a sensitivity of 70.2% (153/218), and a specificity of 90.91% (13,115/14,426).

Conclusions: Our modeling strategy showed excellent generalizability to the UWM, leading to a model with an AUC that is higher than all of the AUCs previously reported in the literature for forecasting asthma hospital encounters. After further optimization, our model could be used to facilitate the efficient and effective allocation of asthma care management resources to improve outcomes.

International Registered Report Identifier (IRRID): RR2-10.2196/resprot.5039

(J Med Internet Res 2021;23(4):e22796) doi:10.2196/22796

KEYWORDS

asthma; forecasting; machine learning; patient care management; risk factors

Introduction

Background

In the United States, 7.7% of people have asthma, which causes 188,968 hospitalizations, 1,776,851 emergency department (ED) visits, and 3441 deaths annually [1]. To reduce asthma hospital encounters covering both hospitalizations and ED visits, many health care systems and health plans deploy predictive models to prospectively find patients at high risk and offer them care management services for preventive care. The University of Washington Medicine (UWM), Intermountain Healthcare, and Kaiser Permanente Northern California [2] are 3 examples of such health care systems. Examples of such health plans include those in 9 of the 12 metropolitan communities [3]. Once a patient is deemed to be at high risk and enrolled in a care management program, a care manager will regularly assess the patient's asthma control, adjust the patient's asthma medications if necessary, and help the patient make appointments for health and related services. Using effective care management, as many as 40% of future hospital encounters by patients with asthma can be avoided [4-7].

Owing to its limited service capacity, a care management program normally enrolls at most 3% of patients with a particular condition [8]. To maximize the benefits of this resource-intensive program, it is crucial for the program to only enroll the patients who are at the highest risk. After all, the deployed predictive model's accuracy (or lack thereof) places an upper bound on the program's effectiveness. Several other research groups have built multiple models for forecasting future asthma hospital encounters of patients with asthma. Every model examined only a few features [2,9-22]. Overlooking some important features in the model degrades model accuracy, making the model miss at least half of the patients who will experience future asthma hospital encounters and incorrectly forecast future asthma hospital encounters for many other patients with asthma. These errors result in impaired patient outcomes and wasted health care spending [23]. In nonmedical fields, people frequently adopt the modeling strategy of examining many candidate features to enhance the accuracy of machine learning models [24-27]. Embracing this modeling strategy for medical data, we built a new machine learning model to forecast future asthma hospital encounters of patients with asthma at Intermountain Healthcare, a nonacademic health care system [23]. Our Intermountain Healthcare model raised the area under the receiver operating characteristic curve (AUC) to 0.859, which is higher than that of every previously published model by 0.049 or more. Although this progress is encouraging, it is unclear how well our modeling strategy generalizes to academic health care systems, which normally care for more complex and sicker patients than nonacademic health care systems [28].

Objective

This study evaluates the generalizability of our modeling strategy to the UWM, an academic health care system. Similar to the Intermountain Healthcare model [23], our UWM model uses clinical and administrative data to forecast future asthma hospital encounters of patients with asthma covering both hospitalizations and ED visits. There are 2 possible values of the categorical dependent variable: whether the patient with asthma will experience asthma hospital encounters in the subsequent 12 months. This paper reports on the development and evaluation of the UWM model.

Our Contributions

This study makes the following 3 innovative contributions:

1. We conducted the first evaluation of the generalizability of our modeling strategy to an academic health care system.
2. We evaluated the predictive power of 71 new features, which were not used in our previous study [23], for forecasting asthma hospital encounters.
3. We evaluated the generalizability of our Intermountain Healthcare model to the UWM and the generalizability of our UWM model to Intermountain Healthcare. To the best of our knowledge, this is the first study to evaluate model generalizability in both directions. Previously, model generalizability was evaluated solely in one direction by assessing the performance of a model built using one site's data on another site's data [17].

Methods

Study Design and Ethics Approval

The institutional review boards of the UWM and Intermountain Healthcare approved this secondary analysis study on clinical and administrative data.

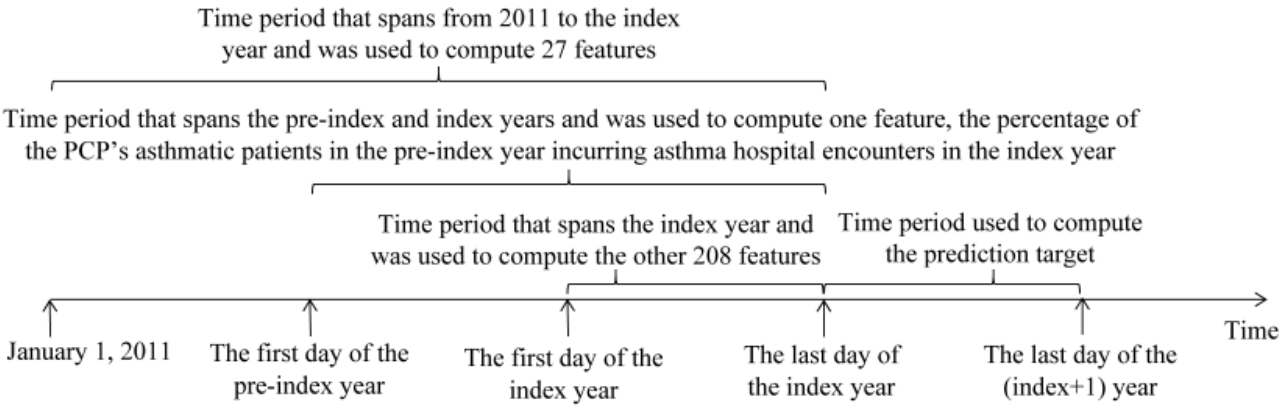
Patient Cohort

The UWM is the largest academic health care system in Washington State. Its enterprise data warehouse contains clinical and administrative data from 3 hospitals and 12 clinics for adults. Our patient cohort covered adult patients with asthma (age ≥18 years) who visited any of these UWM facilities between 2011 and 2018. We defined a patient as having asthma in a specific year if the encounter billing database contained at least one asthma diagnosis code (International Classification of Diseases, Ninth Revision [ICD-9]: 493.0x, 493.1x, 493.8x, 493.9x; International Classification of Diseases, Tenth Revision [ICD-10]: J45.x) record of the patient in that year [10,29,30]. As the sole exclusion criterion, we eliminated patients who passed away in that year.

Prediction Target (Dependent Variable)

The prediction target was from our previous study [23]. We defined an asthma hospital encounter as a hospitalization or an ED visit with asthma as its principal diagnosis (ICD-9: 493.0x, 493.1x, 493.8x, 493.9x; ICD-10: J45.x). As Figure 1 shows, for each patient deemed to have asthma in a specific year, we used any asthma hospital encounter at UWM in the subsequent 12 months, that is, the 12 months after the end of this year, as the outcome of interest. We adopted the patient's data by the end of this year to forecast the patient's outcome in the subsequent 12 months.

Figure 1. The time periods used to compute the features and prediction target for an index year and patient pair. PCP: primary care provider.



Data Set

The UWM enterprise data warehouse supplied a structured data set that contained clinical and administrative data on our patient cohort's encounters at the 3 UWM hospitals and 12 UWM clinics between 2011 and 2019.

Features (Independent Variables)

Similar to our previous study [23], we examined 234 candidate features describing a wide variety of characteristics. Table S1 of Multimedia Appendix 1 describes these features calculated on the structured attributes in our data set, with the 71 new features not used in our previous study [23] marked in italics. Throughout this paper, every mention of the number of a particular kind of items such as medications counts multiplicity whenever the word differing is absent. For instance, consider a patient who was ordered medications twice in a given year. The first time, medications 1 and 2 were ordered for the patient. The second time, medications 2 and 3 were ordered for the patient. Then, the total number of medications ordered for the patient in this year was 4. The total number of differing medications ordered for the patient in this year was 3.

Every input data instance to the predictive model addresses a unique index year and patient pair and is used to forecast the patient's outcome in the subsequent 12 months, that is, the 12 months after the end of the index year. For that pair, we computed the patient's age and primary care provider (PCP) on the last day of the index year. The PCP identified was the patient's last PCP recorded in the electronic medical record system on or before the last day of the index year. As Figure 1 shows, adopting the data in the preindex and index years, we computed 1 feature: the percentage of the PCP's patients with asthma in the preindex year incurring asthma hospital encounters in the index year. Using the data from 2011 to the index year, we computed 25 features: the number of years from the first encounter related to asthma in the data set, the number of years from the first encounter related to chronic obstructive pulmonary disease in the data set, family history of asthma, 15 features related to the problem list, and 7 allergy features. We derived the other 208 features from the data in the index year.

Data Analysis

Data Preparation

Our UWM data set included peak expiratory flow values, which were absent in the Intermountain Healthcare data set adopted in our previous study [23]. Adopting the lower and upper bounds supplied by a clinical expert in our team, we deemed all peak expiratory flow values more than 700 biologically implausible. Adopting the data preparation approach used in our previous paper [23] and this criterion, we pinpointed biologically implausible values, marked them missing, and normalized data. As the outcome of interest came from the subsequent year, our data set included 8 years of effective data (2011-2018) over the 9-year period of 2011-2019. To be consistent with future model use in practice, we used the 2011-2017 data to train the models and the 2018 data to evaluate model performance.

Performance Metrics

As presented in Table 1 and the formulas below, we evaluated model performance using 6 standard metrics: accuracy, AUC, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV).

$$\text{Accuracy} = (TP + TN)/(TP + TN + FP + FN) \quad \textbf{(1)}$$

$$\text{Sensitivity} = TP/(TP + FN) \quad \textbf{(2)}$$

$$\text{Specificity} = TN/(TN + FP) \quad \textbf{(3)}$$

$$PPV = TP/(TP + FP) \quad \textbf{(4)}$$

$$NPV = TN/(TN + FN) \quad \textbf{(5)}$$

Here, TP stands for true positive. TN stands for true negative. FP stands for false positive. FN stands for false negative.

Table 1. The confusion matrix.

Outcome class	Future asthma hospital encounters	No future asthma hospital encounter
Forecasted future asthma hospital encounters	True positive	False positive
Forecasted no future asthma hospital encounter	False negative	True negative

We performed a 1000-fold bootstrap analysis [31] to calculate 95% CIs for the 6 performance measures. For instance, we computed our final UWM model's performance measures for each bootstrap sample of the 2018 data. The 2.5th and 97.5th percentiles of the 1000 values we obtained for every performance metric gave the 95% CI of the corresponding performance measure. We rendered the receiver operating characteristic curve to show the sensitivity-specificity tradeoff.

Classification Algorithms

As in our previous paper [23], our predictive models were built using Waikato Environment for Knowledge Analysis (Weka) Version 3.9 [32]. Weka is a core open-source software package for data mining and machine learning. It integrates a large number of popular feature selection techniques and machine learning algorithms. We checked the extreme gradient boosting (XGBoost) machine learning classification algorithm [33] implemented in the software package XGBoost4J [34] and the 39 native classification algorithms in Weka listed in our previous paper's web-based multimedia appendix [23]. As an efficient and scalable realization of gradient boosting, XGBoost is a form of an ensemble of decision trees. As XGBoost accepts only numerical features, we used one-hot encoding to transform categorical features into numerical features before giving them to XGBoost. We used the 2011-2017 training data and the automatic machine learning model selection method developed in our previous work [35] to automatically select the feature selection technique, classification algorithm, data balancing method for handling imbalanced data, and hyperparameter values among all of the pertinent ones. On average, our method can reduce the model error rate by 11% and search time by 28 times compared with the modern Auto-WEKA automatic machine learning model selection method [35,36].

This study mainly evaluated our modeling strategy's generalizability to the UWM by using the UWM training set to train multiple models and then checking their performance on the UWM test set. In addition, we conducted 2 experiments to evaluate the generalizability of our models across health systems.

Evaluating the Generalizability of Our Intermountain Healthcare Model to the UWM

In the first experiment, we evaluated the generalizability of our Intermountain Healthcare model to the UWM. Previously, we developed both a simplified model and a full model on the Intermountain Healthcare data set [23]. Our simplified Intermountain Healthcare model uses the top 21 features whose importance values calculated by XGBoost on that data set are ≥0.01 [23]. Compared with our full Intermountain Healthcare model using 142 features, our simplified Intermountain Healthcare model retained nearly all of its predictive power. Our UWM data set contained the top 21 features and missed some other features adopted in our full Intermountain Healthcare model. We evaluated the performance of our simplified Intermountain Healthcare model on the UWM test set twice. The first time, we retrained our simplified Intermountain Healthcare model on the UWM training set. The second time, we did not perform retraining and directly applied our original simplified Intermountain Healthcare model trained on the Intermountain Healthcare training set.

Evaluating the Generalizability of Our UWM Model to Intermountain Healthcare

In the second experiment, we evaluated the generalizability of our UWM model to Intermountain Healthcare. We used a simplified UWM model, which used only the top features whose importance values calculated by XGBoost on the UWM training set were ≥0.01. For any top feature that was newly introduced in this study and was not used in our previous study [23], we computed the feature on the Intermountain Healthcare data set. We evaluated our simplified UWM model's performance on the Intermountain Healthcare test set twice. The first time, we retrained our simplified UWM model using the Intermountain Healthcare training set. The second time, we did not perform retraining and directly applied our simplified UWM model trained on the UWM training set.

Results

Demographic and Clinical Characteristics of Our Patient Cohort

Each data instance addresses a unique index year and patient pair. Tables 2 and 3 show the demographic and clinical characteristics of our UWM patient cohort during 2011-2017 and 2018, respectively. The characteristics were similar across the 2 periods. During 2011-2017 and 2018, 1.74% (1184/68,244) and 1.49% (218/14,644) of data instances were linked to asthma hospital encounters in the subsequent 12 months, respectively.

Table 2. Demographic and clinical characteristics of patients with asthma at the University of Washington Medicine during 2011-2017.

Characteristic	Data instances (N=68,244), n (%)	Data instances connecting to asthma hospital encounters in the subsequent 12 months (n=1184), n (%)	Data instances connecting to no asthma hospital encounter in the subsequent 12 months (n=67,060), n (%)
Age (years)			
<40	23,459 (34.38)	466 (39.36)	22,993 (34.29)
40-65	33,889 (49.66)	583 (49.24)	33,306 (49.67)
>65	10,896 (15.97)	135 (11.40)	10,761 (16.05)
Gender			
Male	24,198 (35.46)	551 (46.54)	23,647 (35.26)
Female	44,046 (64.54)	633 (53.46)	43,413 (64.74)
Race			
American Indian or Alaska native	1358 (1.99)	32 (2.70)	1326 (1.98)
Asian	5721 (8.38)	96 (8.11)	5625 (8.39)
Black or African American	8420 (12.34)	520 (43.92)	7900 (11.78)
Native Hawaiian or other Pacific islander	673 (0.99)	14 (1.18)	659 (0.98)
White	47,747 (69.97)	507 (42.82)	47,240 (70.44)
Unknown or not reported	4325 (6.34)	15 (1.27)	4310 (6.43)
Ethnicity			
Hispanic	3526 (5.17)	82 (6.93)	3444 (5.14)
Non-Hispanic	56,309 (82.51)	1062 (89.70)	55,247 (82.38)
Unknown or not reported	8409 (12.32)	40 (3.38)	8369 (12.48)
Insurance			
Private	40,009 (58.63)	424 (35.81)	39,585 (59.03)
Public	28,787 (42.18)	756 (63.85)	28,031 (41.80)
Self-paid or charity	1366 (2.00)	65 (5.49)	1301 (1.94)
Number of years from the first encounter related to asthma in the data set			
≤3	60,873 (89.20)	986 (83.28)	59,887 (89.30)
>3	7371 (10.80)	198 (16.72)	7173 (10.70)
Asthma medication prescription			
Inhaled corticosteroid	28,889 (42.33)	626 (52.88)	28,263 (42.15)
Inhaled corticosteroid and long-acting β-2 agonist combination	22,015 (32.26)	499 (42.15)	21,516 (32.08)
Leukotriene modifier	8171 (11.97)	201 (16.98)	7970 (11.88)
Long-acting β-2 agonist	12,293 (18.01)	374 (31.59)	11,919 (17.77)
Mast cell stabilizer	47 (0.07)	4 (0.34)	43 (0.06)
Short-acting inhaled β-2 agonist	47,808 (70.05)	1010 (85.30)	46,798 (69.79)
Systemic corticosteroid	18,699 (27.40)	614 (51.86)	18,085 (26.97)
Comorbidity			
Allergic rhinitis	11,449 (16.78)	172 (14.53)	11,277 (16.82)
Anxiety or depression	19,885 (29.14)	372 (31.42)	19,513 (29.10)
Bronchopulmonary dysplasia	1 (0)	0 (0)	1 (0)
Chronic obstructive pulmonary disease	3826 (5.61)	133 (11.23)	3693 (5.51)
Cystic fibrosis	61 (0.09)	1 (0.08)	60 (0.09)

Characteristic	Data instances (N=68,244), n (%)	Data instances connecting to asthma hospital encounters in the subsequent 12 months (n=1184), n (%)	Data instances connecting to no asthma hospital encounter in the subsequent 12 months (n=67,060), n (%)
Eczema	3891 (5.70)	66 (5.57)	3825 (5.70)
Gastroesophageal reflux	12,291 (18.01)	238 (20.10)	12,053 (17.97)
Obesity	7845 (11.50)	177 (14.95)	7668 (11.43)
Sinusitis	7261 (10.64)	89 (7.52)	7172 (10.69)
Sleep apnea	4556 (6.68)	88 (7.43)	4468 (6.66)
Smoking status			
Current smoker	14,081 (20.63)	255 (21.54)	13,826 (20.62)
Former smoker	15,530 (22.76)	221 (18.67)	15,309 (22.83)
Never smoker or unknown	38,633 (56.61)	708 (59.80)	37,925 (56.55)

Table 3. Demographic and clinical characteristics of patients with asthma at the University of Washington Medicine in 2018.

Characteristic	Data instances (N=14,644), n (%)	Data instances connecting to asthma hospital encounters in the subsequent 12 months (n=218), n (%)	Data instances connecting to no asthma hospital encounter in the subsequent 12 months (n=14,426), n (%)
Age (years)			
<40	4823 (32.9)	77 (35.3)	4746 (32.9)
40-65	6794 (46.4)	111 (50.9)	6683 (46.3)
>65	3027 (20.7)	30 (13.8)	2997 (20.8)
Gender			
Male	5238 (35.8)	100 (45.9)	5138 (35.6)
Female	9406 (64.2)	118 (54.2)	9288 (64.4)
Race			
American Indian or Alaska native	281 (1.9)	8 (3.7)	273 (1.9)
Asian	1325 (9.1)	18 (8.7)	1307 (9.1)
Black or African American	1570 (10.7)	79 (36.2)	1491 (10.3)
Native Hawaiian or other Pacific islander	131 (0.9)	2 (0.9)	129 (0.9)
White	10,213 (69.7)	110 (50.5)	10,103 (70)
Unknown or not reported	1124 (7.7)	1 (0.5)	1123 (7.8)
Ethnicity			
Hispanic	850 (5.8)	20 (9.2)	830 (5.7)
Non-Hispanic	12,566 (85.8)	196 (89.9)	12,370 (85.7)
Unknown or not reported	1228 (8.4)	2 (0.9)	1226 (8.5)
Insurance			
Private	10,800 (73.7)	108 (49.5)	10,692 (74.1)
Public	8023 (54.8)	182 (83.5)	7841 (54.3)
Self-paid or charity	484 (3.3)	25 (11.5)	459 (3.2)
Number of years from the first encounter related to asthma in the data set			
≤3	10,566 (72.1)	124 (56.9)	10,442 (72.4)
>3	4078 (27.8)	94 (43.1)	3984 (27.6)
Asthma medication prescription			
Inhaled corticosteroid	6177 (42.2)	108 (49.5)	6069 (42.1)
Inhaled corticosteroid and long-acting β-2 agonist combination	4508 (30.8)	83 (38.1)	4425 (30.7)
Leukotriene modifier	2176 (14.9)	46 (21.1)	2130 (14.77)
Long-acting β-2 agonist	2518 (17.2)	62 (28.4)	2456 (17.02)
Mast cell stabilizer	14 (0.1)	1 (0.5)	13 (0.09)
Short-acting inhaled β-2 agonist	9704 (66.3)	164 (75.2)	9540 (66.1)
Systemic corticosteroid	4163 (28.4)	120 (55.1)	4043 (28)
Comorbidity			
Allergic rhinitis	2095 (14.3)	26 (11.9)	2069 (14.3)
Anxiety or depression	4346 (29.7)	62 (28.4)	4284 (29.7)
Bronchopulmonary dysplasia	4 (0)	0 (0)	4 (0)
Chronic obstructive pulmonary disease	932 (6.4)	30 (13.8)	902 (6.2)
Cystic fibrosis	17 (0.1)	0 (0)	17 (0.1)

Characteristic	Data instances (N=14,644), n (%)	Data instances connecting to asthma hospital encounters in the subsequent 12 months (n=218), n (%)	Data instances connecting to no asthma hospital encounter in the subsequent 12 months (n=14,426), n (%)
Eczema	743 (5.1)	11 (5.1)	732 (5.1)
Gastroesophageal reflux	2657 (18.1)	46 (21.1)	2611 (18.1)
Obesity	1604 (10.9)	25 (11.5)	1579 (10.9)
Sinusitis	1372 (9.4)	15 (6.9)	1357 (9.4)
Sleep apnea	1499 (10.2)	24 (11.0)	1475 (10.2)
Smoking status			
Current smoker	3242 (22.1)	49 (22.5)	3193 (22.1)
Former smoker	3494 (23.9)	41 (18.8)	3453 (23.9)
Never smoker or unknown	7908 (54.0)	128 (58.7)	7780 (53.9)

As the Chi-square 2-sample test showed, for both the 2011-2017 and 2018 data, the data instances connecting to future asthma hospital encounters and those connecting to no future asthma hospital encounter exhibited the same distribution for anxiety or depression occurrence (P=.74 for the 2018 data and P=.09 for the 2011-2017 data), bronchopulmonary dysplasia occurrence (P=.99), cystic fibrosis occurrence (P=.99), eczema occurrence (P=.99 for the 2018 data and P=.90 for the 2011-2017 data), gastroesophageal reflux occurrence (P=.29 for the 2018 data and P=.06 for the 2011-2017 data), and sleep apnea occurrence (P=.79 for the 2018 data and P=.32 for the 2011-2017 data). These 2 sets of data instances exhibited differing distributions for gender (P=.002 for the 2018 data and P<.001 for the 2011-2017 data), ethnicity (P<.001), insurance category (P<.001), race (P<.001), systemic corticosteroid prescription (P<.001), inhaled corticosteroid prescription (P=.02 for the 2018 data and P<.001 for the 2011-2017 data), inhaled corticosteroid and long-acting β-2 agonist combination prescription (P=.02 for the 2018 data and P<.001 for the 2011-2017 data), short-acting inhaled β-2 agonist prescription (P=.006 for the 2018 data and P<.001 for the 2011-2017 data), long-acting β-2 agonist prescription (P<.001), leukotriene modifier prescription (P=.01 for the 2018 data and P<.001 for the 2011-2017 data), and chronic obstructive pulmonary disease occurrence (P<.001). For the 2011-2017 data, these 2 sets of data instances exhibited differing distributions for mast cell stabilizer prescription (P=.003), obesity occurrence (P<.001), sinusitis occurrence (P<.001), allergic rhinitis occurrence (P=.04), and smoking status (P=.003). For the 2018 data, these 2 sets of data instances exhibited the same distribution for mast cell stabilizer prescription (P=.52), obesity occurrence (P=.89), sinusitis occurrence (P=.25), allergic rhinitis occurrence (P=.36), and smoking status (P=.19).

As the Cochran-Armitage trend test [37] showed, the data instances connecting to future asthma hospital encounters and those connecting to no future asthma hospital encounter exhibited the same distribution for age (P=.06) in the 2018 data and differing distributions for age (P<.001) in the 2011-2017 data. With regard to the 2018 and 2011-2017 data, these 2 sets of data instances exhibited differing distributions for the number of years from the first encounter related to asthma in the data set (P<.001).

Table 4 shows the number of patients with asthma and their number of visits in each year between 2011 and 2018.

Table 4. The number of patients with asthma and their number of visits in each year between 2011 and 2018.

Year	Number of patients with asthma	Number of visits by patients with asthma
2011	6852	32,910
2012	7768	40,730
2013	7754	39,385
2014	9785	58,953
2015	10,587	69,285
2016	12,072	78,605
2017	13,426	87,403
2018	14,644	94,875

Classification Algorithm and Features Adopted by Our Final UWM Model

Our automatic machine learning model selection method [35] selected the XGBoost classification algorithm [33]. XGBoost is a form of an ensemble of decision trees that can naturally deal with missing feature values. As described in Hastie et al [38] in detail, XGBoost automatically calculates the importance value of each feature based on its apportioned contribution to the model. Our final UWM model was formed using XGBoost

and 71 features displayed in descending order of their importance values in Table S2 of Multimedia Appendix 1. XGBoost automatically removed the other features because they had no additional predictive power.

Performance Measures Yielded by Our Final UWM Model

On the UWM test set, our final model yielded an AUC of 0.902 (95% CI 0.879-0.924). Figure 2 shows the receiver operating characteristic curve of the model. Table 5 lists the model's performance measures when the cutoff point for making binary classification was placed at different top percentages of patients with asthma with the largest forecasted risk. When the cutoff point was placed at the top 10% (1464/14,644), the model yielded an accuracy of 90.6% (13,268/14,644; 95% CI 90.13-91.06), a sensitivity of 70.2% (153/218; 95% CI 63.8-76.0), a specificity of 90.91% (13,115/14,426; 95% CI 90.45-91.38), a PPV of 10.45% (153/1464; 95% CI 8.90-11.97), and an NPV of 99.51% (13,115/13,180; 95% CI 99.39-99.62). Table 6 presents the confusion matrix of the model in this case.

Several features, such as a family history of asthma, were calculated on 2 or more years of data. When we dropped these features and checked solely those features calculated on 1 year of data, the AUC of the model decreased from 0.902 to 0.899. If we used only the top 17 features in Table S2 of Multimedia Appendix 1 whose importance values are ≥0.01 and ignored the other 217 features, the model's AUC decreased from 0.902 to 0.898 (95% CI 0.874-0.919). In this case, when we placed the cutoff point for making binary classification at the top 10% (1464/14,644) of patients with asthma with the largest forecasted risk, the model's accuracy decreased from 90.6% (13,268/14,644) to 90.59% (13,266/14,644; 95% CI 90.11-91.06), sensitivity decreased from 70.2% (153/218) to 69.7% (152/218; 95% CI 63.6-75.5), specificity remained at 90.91% (13,114/14,426; 95% CI 90.42-91.37), PPV decreased from 10.45% (153/1464) to 10.38% (152/1464; 95% CI 8.82-11.97), and NPV decreased from 99.51% (13,115/13,180) to 99.5% (13,114/13,180; 95% CI 99.38-99.61).

Figure 2. The receiver operating characteristic curve of our final University of Washington Medicine model.

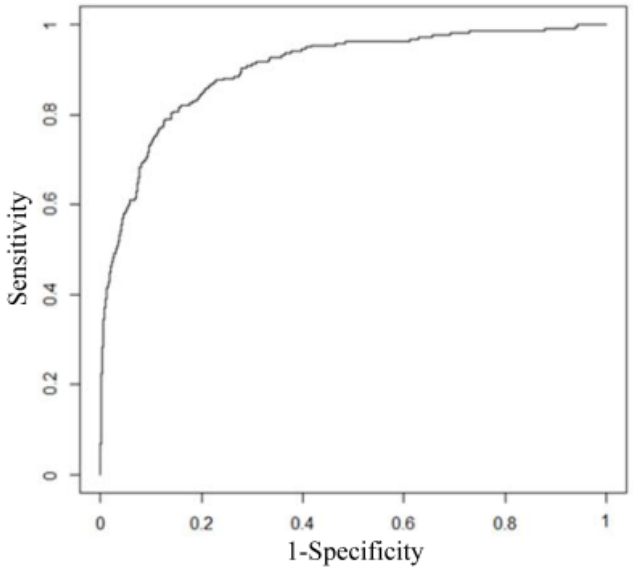


Table 5. Our final UWM model's performance measures when the cutoff point for making binary classification was placed at different top percentages of patients with asthma with the largest forecasted risk.

Top percentage of patients with asthma with the largest forecasted risk (%)	Accuracy (N=14,644), n (%)	Sensitivity (N=218), n (%)	Specificity (N=14,426), n (%)	PPV[a]		NPV[b]	
				n (%)	N	n (%)	N
1	14,410 (98.4)	65 (29.8)	14,345 (99.4)	65 (44.5)	146	14,345 (98.9)	14,498
2	14,316 (97.8)	91 (41.7)	14,225 (98.6)	91 (31.2)	292	14,225 (99.1)	14,352
3	14,193 (96.9)	103 (47.3)	14,090 (97.7)	103 (23.5)	439	14,090 (99.2)	14,205
4	14,061 (96)	110 (50.5)	13,951 (96.7)	110 (18.8)	585	13,951 (99.2)	14,059
5	13,936 (95.2)	121 (55.5)	13,815 (95.8)	121 (16.5)	732	13,815 (99.3)	13,912
6	13,806 (94.3)	129 (59.2)	13,677 (94.8)	129 (14.7)	878	13,677 (99.3)	13,766
7	13,667 (93.3)	133 (61)	13,534 (93.8)	133 (13)	1025	13,534 (99.4)	13,619
8	13,529 (92.4)	137 (62.8)	13,392 (92.8)	137 (11.7)	1171	13,392 (99.4)	13,473
9	13,411 (91.6)	151 (69.3)	13,260 (91.9)	151 (11.5)	1317	13,260 (99.5)	13,327
10	13,268 (90.6)	153 (70.2)	13,115 (90.9)	153 (10.5)	1464	13,115 (99.5)	13,180
15	12,576 (85.9)	173 (79.4)	12,403 (86)	173 (7.9)	2196	12,403 (99.6)	12,448
20	11,860 (81)	181 (83)	11,679 (81)	181 (6.2)	2928	11,679 (99.7)	11,716
25	11,147 (76.1)	191 (87.6)	10,956 (75.9)	191 (5.2)	3661	10,956 (99.7)	10,983

[a]PPV: positive predictive value.

[b]NPV: negative predictive value.

Table 6. The confusion matrix of our final University of Washington Medicine model when the cutoff point for making binary classification was placed at the top 10% (1464/14,644) of patients with asthma with the largest forecasted risk.

Outcome class	Future asthma hospital encounter, n	No future asthma hospital encounter, n
Forecasted future asthma hospital encounters	153	1311
Forecasted no future asthma hospital encounter	65	13,115

Performance Measures Yielded by Our Simplified Intermountain Healthcare Model on UWM Data

For our original simplified Intermountain Healthcare model trained on the Intermountain Healthcare training set [23], when we did not retrain the model and applied the model directly to the UWM test set, the model yielded an AUC of 0.861 (95% CI 0.835-0.885). When we placed the cutoff point for making binary classification at the top 10% (1464/14,644) of patients with asthma with the largest forecasted risk, the model yielded an accuracy of 90.29% (13,222/14,644; 95% CI 89.81-90.77), a sensitivity of 59.6% (130/218; 95% CI 53.4-65.7), a specificity of 90.75% (13,092/14,426; 95% CI 90.28-91.20), a PPV of 8.88% (130/1464; 95% CI 7.46-10.34), and an NPV of 99.33% (13,092/13,180; 95% CI 99.20-99.46).

After we used the UWM training set to retrain our simplified Intermountain Healthcare model [23], the retrained model yielded on the UWM test set an AUC of 0.874 (95% CI 0.848-0.896). When we placed the cutoff point for making binary classification at the top 10% (1464/14,644) of patients with asthma with the largest forecasted risk, the model yielded an accuracy of 90.34% (13,230/14,644; 95% CI 89.85-90.80), a sensitivity of 61.5% (134/218; 95% CI 54.6-67.7), a specificity of 90.78% (13,096/14,426; 95% CI 90.32-91.23), a PPV of 9.15% (134/1464; 95% CI 7.62-10.66), and an NPV of 99.36% (13,096/13,180; 95% CI 99.22-99.49).

Performance Measures Yielded by Our Simplified UWM Model on Intermountain Healthcare Data

Our simplified UWM model used only the top 17 features with importance values of ≥0.01. For our simplified UWM model trained on the UWM training set, when we did not retrain the model and applied the model directly to the Intermountain Healthcare test set, the model yielded an AUC of 0.814 (95% CI 0.798-0.830). When we placed the cutoff point for making binary classification at the top 10% (1926/19,256) of patients with asthma with the largest forecasted risk, the model yielded an accuracy of 89.76% (17,285/19,256; 95% CI 89.32-90.18), a sensitivity of 47.2% (383/812; 95% CI 43.8-50.6), a specificity of 91.64% (16,902/18,444; 95% CI 91.24-92.03), a PPV of 19.90% (383/1925; 95% CI 18.16-21.60), and an NPV of 97.52% (16,902/17,331; 95% CI 97.28-97.75).

After we used the Intermountain Healthcare training set to retrain our simplified UWM model, the retrained model yielded on the Intermountain Healthcare test set an AUC of 0.846 (95% CI 0.831-0.859). When we placed the cutoff point for making

binary classification at the top 10% (1926/19,256) of patients with asthma with the largest forecasted risk, the model yielded an accuracy of 90.11% (17,351/19,256; 95% CI 89.64-90.56), a sensitivity of 51.2% (416/812; 95% CI 47.6-54.5), a specificity of 91.82% (16,935/18,444; 95% CI 91.43-92.21), a PPV of 21.62% (416/1,925; 95% CI 19.81-23.41), and an NPV of 97.72% (16,935/17,331; 95% CI 97.48-97.93).

Discussion

Principal Findings

We built a model on UWM data to forecast asthma hospital encounters of patients with asthma in the subsequent 12 months. Table 7 reveals that our final UWM model yielded an AUC that was higher than the previously reported AUC of every existing model [2,9-23], that is, our modeling strategy of examining many candidate features to enhance model accuracy showed excellent generalizability to the UWM. After further optimization to boost its accuracy and automatically provide explanations of its predictions [39,40] to allow clinical interpretability, our UWM model could be used to facilitate efficient and effective allocation of asthma care management resources to improve outcomes.

In Table S2 of Multimedia Appendix 1, both the 5 most important features and multiple other features within the top 17 indicate a loss of asthma control. It is important to note that the loss of asthma control could be partly because of factors not well captured in our data, such as socioeconomic circumstances, variable management practices among providers, access to subspecialty clinicians, and nonadherence to medications and treatments. Variable asthma severity across patients over time also influences this process.

We checked 234 candidate features. Our final UWM model used 30.3% (71/234) of them. Despite being correlated with the outcome, many unused features had no extra predictive power on the UWM data set over the features adopted in our final UWM model.

For our original simplified Intermountain Healthcare model trained on the Intermountain Healthcare training set [23], when we did not retrain the model on the UWM data and directly applied the model, the model yielded an AUC of 0.861 on the UWM test set. This AUC is 0.041 lower than our final UWM model's AUC, but is still larger than the previously reported AUC of every existing model for forecasting future hospitalizations and ED visits of patients with asthma (Table 7). Therefore, our simplified Intermountain Healthcare model showed excellent generalizability to the UWM.

Compared with our full UWM model using 71 features, our simplified UWM model retained nearly all of its predictive power. For our simplified UWM model trained on the UWM training set, when we did not retrain the model on the Intermountain Healthcare data and directly applied the model, the model yielded an AUC of 0.814 on the Intermountain Healthcare test set. This AUC is 0.045 lower than our full Intermountain Healthcare model's AUC but is still larger than the previously reported AUC of every existing model developed by others for forecasting future hospitalizations and ED visits of patients with asthma (Table 7). Therefore, our simplified UWM model shows excellent generalizability to Intermountain Healthcare.

Table 7. A comparison of our final University of Washington Medicine model and several existing models for forecasting future hospitalizations and emergency department (ED) visits of patients with asthma.

Model	Prediction target	Number of data instances	Number of features the model adopted	Classification algorithm	Sensitivity (%)	Specificity (%)	PPV[a] (%)	NPV[b] (%)	AUC[c]
Our final UWM model	Asthma hospital encounters	82,888	71	XGBoost[d]	70.2	90.91	10.45	99.51	0.902
Our Intermountain Healthcare model [23]	Asthma hospital encounters	334,564	142	XGBoost	53.69	91.93	22.65	97.83	0.859
Loymans et al [9]	Asthma exacerbation	611	7	Logistic regression	—[e]	—	—	—	0.8
Schatz et al [10]	Asthma-induced hospitalization in children	4197	5	Logistic regression	43.9	89.8	5.6	99.1	0.781
Schatz et al [10]	Asthma-induced hospitalization in adults	6904	3	Logistic regression	44.9	87	3.9	99.3	0.712
Eisner et al [11]	Asthma-induced hospitalization	2858	1	Logistic regression	—	—	—	—	0.689
Eisner et al [11]	Asthma-induced ED visit	2415	3	Logistic regression	—	—	—	—	0.751
Sato et al [12]	Severe asthma exacerbation	78	3	Classification and regression tree	—	—	—	—	0.625
Miller et al [14]	Asthma hospital encounters	2821	17	Logistic regression	—	—	—	—	0.81
Yurk et al [16]	Lost day or hospital encounters for asthma	4888	11	Logistic regression	77	63	82	56	0.78
Lieu et al [2]	Asthma-induced hospitalization	16,520	7	Proportional-hazards regression	—	—	—	—	0.79
Lieu et al [2]	Asthma-induced ED visit	16,520	7	Proportional-hazards regression	—	—	—	—	0.69
Lieu et al [18]	Asthma hospital encounters	7141	4	Classification and regression tree	49	83.6	18.5	—	—
Schatz et al [19]	Asthma hospital encounters	14,893	4	Logistic regression	25.4	92	22	93.2	0.614
Forno et al [21]	Severe asthma exacerbation	615	17	Scoring	—	—	—	—	0.75
Xiang et al [22]	Asthma exacerbation	31,433	—	Recurrent neural network	—	—	—	—	0.70

[a]PPV: positive predictive value.

[b]NPV: negative predictive value.

[c]AUC: area under the receiver operating characteristic curve.

[d]XGBoost: extreme gradient boosting.

[e]The initial paper showing the model did not give the performance measure.

Comparison With the Previous Work

Researchers have built multiple models to forecast future hospitalizations and ED visits of patients with asthma [2,9-23]. Table 7 compares our final UWM model with these models, which cover all the relevant models described in the systematic review by Loymans et al [17]. The final UWM model's AUC was 0.902. The AUC of our Intermountain Healthcare model was 0.859. All other existing models have a previously reported AUC 0.81 [2,9-22], which is lower than our final UWM model's AUC by at least 0.091.

It is important to consider the prevalence of the outcome of interest when comparing the performance of different predictive models. Compared with other existing models, the model by Yurk et al [16] achieved a higher sensitivity and PPV mainly because it adopted a different prediction target: asthma hospital encounters or at least 1 day lost for diminished activities or missing work for asthma. This prediction target had a 54%

prevalence rate in patients with asthma and was therefore easier to forecast. If the model by Yurk et al [16] were used to forecast asthma hospital encounters, an outcome that had a <2% prevalence rate in patients with asthma, the model's sensitivity and PPV would likely drop.

The recurrent neural network model by Xiang et al [22] reached a low AUC of 0.7, mainly because it used mostly inpatient data with little outpatient data; adopted only 3 types of attributes: medication, diagnosis, and demographics; and did not merge individual asthma medications into asthma medication categories such as nebulizers and short-acting β-2 agonists, that is, the low AUC does not prove that the recurrent neural network is ineffective at predicting asthma outcomes, but is mainly because of incomplete data and insufficient feature modeling. In comparison, to build our final UWM model, we used both inpatient and outpatient data, adopted many types of attributes, and merged individual asthma medications into asthma medication categories to better capture and model the relationship among different asthma medications.

Excluding the model by Yurk et al [16], every existing published model has a sensitivity 53.69%, which is significantly lower than our final UWM model's sensitivity of 70.2%. For patients with asthma who will have future asthma hospital encounters, sensitivity is the percentage of them identified by the model. The difference in sensitivity can have a significant impact on health care use. Owing to the high prevalence rate of asthma, for every 10% increase in the identified percentage of patients with asthma who would have future asthma hospital encounters, up to 7759 more hospitalizations and 71,074 more ED visits could be avoided in the United States each year with effective care management [1,4-7].

The prevalence rate of targeted poor outcomes greatly impacts the PPV of any predictive model [41]. In our UWM test data set, 1.49% (218/14,644) of patients with asthma had future asthma hospital encounters. When we placed the cutoff point for making binary classification at the top 10% (1464/14,644) of patients with asthma with the largest forecasted risk, an impeccable model in theory would yield the highest possible PPV of 14.89% (218/1464). Our final UWM model yielded a PPV of 10.45% (153/1464), which is 70.18% of the highest possible PPV in theory. In comparison, our Intermountain Healthcare model achieved a PPV of 22.65% [23]. This is 53.69% of the highest possible PPV that an impeccable model in theory would yield on the Intermountain Healthcare test set. The model by Lieu et al [18] yielded a PPV of 18.5% on a data set where 6.9% of patients with asthma had future asthma hospital encounters. The model by Schatz et al [19] yielded a PPV of 22% on a data set where 6.5% of patients with asthma had future asthma hospital encounters. Compared with our case with UWM, both populations have a higher prevalence of asthma hospital encounters, which allows the PPV to be higher. Excluding these PPVs and the PPV by the Yurk et al [16] model, no other existing published model's PPV exceeds 5.6%.

Our final UWM model and our Intermountain Healthcare model [23] have similar top features with importance values of ≥0.01. In both models, many top features are related to previous ED visits and asthma medications. We did not identify several candidate features at the time of constructing our Intermountain Healthcare model. They appeared as top features and affected the ranks and importance values of the other top features in our final UWM model.

Differing models in Table 7 were built using different patient cohorts and used similar but not necessarily identical prediction targets. Some features used in the models built by other researchers, such as certain features computed from patient-reported outcomes and patient surveys, are unavailable in our UWM data set. Therefore, we were unable to show the performance measures that the models built by other researchers would achieve on our UWM data set. However, we are confident that the techniques used in this study improved prediction accuracy. Our final UWM model was built using a state-of-the-art machine learning algorithm, XGBoost. Compared with statistical methods such as logistic regression, machine learning can enhance prediction accuracy with less strict assumptions on data distribution [8,42,43]. Compared with the models built by other researchers, our final UWM model was built using more patients and a more extensive set of candidate features constructed with careful feature engineering, both of which are known to often help improve prediction accuracy [24-27,32]. As partial evidence for this, we built predictive models for asthma hospital encounters using data from 3 health care systems: UWM, Intermountain Healthcare [23], and Kaiser Permanente Southern California [44]. For each of the 3 health care systems, we started model building with approximately 20 candidate features and obtained unsatisfactory accuracy. This motivated us to examine several hundred candidate features. Ultimately, for each of the 3 health care systems, we built a model with an AUC that is higher than all of the AUCs other researchers previously reported in the literature for forecasting asthma hospital encounters [23,44]. This demonstrates the generalizability of our modeling strategy for forecasting asthma hospital encounters.

Considerations Concerning the Potential Clinical Use

Our final UWM model has an AUC that is higher than all of the AUCs previously reported in the literature for forecasting asthma hospital encounters, but still had a seemingly low PPV of 10.4% (153/1464). Nevertheless, this model could be valuable in clinical care. First, health care systems such as UWM, Intermountain Healthcare, and Kaiser Permanente Northern California [2] use proprietary models to allocate asthma care management resources. These models and the models that were formerly built by others have similar performance measures. Our final UWM model has an AUC that is higher than the previously reported AUCs of all these models.

Second, as explained earlier, even an impeccable model in theory would reach a low PPV because the poor outcome of interest has a low prevalence rate in our data set. For such an outcome, sensitivity better reflects the model's potential clinical value than PPV. Our final UWM model had a higher sensitivity than the previously reported sensitivity of every existing model using a comparable prediction target. It is important to note that while asthma hospital encounters have an overall low prevalence rate in the population of patients with asthma, they have

significant financial and clinical impacts at both the population and individual patient levels.

Third, a PPV of 10.45% (153/1464) is useful for identifying high-risk patients with asthma to receive low-cost preventive interventions. The following are 4 examples of such interventions: training the patient to record a diary about environmental triggers, coaching the patient to use an asthma inhaler correctly, coaching the patient to use a peak flow meter correctly and giving it to the patient to self-monitor symptoms at home, and asking a nurse to do extra follow-up phone calls with the patient or the patient's caregiver. These interventions could have a significant impact on patient outcomes.

The final UWM model used 71 features. Reducing the number of features could ease the clinical deployment of our model. To this end, if a minor decrease in prediction accuracy could be tolerated, one could adopt the top few features whose importance values are greater than a given threshold, such as 0.01, and drop the other features. The importance value of a feature varies across health care systems. Ideally, the importance values of the features should first be calculated on a data set from the target health care system before choosing the features to retain.

As is typical with complex machine learning models, an XGBoost model using many features is difficult to interpret. This can limit clinical understandability and adoption, particularly by clinicians who are resistant to using automated tools. In the future, we plan to adopt our previously developed method [39,40] to automatically explain the prediction results of our final UWM model.

The final UWM model was constructed using XGBoost [33]. For binary classification of imbalanced data, XGBoost leverages a hyperparameter, scale_pos_weight, to balance the weights of the 2 outcome classes [45]. To maximize the AUC of our UWM model, our automatic model selection method [35] altered scale_pos_weight to a nondefault value to balance the 2 outcome classes [46]. This incurs a side effect of significantly shrinking the model's forecasted probabilities of having future asthma hospital encounters to values much less than the actual probabilities [46]. This does not preclude us from choosing the top few percent of patients with asthma with the greatest forecasted risk to receive various preventive interventions. To prevent this side effect from occurring, we could retain scale_pos_weight at its default value of 1 without doing any balancing. As a tradeoff, the AUC of the model would decrease from 0.902 to 0.885 (95% CI 0.861-0.907); however, this decreased AUC is still larger than all of the AUCs previously reported in the literature for forecasting asthma hospital encounters.

Limitations

This study has at least 4 limitations that could be interesting topics for future work, as follows:

1. It is possible to further increase the model accuracy by using features other than those checked in this study. For example, features derived from environmental and physiological data gathered by intelligent wearable devices can have this potential.
2. This study used purely structured data and checked only nondeep learning classification algorithms. It is possible to further increase the model accuracy by using deep learning as well as features derived from unstructured clinical notes using natural language processing techniques [40,47].
3. Our UWM data set contained no data on patients' health care use outside of UWM. Therefore, we limited the prediction target to asthma hospital encounters at UWM instead of asthma hospital encounters anywhere. In addition, the features we checked were derived from patients' incomplete administrative and clinical data [48-51]. It would be worth investigating how model accuracy would vary if we have more complete administrative and clinical data of patients [52].
4. This study evaluated the generalizability of our modeling strategy to an academic health care system on a single outcome of a complex chronic disease. We recently showed that our modeling strategy also generalizes well to Kaiser Permanente Southern California for the same predictive modeling problem [44]. We plan to investigate our modeling strategy's generalizability to other diseases, outcomes, and health care systems in the future.

Conclusions

In the first evaluation of its generalizability to an academic health care system, our modeling strategy of examining many candidate features to enhance prediction accuracy showed excellent generalizability to the UWM and led to a model with an AUC that is higher than all of the AUCs previously reported in the literature for forecasting asthma hospital encounters. After further optimization, our UWM model could be used to facilitate the efficient and effective allocation of asthma care management resources to improve outcomes.

Acknowledgments

The authors would like to thank Katy Atwood for helping with the retrieval of the UWM data set and Michael D Johnson for useful discussions. GL, SM, and GD were partially supported by the National Heart, Lung, and Blood Institute of the National Institutes of Health under award R01HL142503. The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. YT did the work at the University of Washington when she was a visiting PhD student.

Authors' Contributions

YT participated in data analysis and the writing of the first draft of the paper. GL conceptualized and designed the study, performed a literature review, participated in data analysis, and rewrote the whole paper. AM, AW, SM, GD, and PS provided feedback on various medical issues, contributed to conceptualizing the presentation, and revised the paper.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The list of candidate features considered in this study.
[PDF File (Adobe PDF File), 105 KB - jmir_v23i4e22796_app1.pdf]

References

1. Most recent national asthma data. Centers for Disease Control and Prevention. 2020. URL: https://www.cdc.gov/asthma/most_recent_national_asthma_data.htm [accessed 2021-03-26]
2. Lieu TA, Quesenberry CP, Sorel ME, Mendoza GR, Leong AB. Computer-based models to identify high-risk children with asthma. Am J Respir Crit Care Med 1998 Apr;157(4 Pt 1):1173-1180. [doi: 10.1164/ajrccm.157.4.9708124] [Medline: 9563736]
3. Mays GP, Claxton G, White J. Managed care rebound? Recent changes in health plans' cost containment strategies. Health Aff (Millwood) 2004;Suppl Web Exclusives:427-436. [doi: 10.1377/hlthaff.w4.427] [Medline: 15451964]
4. Caloyeras JP, Liu H, Exum E, Broderick M, Mattke S. Managing manifest diseases, but not health risks, saved PepsiCo money over seven years. Health Aff (Millwood) 2014 Jan;33(1):124-131. [doi: 10.1377/hlthaff.2013.0625] [Medline: 24395944]
5. Greineder DK, Loane KC, Parks P. A randomized controlled trial of a pediatric asthma outreach program. J Allergy Clin Immunol 1999 Mar;103(3 Pt 1):436-440. [doi: 10.1016/s0091-6749(99)70468-9] [Medline: 10069877]
6. Kelly CS, Morrow AL, Shults J, Nakas N, Strope GL, Adelman RD. Outcomes evaluation of a comprehensive intervention program for asthmatic children enrolled in Medicaid. Pediatrics 2000 May;105(5):1029-1035. [doi: 10.1542/peds.105.5.1029] [Medline: 10790458]
7. Axelrod RC, Zimbro KS, Chetney RR, Sabol J, Ainsworth VJ. A disease management program utilising life coaches for children with asthma. J Clin Outcomes Manag 2001;8(6):38-42.
8. Axelrod RC, Vogel D. Predictive modeling in health plans. Dis Manag Health Outcomes 2003;11(12):779-787. [doi: 10.2165/00115677-200311120-00003]
9. Loymans RJ, Honkoop PJ, Termeer EH, Snoeck-Stroband JB, Assendelft WJ, Schermer TR, et al. Identifying patients at risk for severe exacerbations of asthma: development and external validation of a multivariable prediction model. Thorax 2016 Sep;71(9):838-846. [doi: 10.1136/thoraxjnl-2015-208138] [Medline: 27044486]
10. Schatz M, Cook EF, Joshua A, Petitti D. Risk factors for asthma hospitalizations in a managed care organization: development of a clinical prediction rule. Am J Manag Care 2003 Aug;9(8):538-547 [FREE Full text] [Medline: 12921231]
11. Eisner MD, Yegin A, Trzaskoma B. Severity of asthma score predicts clinical outcomes in patients with moderate to severe persistent asthma. Chest 2012 Jan;141(1):58-65. [doi: 10.1378/chest.11-0020] [Medline: 21885725]
12. Sato R, Tomita K, Sano H, Ichihashi H, Yamagata S, Sano A, et al. The strategy for predicting future exacerbation of asthma using a combination of the Asthma Control Test and lung function test. J Asthma 2009 Sep;46(7):677-682. [doi: 10.1080/02770900902972160] [Medline: 19728204]
13. Osborne ML, Pedula KL, O'Hollaren M, Ettinger KM, Stibolt T, Buist AS, et al. Assessing future need for acute care in adult asthmatics: the Profile of Asthma Risk Study: a prospective health maintenance organization-based study. Chest 2007 Oct;132(4):1151-1161. [doi: 10.1378/chest.05-3084] [Medline: 17573515]
14. Miller MK, Lee JH, Blanc PD, Pasta DJ, Gujrathi S, Barron H, TENOR Study Group. TENOR risk score predicts healthcare in adults with severe or difficult-to-treat asthma. Eur Respir J 2006 Dec;28(6):1145-1155 [FREE Full text] [doi: 10.1183/09031936.06.00145105] [Medline: 16870656]
15. Peters D, Chen C, Markson LE, Allen-Ramey FC, Vollmer WM. Using an asthma control questionnaire and administrative data to predict health-care utilization. Chest 2006 Apr;129(4):918-924. [doi: 10.1378/chest.129.4.918] [Medline: 16608939]
16. Yurk RA, Diette GB, Skinner EA, Dominici F, Clark RD, Steinwachs DM, et al. Predicting patient-reported asthma outcomes for adults in managed care. Am J Manag Care 2004 May;10(5):321-328 [FREE Full text] [Medline: 15152702]
17. Loymans RJ, Debray TP, Honkoop PJ, Termeer EH, Snoeck-Stroband JB, Schermer TR, et al. Exacerbations in adults with asthma: a systematic review and external validation of prediction models. J Allergy Clin Immunol Pract 2018;6(6):1942-1952. [doi: 10.1016/j.jaip.2018.02.004] [Medline: 29454163]
18. Lieu TA, Capra AM, Quesenberry CP, Mendoza GR, Mazar M. Computer-based models to identify high-risk adults with asthma: is the glass half empty of half full? J Asthma 1999 Jun;36(4):359-370. [doi: 10.3109/02770909909068229] [Medline: 10386500]

19. Schatz M, Nakahiro R, Jones CH, Roth RM, Joshua A, Petitti D. Asthma population management: development and validation of a practical 3-level risk stratification scheme. Am J Manag Care 2004 Jan;10(1):25-32 [FREE Full text] [Medline: 14738184]
20. Grana J, Preston S, McDermott PD, Hanchak NA. The use of administrative data to risk-stratify asthmatic patients. Am J Med Qual 1997;12(2):113-119. [doi: 10.1177/0885713X9701200205] [Medline: 9161058]
21. Forno E, Fuhlbrigge A, Soto-Quirós ME, Avila L, Raby BA, Brehm J, et al. Risk factors and predictive clinical scores for asthma exacerbations in childhood. Chest 2010 Nov;138(5):1156-1165 [FREE Full text] [doi: 10.1378/chest.09-2426] [Medline: 20472862]
22. Xiang Y, Ji H, Zhou Y, Li F, Du J, Rasmy L, et al. Asthma exacerbation prediction and risk factor analysis based on a time-sensitive, attentive neural network: retrospective cohort study. J Med Internet Res 2020 Jul 31;22(7):e16981 [FREE Full text] [doi: 10.2196/16981] [Medline: 32735224]
23. Luo G, He S, Stone BL, Nkoy FL, Johnson MD. Developing a model to predict hospital encounters for asthma in asthmatic patients: secondary analysis. JMIR Med Inform 2020 Jan 21;8(1):e16080 [FREE Full text] [doi: 10.2196/16080] [Medline: 31961332]
24. Mayfield J, McNamee P, Piatko CD. Named entity recognition using hundreds of thousands of features. In: Proceedings of the Seventh Conference on Natural Language Learning at HLT-NAACL 2003 - Volume 4. 2003 Presented at: CoNLL'03; May 31-June 1, 2003; Edmonton, Canada p. 184-187. [doi: 10.3115/1119176.1119205]
25. Cao Y, Yu H, Abbott NL, Zavala VM. Machine learning algorithms for liquid crystal-based sensors. ACS Sens 2018 Nov 26;3(11):2237-2245. [doi: 10.1021/acssensors.8b00100] [Medline: 30289249]
26. Zhai Y, Ong Y, Tsang IW. The emerging "Big Dimensionality". IEEE Comput Intell Mag 2014 Aug;9(3):14-26. [doi: 10.1109/mci.2014.2326099]
27. Hansson K, Yella S, Dougherty M, Fleyeh H. Machine learning algorithms in heavy process manufacturing. Am J Intell Syst 2016;6(1):1-13. [doi: 10.5923/j.ajis.20160601.01]
28. Liu LL, Forgione DA, Younis MZ. A comparative analysis of the CVP structure of nonprofit teaching and for-profit non-teaching hospitals. J Health Care Finance 2012;39(1):12-38. [Medline: 23155742]
29. Desai JR, Wu P, Nichols GA, Lieu TA, O'Connor PJ. Diabetes and asthma case identification, validation, and representativeness when using electronic health data to construct registries for comparative effectiveness and epidemiologic research. Med Care 2012 Jul;50 Suppl:30-35 [FREE Full text] [doi: 10.1097/MLR.0b013e318259c011] [Medline: 22692256]
30. Wakefield DB, Cloutier MM. Modifications to HEDIS and CSTE algorithms improve case recognition of pediatric asthma. Pediatr Pulmonol 2006 Oct;41(10):962-971. [doi: 10.1002/ppul.20476] [Medline: 16871628]
31. Steyerberg EW. Clinical Prediction Models: a Practical Approach to Development, Validation, and Updating, 2nd Ed. New York, NY: Springer; 2019.
32. Witten IH, Frank E, Hall MA, Pal CJ. Data Mining: Practical Machine Learning Tools and Techniques, 4th Ed. Burlington, MA: Morgan Kaufmann; 2016.
33. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. In: Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2016 Presented at: KDD '16: 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; August 13-17, 2016; San Francisco, CA p. 785-794. [doi: 10.1145/2939672.2939785]
34. XGBoost JVM Package. 2020. URL: https://xgboost.readthedocs.io/en/latest/jvm/index.html [accessed 2021-03-26]
35. Zeng X, Luo G. Progressive sampling-based Bayesian optimization for efficient and automatic machine learning model selection. Health Inf Sci Syst 2017 Dec;5(1):2 [FREE Full text] [doi: 10.1007/s13755-017-0023-z] [Medline: 29038732]
36. Thornton C, Hutter F, Hoos HH, Leyton-Brown K. Auto-WEKA: combined selection and hyperparameter optimization of classification algorithms. In: Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2013 Presented at: KDD'13: 19th ACM SIGKDD international conference on Knowledge discovery and data mining; August 11-14, 2013; Chicago Illinois USA p. 847-855. [doi: 10.1145/2487575.2487629]
37. Agresti A. Categorical Data Analysis, 3rd ed. Hoboken, NJ: Wiley; 2012.
38. Hastie T, Tibshirani R, Friedman J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2nd ed. New York, NY: Springer; 2016.
39. Luo G. Automatically explaining machine learning prediction results: a demonstration on type 2 diabetes risk prediction. Health Inf Sci Syst 2016;4:2 [FREE Full text] [doi: 10.1186/s13755-016-0015-4] [Medline: 26958341]
40. Luo G. A roadmap for semi-automatically extracting predictive and clinically meaningful temporal features from medical data for predictive modeling. Glob Transit 2019;1:61-82 [FREE Full text] [doi: 10.1016/j.glt.2018.11.001] [Medline: 31032483]
41. Ranganathan P, Aggarwal R. Common pitfalls in statistical analysis: understanding the properties of diagnostic tests – Part 1. Perspect Clin Res 2018;9(1):40-43. [doi: 10.4103/picr.picr_170_17]
42. Kuhn M, Johnson K. Applied Predictive Modeling. New York, NY: Springer; 2013.
43. Asadi H, Dowling R, Yan B, Mitchell P. Machine learning for outcome prediction of acute ischemic stroke post intra-arterial therapy. PLoS One 2014;9(2):e88225 [FREE Full text] [doi: 10.1371/journal.pone.0088225] [Medline: 24520356]

44. Luo G, Nau CL, Crawford WW, Schatz M, Zeiger RS, Rozema E, et al. Developing a predictive model for asthma-related hospital encounters in patients with asthma in a large, integrated health care system: secondary analysis. JMIR Med Inform 2020 Nov 09;8(11):e22689 [FREE Full text] [doi: 10.2196/22689] [Medline: 33164906]
45. XGBoost Parameters. 2020. URL: https://xgboost.readthedocs.io/en/latest/parameter.html [accessed 2021-03-26]
46. Notes on Parameter Tuning. 2020. URL: https://xgboost.readthedocs.io/en/latest/tutorials/param_tuning.html [accessed 2021-03-26]
47. Luo G, Stone BL, Koebnick C, He S, Au DH, Sheng X, et al. Using temporal features to provide data-driven clinical early warnings for chronic obstructive pulmonary disease and asthma care management: protocol for a secondary analysis. JMIR Res Protoc 2019 Jun 06;8(6):e13783 [FREE Full text] [doi: 10.2196/13783] [Medline: 31199308]
48. Bourgeois FC, Olson KL, Mandl KD. Patients treated at multiple acute health care facilities: quantifying information fragmentation. Arch Intern Med 2010 Dec 13;170(22):1989-1995. [doi: 10.1001/archinternmed.2010.439] [Medline: 21149756]
49. Finnell JT, Overhage JM, Grannis S. All health care is not local: an evaluation of the distribution of emergency department care delivered in Indiana. AMIA Annu Symp Proc 2011;2011:409-416 [FREE Full text] [Medline: 22195094]
50. Luo G, Tarczy-Hornoch P, Wilcox AB, Lee ES. Identifying patients who are likely to receive most of their care from a specific health care system: demonstration via secondary analysis. JMIR Med Inform 2018 Nov 05;6(4):e12241 [FREE Full text] [doi: 10.2196/12241] [Medline: 30401670]
51. Kern LM, Grinspan Z, Shapiro JS, Kaushal R. Patients' use of multiple hospitals in a major US city: implications for population management. Popul Health Manag 2017 Apr;20(2):99-102 [FREE Full text] [doi: 10.1089/pop.2016.0021] [Medline: 27268133]
52. Samuels-Kalow ME, Faridi MK, Espinola JA, Klig JE, Camargo CAJ. Comparing statewide and single-center data to predict high-frequency emergency department utilization among patients with asthma exacerbation. Acad Emerg Med 2018 Jun;25(6):657-667 [FREE Full text] [doi: 10.1111/acem.13342] [Medline: 29105238]

Abbreviations

AUC: area under the receiver operating characteristic curve
ED: emergency department
FN: false negative
FP: false positive
ICD-10: International Classification of Diseases, Tenth Revision
ICD-9: International Classification of Diseases, Ninth Revision
NPV: negative predictive value
PCP: primary care provider
PPV: positive predictive value
TN: true negative
TP: true positive
UWM: University of Washington Medicine
Weka: Waikato Environment for Knowledge Analysis
XGBoost: extreme gradient boosting

Edited by R Kukafka; submitted 26.08.20; peer-reviewed by Y Xiang, J Fonseca; comments to author 20.10.20; revised version received 31.10.20; accepted 22.03.21; published 16.04.21.

Please cite as:
Tong Y, Messinger AI, Wilcox AB, Mooney SD, Davidson GH, Suri P, Luo G
Forecasting Future Asthma Hospital Encounters of Patients With Asthma in an Academic Health Care System: Predictive Model Development and Secondary Analysis Study
J Med Internet Res 2021;23(4):e22796
URL: https://www.jmir.org/2021/4/e22796
doi:10.2196/22796
PMID:33861206

Original Paper

Machine Learning–Driven Models to Predict Prognostic Outcomes in Patients Hospitalized With Heart Failure Using Electronic Health Records: Retrospective Study

Haichen Lv[1*], MD, PhD; Xiaolei Yang[1*], MD, PhD; Bingyi Wang[2*], MA; Shaobo Wang[3,4], PhD; Xiaoyan Du[2], MA; Qian Tan[5], MD, PhD; Zhujing Hao[1], MA; Ying Liu[1], MD, PhD; Jun Yan[4], PhD; Yunlong Xia[1], MD, PhD

[1]Department of Cardiology, The First Affiliated Hospital of Dalian Medical University, Dalian, China

[2]Medical Department, Yidu Cloud (Beijing) Technology Co Ltd, Beijing, China

[3]College of Life Science and Bioengineering, Beijing University of Technology, Beijing, China

[4]AI Lab, Yidu Cloud (Beijing) Technology Co Ltd, Beijing, China

[5]Medical Department, Happy Life Technology Co Ltd, Beijing, China

[*]these authors contributed equally

Corresponding Author:
Yunlong Xia, MD, PhD
Department of Cardiology
The First Affiliated Hospital of Dalian Medical University
193 Lianhe Road, Shahekou District
Dalian, 116014
China
Phone: 86 18098875555
Email: yunlongxia01@163.com

Abstract

Background: With the prevalence of cardiovascular diseases increasing worldwide, early prediction and accurate assessment of heart failure (HF) risk are crucial to meet the clinical demand.

Objective: Our study objective was to develop machine learning (ML) models based on real-world electronic health records to predict 1-year in-hospital mortality, use of positive inotropic agents, and 1-year all-cause readmission rate.

Methods: For this single-center study, we recruited patients with newly diagnosed HF hospitalized between December 2010 and August 2018 at the First Affiliated Hospital of Dalian Medical University (Liaoning Province, China). The models were constructed for a population set (90:10 split of data set into training and test sets) using 79 variables during the first hospitalization. Logistic regression, support vector machine, artificial neural network, random forest, and extreme gradient boosting models were investigated for outcome predictions.

Results: Of the 13,602 patients with HF enrolled in the study, 537 (3.95%) died within 1 year and 2779 patients (20.43%) had a history of use of positive inotropic agents. ML algorithms improved the performance of predictive models for 1-year in-hospital mortality (areas under the curve [AUCs] 0.92-1.00), use of positive inotropic medication (AUCs 0.85-0.96), and 1-year readmission rates (AUCs 0.63-0.96). A decision tree of mortality risk was created and stratified by single variables at levels of high-sensitivity cardiac troponin I (<0.068 μg/L), followed by percentage of lymphocytes (<14.688%) and neutrophil count (4.870×10^9/L).

Conclusions: ML techniques based on a large scale of clinical variables can improve outcome predictions for patients with HF. The mortality decision tree may contribute to guiding better clinical risk assessment and decision making.

(J Med Internet Res 2021;23(4):e24996) doi:10.2196/24996

KEYWORDS

heart failure; machine learning; predictive modeling; mortality; positive inotropic agents; readmission

Introduction

Heart failure (HF) syndrome is a life-threatening chronic disorder with a global prevalence that has been rising consistently over recent decades because of population aging, shifts in disease spectrums, and improved survival rates among patients with various cardiovascular diseases [1,2]. HF is characterized by complex therapeutic regimens, frequent hospitalizations, and a poor prognosis, resulting in a tremendous health care burden [3]. In certain instances, these issues are also fundamental targets of a strategy for HF prevention and treatment [4,5]. It is crucial to discriminate accurately among patients with HF to identify those who have a high risk of in-hospital mortality and readmission, as well as to guide the use of different therapies based on patients' features.

A prediction model was developed as an important risk assessment tool and used in various health care areas over the past decades. It has been recognized to facilitate early identification of patients at disease or event risk and enables effective interventions for those who might benefit most from identifying specific risk factors [6,7]. Previous studies in the field of cardiology, generally based on different populations with HF, have constructed models that are relevant to prognosis prediction, including the Seattle Heart Failure Model (SHFM) [8], Munich score [9], Enhanced Feedback for Effective Cardiac Treatment (EFFECT) scale [10], and Acute Decompensated Heart Failure National Registry (ADHERE) [11]. Moreover, several available parameters related to increased mortality have been identified as well, such as age [10,12,13], concentration of B-type natriuretic peptide (BNP) [14,15], urea nitrogen level [10,12], and systolic blood pressure (SBP) [9,13]. Although the model's construction from such cohorts or databases provides a level of concrete evidence, it is typically limited to large volumes of clinical resources and unstructured data [16]. Given the growing popularity of big data use and mining, data derived from electronic health records (EHRs) are becoming more available and accessible for clinical research.

Besides this, predictive models have been constructed in various health care domains with a certain degree of success by automated mining of EHRs, combined with machine learning (ML) approaches [17], specifically for the prediction of HF outcomes [18-20]. In contrast to previous studies of predictive models for HF outcomes that apply traditional methods, recent research is adopting ML techniques for predicting HF mortality, readmission, and medication adherence, which might demonstrate better performance in their predictions because of their consideration of higher order and nonlinear relationships between multidimensional variables [21].

In our study, we explored the use of the traditional method—logistic regression (LR)—and the four novel ML approaches—support vector machine (SVM), artificial neural network (ANN), random forest (RF), and extreme gradient boosting (XGBoost)—to predict prognostic outcomes for subjects with HF in real-world settings. We demonstrated the development of EHR-based models to predict the 1-year in-hospital mortality, use of positive inotropic agents, and all-cause readmissions in a single year.

Methods

Patients and Data Source

We collected the EHR data from hospitalized patients diagnosed with HF at the First Affiliated Hospital of Dalian Medical University (Liaoning Province, China) over 7 years between December 2010 and August 2018. All patients with HF were diagnosed and treated according to institutional guidelines. Because the diagnostic terminologies of EHRs have been structured and normalized according to the International Statistical Classification of Diseases and Related Health Problems, 10th Revision (ICD-10 [22]), newly diagnosed patients with HF (aged >18 years) were screened by a dynamically updated big data intelligence platform developed by an artificial intelligence technology company in collaboration with hospitals (Yidu Cloud Technology Co., Ltd). Each patient's data were extracted from various EHR systems integrated into a single platform, including the hospital information system, electronic medical record, radiation information system, laboratory information system, ultrasound system, and electrocardiogram system. The study was finally approved by the ethics committee of the First Affiliated Hospital of Dalian Medical University. Written informed consent was waived due to the retrospective design.

Selection of Variables

The platform incorporates comprehensive and detailed data on patients' routine health care. Common and cardiovascular-specific variables were structured and normalized by natural language processing, ML techniques, and well-defined logical rules. Considering the large scale of EHR data and the advantages of ML techniques, candidate variables with known clinical significance and inaccessible through traditional medical records were collected in our study. We excluded variables with missing values greater than 20%. Finally, a total of 79 variables related to the first hospitalization were extracted from a big data intelligence platform. The features were as follows: demographics (age and sex), personal history (smoking and drinking), history (comorbidities and surgery), etiology, vital signs, routine laboratory examinations, interventions, and medication use on admission (Multimedia Appendix 1). Repeated measurements of vital signs and laboratory tests from patients with HF were taken over different periods to ensure data accuracy.

Outcomes

We established models using 5 algorithms separately to predict the primary outcome: all-cause in-hospital mortality within 1 year. The secondary outcomes were (1) use of positive inotropic agents for patients with HF and (2) all-cause readmission. Mortality was defined as a clear record of death of an inpatient within 1 year of hospitalization. The following commonly used positive inotropic agents in clinical practice were selected for our study: dopamine, dobutamine, milrinone, levosimendan, and deslanoside. Readmission was defined as any patient with an interval of more than 1 day from the last discharge until the next admission; no readmitted patients died within 1 year of hospitalization. Therefore, enrolled patients were labeled as "deceased" or "survivors."

ML Model Development and Performance Evaluation

The synthetic minority oversampling technique was first employed to address the imbalance in the data set, with a ratio of 1:1 between deceased patients and survivors. It is common for a minority to be oversampled by deriving new, "synthetic" samples to alleviate imbalance [23].

The hold-out method was used to divide the data set into a training/validation data set (90%) and a hold-out test set (10%). Five approaches were explored with 10-fold cross-validation for building models and adjusting parameters: LR, RF, SVM, ANN, and XGBoost. The hold-out test set was used to evaluate the best-performing models created with the training set. The area under the curve (AUC) of the receiver operating characteristic (ROC) curve was chosen as the primary evaluation metric for our models, including accuracy, precision, and recall. The Brier score (range of 0 to 1)—the average squared error between the predicted and the actual value—was also commonly represented as a "calibration" for overall measurement. Shapley additive explanation (SHAP) values were used to evaluate feature importance.

LR Method

LR is the most basic dichotomous linear method of model selection that makes classification decisions. LR is superior in measuring the probability between "0" and "1" based on the relationships of binary classifications in continuous or categorical variables [24]. We used the sklearn.linear module to develop the LR models.

RF Method

RF is an algorithm that integrates multiple decision tree classifiers. Each node of the decision tree represents a predictive variable to separate the outcome classes by setting the optimal threshold. The importance of features can also be obtained with the sum of weights of the classifier's nodes [25]. The RF method is generally used to deal with thousands of input variables without dimension reduction. We used the RF classifier from the sklearn.ensemble module to develop the models. The related parameters (n_estimators, max_features, max_depth, and min_samples_split) were adjusted to prevent poor results and overfitting.

SVM Method

As a dichotomous supervised algorithm, SVM can be used in high-dimensional feature space. The best hyperplane can be achieved using a kernel-based function for separating two classes at maximum intervals [26].

ANN Method

A multilayer perceptron (MLP) classifier was implemented using the sklearn.neural_network module to develop models [27]. MLP is a popular ANN method that generally consists of neurons from the input, hidden, and output layers. The data were processed through weighted connections and activation functions in the hidden layers [28]. In our study, the two hidden layers had 40 and 20 neurons. The rectified linear unit was chosen as their activation function.

XGBoost Method

XGBoost is one of the boosting methods; this algorithm aims to integrate weak classifiers into a single robust classifier in an iterative fashion [29]. This algorithm constructs a scalable classification and regression tree in a boosting ensemble manner on a gradient boosting decision tree basis, which can learn nonlinear relationships among variables and outcomes flexibly and accurately [30]. The small learning rate (0.1) indicates better generalization. The tree number and maximum depth were limited to 80 (estimators) and 3, respectively.

Statistical Analysis

All ML algorithms were performed using the scikit-learn (version 0.21.1) package in Python (version 3.6.5; Python Software Foundation), and statistical analysis was conducted using an open-source Scipy (version 1.3.0) database from Python (version 3.6.3). All categorical data were presented as percentages. All continuous data performing a normal distribution were presented as mean (SD); otherwise, they were expressed as median (IQR). Student *t* tests or chi-square tests were applied for group comparisons. *P* values <.05 were considered statistically significant.

Results

Baseline Characteristics

A total of 13,602 hospitalized patients with newly diagnosed HF were enrolled in this study, of whom 3.95% (n=537) died in hospital within 1 year. The usage rate of positive inotropic agents was 20.43% (n=2779). The overall all-cause readmission rates of 30 days, 60 days, and 1 year for patients with HF were 4.83%, 14.77%, and 21.16%, respectively (n=657, n=2009, and n=2878, respectively, of 13,602 cases). The eligible population's baseline characteristics were compared between 2 groups according to survival status (Table 1). Patients in the deceased group were older than those in the survivor group (77 years, IQR 66.5-83.0 years versus 72 years, IQR 63.0-80.0 years, respectively). The proportion of male patients in both groups was 52.5% (282/537 and 6860/13,065 in the deceased group and survivor group, respectively). The number of patients with comorbid diagnoses of diabetes, hypertension, and tumors was remarkably different between the 2 groups (all *P*<.001). Concerning the etiology of HF, the number of patients with cardiomyopathy and cardiac arrhythmia was found to be different between the deceased and survivor groups (both *P*<.001). There was a significant difference in the vital signs (heart rate and blood pressure) of patients with HF between the 2 groups (all *P*<.001). In-hospital medication use (angiotensin-converting enzyme inhibitors [ACEIs] and aldosterone receptor antagonists [ARBs]) was higher among subjects who survived in the hospital (*P*<.001).

Table 1. Demographic and clinical variables of the deceased and survivor groups (N=13,602).

Variables	Deceased group	Survivor group	*P* value
Total patients, n (%)	537 (3.9)	13,065 (96.1)	
Demographic information			
Age (years), median (IQR)	77.0 (66.5-83.0)	72.0 (63.0-80.0)	.003
Gender (male), n (%)	282 (52.5)	6860 (52.5)	.997
Smoking history, n (%)	115 (21.4)	3282 (25.1)	.05
Drinking history, n (%)	64 (11.9)	1820 (13.9)	.19
Comorbidities, n (%)			
Diabetes mellitus	202 (37.6)	3593 (27.5)	<.001
Hypertension	338 (62.9)	7073 (54.1)	<.001
Dyslipidemia	313 (58.2)	8449 (64.7)	.003
COPD[a]	3 (0.5)	50 (0.4)	.47
Chronic renal disease	13 (2.4)	144 (1.1)	.005
Tumors	46 (8.6)	534 (4.1)	<.001
Etiology of heart failure, n (%)			
Coronary heart disease	349 (65.0)	7810 (59.8)	.02
Cardiomyopathy	22 (4.1)	1207 (9.2)	<.001
Valvular heart disease	91 (16.9)	2389 (18.3)	.43
Cardiac arrhythmia	193 (35.9)	5780 (44.2)	<.001
History of cardiovascular surgery	115 (21.4)	2519 (19.3)	.22
Vital signs, median (IQR)			
Blood pressure (mmHg)			
Diastolic	77.0 (68.0-84.0)	80.0 (70.0-90.0)	<.001
Systolic	130.0 (115.0-150.0)	140.0 (120.0-152.0)	<.001
Heart rate (beats/min)	84.0 (72.0-99.0)	76.0 (68.0-90.0)	<.001
Respiratory rate (breaths/min)	19.0 (18.0-20.0)	18.0 (17.0-19.0)	.66
Temperature	36.2 (36.0-36.5)	36.2 (36.0-36.4)	.003
NYHA[b] classification, n (%)			<.001
IV	113 (21.0)	1424 (10.9)	
III	96 (17.9)	4006 (30.7)	
II	22 (4.1)	1738 (13.3)	
I	0 (0)	12 (0.1)	
None	297 (55.3)	5291 (40.5)	
Laboratory indicators at admission, median (IQR)			
BNP[c]	1053.5 (399.5-2383.3)	322.9 (106.6-845.0)	<.001
hs-cTnl[d]	0.4 (0.1-5.2)	0.03 (0.01-0.11)	<.001
Creatine kinase MB (U/L)	2.8 (1.4-8.0)	1.5 (0.8-2.6)	<.001
Hemoglobin (g/L)	115.0 (94.0-133.0)	131.0 (117.0-144.0)	<.001
Platelets	180.5 (125.8-242.0)	193.0 (155.0-235.0)	.001
White blood cells ($\times 10^9$/L)	9.5 (6.4-14.0)	6.6 (5.3-8.2)	<.001
Red blood cells	3.9 (3.2-4.4)	4.3 (3.9-4.8)	<.001
Lymphocytes	1.1 (0.7-1.7)	1.6 (1.1-2.1)	.59

Variables	Deceased group	Survivor group	*P* value
Neutrophils	7.1 (4.5-11.3)	4.1 (3.1-5.5)	<.001
Mean platelet volume (fL)	10.8 (10.0-11.7)	10.7 (10.0-11.4)	.004
Hematocrit	34.1 (26.1-39.7)	38.7 (33.2-42.7)	<.001
Basophils ($\times10^9$/L)	0.02 (0.01-0.03)	0.02 (0.01-0.04)	.55
Monocytes ($\times10^9$/L)	0.6 (0.4-0.9)	0.5 (0.4-0.7)	<.001
Monocytes (%)	6.7 (4.5-9.0)	7.9 (6.4-9.7)	<.001
Mean corpuscular volume (fL)	91.1 (87.8-94.2)	91.0 (87.9-94.2)	.62
Procalcitonin	0.4 (0.1-1.9)	0.1 (0.1-0.3)	.20
Neutrophils (%)	78.3 (67.6-87.2)	63.1 (55.5-71.1)	<.001
Basophils (%)	0.2 (0.1-0.4)	0.4 (0.2-0.5)	<.001
Eosinophils (%)	0.8 (0.2-1.9)	1.7 (0.9-2.9)	<.001
Eosinophils ($\times10^9$/L)	0.1 (0.03-0.2)	0.1 (0.1-0.2)	.002
Lymphocytes (%)	12.7 (6.7-21.4)	25.6 (18.2-32.7)	<.001
Total bilirubin (μmol/L)	16.3 (11.2-27.3)	14.5 (10.5-20.5)	<.001
Direct bilirubin (μmol/L)	5.4 (3.4-9.4)	4.6 (3.2-6.9)	<.001
Glucose (mmol/L)	6.5 (5.1-9.3)	5.5 (4.9-6.8)	<.001
Lipoprotein(a) (mg/L)	165.3 (84.7-307.5)	152.1 (83.0-277.0)	.08
High-density lipoprotein cholesterol (mmol/L)	1.3 (0.9-34.0)	1.5 (1.0-39.0)	.01
Low-density lipoprotein cholesterol (mmol/L)	3.2 (2.2-80.0)	3.5 (2.3-90.0)	.08
Total cholesterol (mmol/L)	5.3 (3.9-137.3)	5.8 (4.2-155.0)	.03
Triglycerides (mmol/L)	1.8 (1.0-80.0)	2.1 (1.1-91.0)	.01
Alanine aminotransferase (U/L)	27.0 (14.0-61.0)	20.0 (13.0-33.0)	<.001
Aspartate aminotransferase (U/L)	36.0 (20.0-101.0)	21.0 (16.0-30.0)	<.001
Gamma-glutamyl transferase (U/L)	46.0 (25.0-87.0)	35.0 (22.0-63.0)	<.001
Albumin (g/L)	35.0 (30.6-38.6)	39.3 (36.2-41.9)	<.001
Globulin (g/L)	28.6 (24.4-32.9)	26.7 (23.6-30.3)	<.001
Albumin/globulin ratio	1.2 (1.0-1.5)	1.5 (1.3-1.7)	<.001
Total protein (g/L)	63.1 (58.1-69.3)	65.9 (61.5-70.4)	<.001
Creatinine (μmol/L)	110.0 (77.0-201.0)	77.0 (62.0-98.0)	<.001
Sodium (mmol/L)	138.5 (135.0-142.0)	141.0 (138.8-143.0)	<.001
Potassium (mmol/L)	4.0 (3.7-4.5)	4.0 (3.7-4.3)	<.001
Calcium (mmol/L)	2.1 (2.0-2.2)	2.2 (2.1-2.3)	<.001
Uric acid (μmol/L)	446.0 (321.8-611.3)	390.0 (311.0-489.0)	<.001
Urea (mmol/L)	11.4 (7.4-19.3)	7.1 (5.6-9.5)	<.001
Alkaline phosphatase (U/L)	84.0 (69.0-116.0)	74.0 (62.0-92.0)	<.001
Acetylcholinesterase (U/L)	210.5 (148.0-281.0)	292.0 (229.0-363.0)	<.001
International normalized ratio	1.2 (1.1-1.4)	1.1 (1.0-1.21)	<.001
Prothrombin time (s)	13.2 (11.9-15.8)	11.9 (11.0-13.2)	<.001
Fasting blood glucose (g/L)	3.6 (2.7-4.5)	3.0 (2.5-3.7)	<.001
Activated partial thromboplastin time (s)	31.7 (26.5-41.8)	26.8 (23.9-30.8)	<.001
Use of devices during hospitalization, n (%)			
Cardiac resynchronization therapy	2 (0.4)	42 (0.3)	.69

Variables	Deceased group	Survivor group	*P* value
ICD[e] implantation	0 (0)	32 (0.2)	.64
Permanent pacemaker	3 (0.6)	350 (2.7)	.002
Temporary pacemaker	0 (0)	15 (0.1)	>.99
Medication use during hospitalization, n (%)			
ACEI[f]/ARB[g]	207 (38.5)	7967 (61.0)	<.001
β-blocker	408 (76.0)	10,199 (78.1)	.25
Aldosterone antagonist	266 (49.5)	8364 (64.0)	<.001
Statin	310 (57.7)	8399 (64.3)	.002
Aspirin	331 (61.6)	8255 (63.2)	.47
Diuretic	508 (94.6)	10,927 (83.6)	<.001
Digoxin	73 (13.6)	2424 (18.6)	.004
Outcome, n (%)			
Positive inotropic agents use			
Dopamine	319 (59.4)	1407 (10.8)	<.001
Dobutamine hydrochloride	39 (7.3)	225 (1.7)	<.001
Milrinone	52 (9.7)	325 (2.5)	<.001
Levosimendan	7 (1.3)	37 (0.3)	.002
Lanatoside C	129 (24.0)	1132 (8.7)	<.001
Readmissions			
30 days	53 (9.9)	604 (4.6)	<.001
180 days	129 (24.0)	1880 (14.4)	<.001
1 year	162 (30.2)	2716 (20.8)	<.001

[a]COPD: chronic obstructive pulmonary disease.

[b]NYHA: New York Heart Association.

[c]BNP: B-type natriuretic peptide.

[d]hs-cTnI: high-sensitivity cardiac troponin I.

[e]ICD: implantable cardioverter defibrillator.

[f]ACEI: angiotensin-converting enzyme inhibitor.

[g]ARB: angiotensin receptor blocker.

1-Year In-Hospital Mortality Models

Predictive models for 1-year in-hospital mortality risk assessment were conducted using 5 algorithms. Figure 1A shows the performances of models in the form of AUC. The AUC values for LR, RF, SVM, ANN, and XGBoost were 0.91, 1.00, 0.99, 0.99, and 0.99, respectively. RF had relatively higher AUC than the other algorithms. The calibration plots of our 5 methods are presented in Figure 2A. Four ML models had an accuracy of higher than 95%. Regarding precision, the RF and ANN algorithms emerged as the best, achieving the highest precision (0.96), followed by the SVM (0.93) and XGBoost (0.91) algorithms. The Brier score for RF and ANN was the lowest (0.03) (Table 2).

Figure 1. Receiver operating characteristic (ROC) curves using the synthetic minority oversampling technique for the logistic regression, random forest, support vector machine, artificial neural network (ANN), and extreme gradient boosting (XGBoost) models in predicting (A) 1-year in-hospital mortality, (B) use of positive inotropic agents, and (C) 1-year all-cause readmission. AUC: area under the curve.

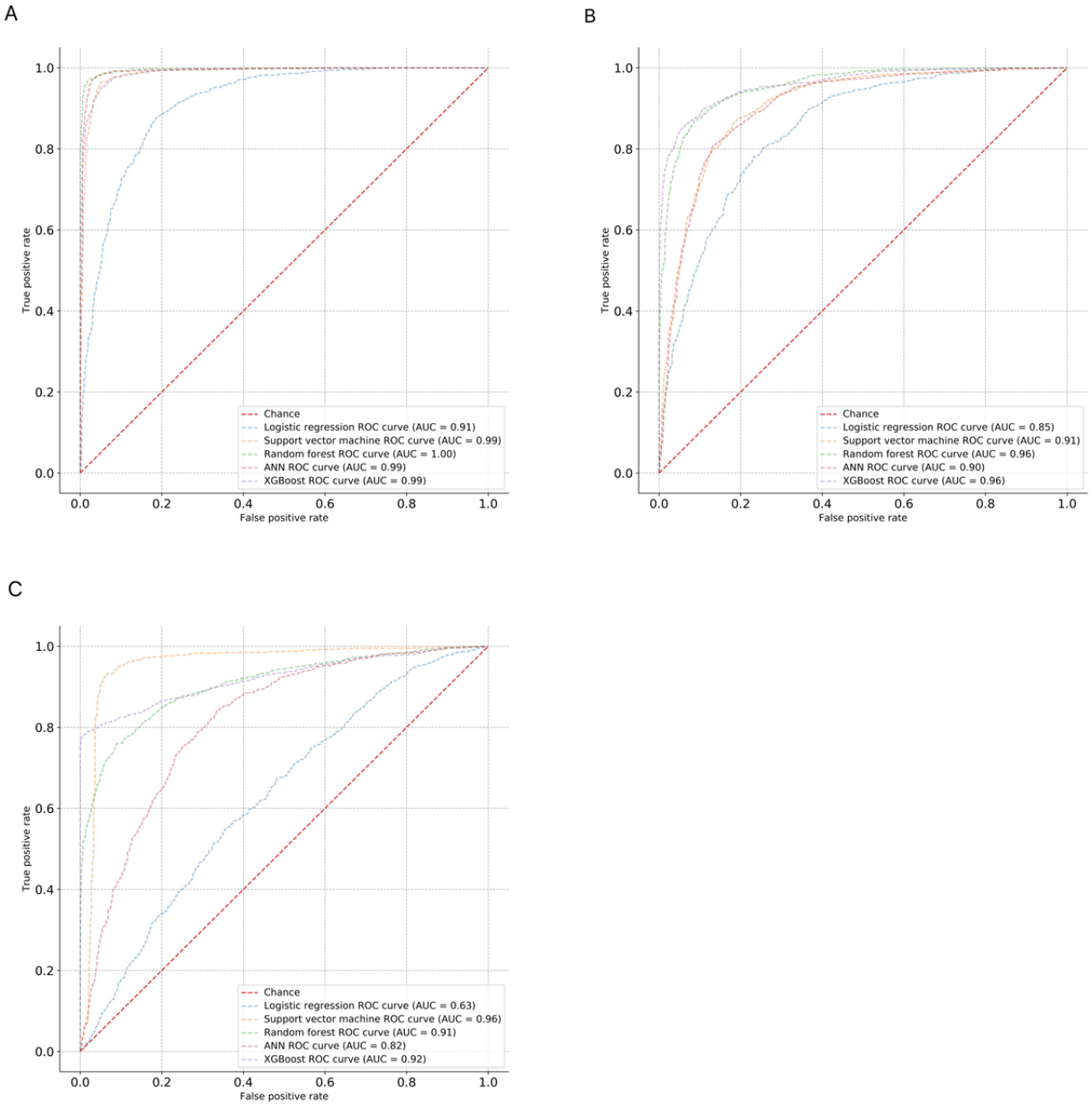


Figure 2. Calibration plots using the synthetic minority oversampling technique for the logistic regression, random forest, support vector machine, artificial neural network (ANN), and extreme gradient boosting (XGBoost) models in predicting (A) 1-year in-hospital mortality, (B) use of positive inotropic agents, and (C) 1-year all-cause readmission.

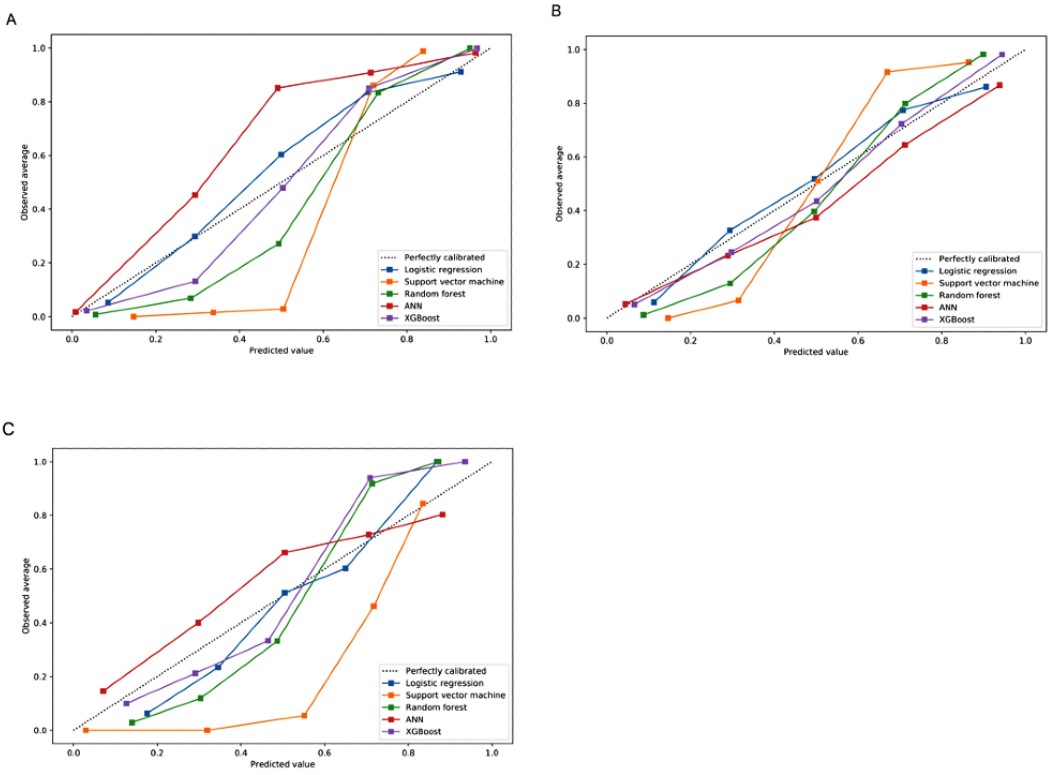


Table 2. Performance of the machine learning approaches for the estimation of 1-year in-hospital all-cause mortality.

Model	AUC[a]	Accuracy	Precision	Recall	F1	Brier score
LR[b]	0.91	0.83	0.86	0.80	0.83	0.12
RF[c]	1.00	0.97	0.96	0.98	0.97	0.03
SVM[d]	0.99	0.94	0.93	0.96	0.94	0.16
ANN[e]	0.99	0.97	0.96	0.98	0.97	0.03
XGBoost[f]	0.99	0.94	0.91	0.98	0.94	0.05

[a]AUC: area under the curve.

[b]LR: logistic regression.

[c]RF: random forest.

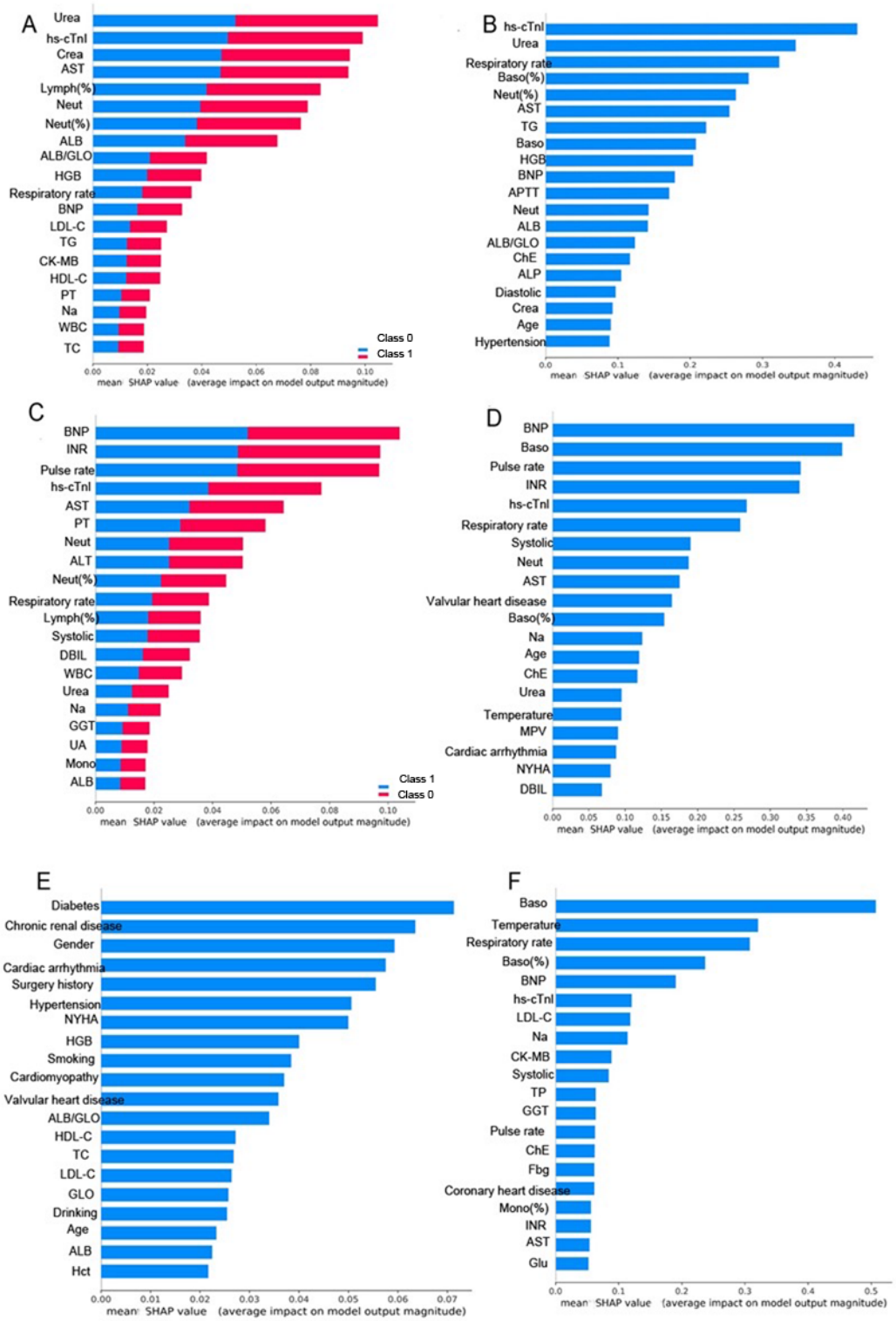
[d]SVM: support vector machine.

[e]ANN: artificial neural network.

[f]XGBoost: extreme gradient boosting.

Furthermore, we explored the importance of the features that affect mortality prediction by applying RF and XGBoost approaches (Figure 3), with the weight assignment of each feature expressed as a SHA*P* value based on whether it favored judgment in survival. As shown in Figure 3A and 3B, blood urea, high-sensitivity cardiac troponin I (hs-cTnI), creatinine, aspartate aminotransferase (AST), and percentage of lymphocytes were the top 5 related variables in mortality prediction. In contrast, hs-cTnI was the most crucial marker identified using the XGBoost algorithm, followed by urea, respiration rate, percentage of basophils, and percentage of neutrophils.

Figure 3. Shapley additive explanation (SHAP) plots for the machine learning models in predicting (A) 1-year in-hospital mortality using the random forest (RF) method, (B) 1-year in-hospital mortality using the extreme gradient boosting (XGBoost) method, (C) use of positive inotropic agents using the RF method, (D) use of positive inotropic agents using the XGBoost method, (E) 1-year all-cause readmission using the support vector machine (SVM) method, and (F) 1-year all-cause readmission using the XGBoost method. ALB: albumin; ALP: alkaline phosphatase; APTT: activated partial thromboplastin time; AST: aspartate aminotransferase; Baso: basophils; BNP: B-type natriuretic peptide; ChE: cholinesterase: CK-MB: creatine kinase MB; COPD: chronic obstructive pulmonary disease; Crea: creatinine; DBIL: direct bilirubin; Fbg: fibrinogen; GGT: gamma-glutamyl transferase; GLO: globulin; Glu: glucose; Hct: hematocrit; HDL-C: high-density lipoprotein cholesterol; HGB: hemoglobin; hs-cTnI: high-sensitivity cardiac troponin I; INR: international normalized ratio; LDL-C: low-density lipoprotein cholesterol; Lymph(%): percentage of lymphocytes; Mono: monocytes; MPV: mean platelet volume; Na: sodium; Neut: neutrophils; NYHA: New York Heart Association; PT: prothrombin time; TC: total cholesterol; Systolic: systolic blood pressure; TG: triglycerides; TP: total protein; UA: uric acid; WBC: white blood cells.



Positive Inotropic Agent Use Models

Figure 1B demonstrates the ROC of the predictive models in patients' use of positive inotropic agents. In comparing AUCs among the 5 models, the XGBoost and RF models had the highest AUC value (0.96), followed by the SVM model (0.91), ANN model (0.90), and LR model (0.85). In consideration of precision (0.85) and recall (0.91) values, RF was determined to be the best method for prediction (Table 3). As shown in calibration curves (Figure 2B), a nonparametric plot of the RF algorithm was close along the ideal diagonal line and had the lowest Brier score (0.10).

Table 3. Performance of the machine learning approaches for the estimation of use of positive inotropic agents.

Model	AUC[a]	Accuracy	Precision	Recall	F1	Brier score
LR[b]	0.85	0.78	0.77	0.79	0.78	0.16
RF[c]	0.96	0.87	0.85	0.91	0.88	0.10
SVM[d]	0.91	0.85	0.83	0.88	0.84	0.17
ANN[e]	0.90	0.83	0.78	0.94	0.84	0.12
XGBoost[f]	0.96	0.84	0.79	0.94	0.86	0.11

[a]AUC: area under the curve.

[b]LR: logistic regression.

[c]RF: random forest.

[d]SVM: support vector machine.

[e]ANN: artificial neural network.

[f]XGBoost: extreme gradient boosting.

Interestingly, BNP, international normalized ratio (INR), pulse rate, hs-cTnI, and AST were the top 5 markers predicting use of positive inotropic agents by the RF approach. BNP was also identified as a critical marker for forecasting positive inotropic agent use in the XGBoost model, followed by basophil counts, pulse rate, INR, and hs-cTnI (Figure 3C and 3D).

1-Year All-Cause Readmission Models

The discrimination of different models for 1-year all-cause readmissions represented by AUCs is shown in Figure 1C. The SVM method achieved the best performance in terms of 1-year readmission prediction (AUC 0.96). XGBoost showed the lowest Brier score for calibration plots (0.12), followed by RF (0.13), SVM (0.16), ANN (0.18), and LR (0.24) (Figure 2C and Table 4).

Table 4. Performance of the machine learning approaches for the estimation of 1-year all-cause readmission.

Model	AUC[a]	Accuracy	Precision	Recall	F1	Brier score
LR[b]	0.63	0.57	0.57	0.59	0.58	0.24
RF[c]	0.91	0.82	0.83	0.81	0.82	0.13
SVM[d]	0.96	0.90	0.86	0.96	0.91	0.16
ANN[e]	0.82	0.74	0.74	0.75	0.74	0.18
XGBoost[f]	0.92	0.83	0.82	0.84	0.83	0.12

[a]AUC: area under the curve.

[b]LR: logistic regression.

[c]RF: random forest.

[d]SVM: support vector machine.

[e]ANN: artificial neural network.

[f]XGBoost: extreme gradient boosting.

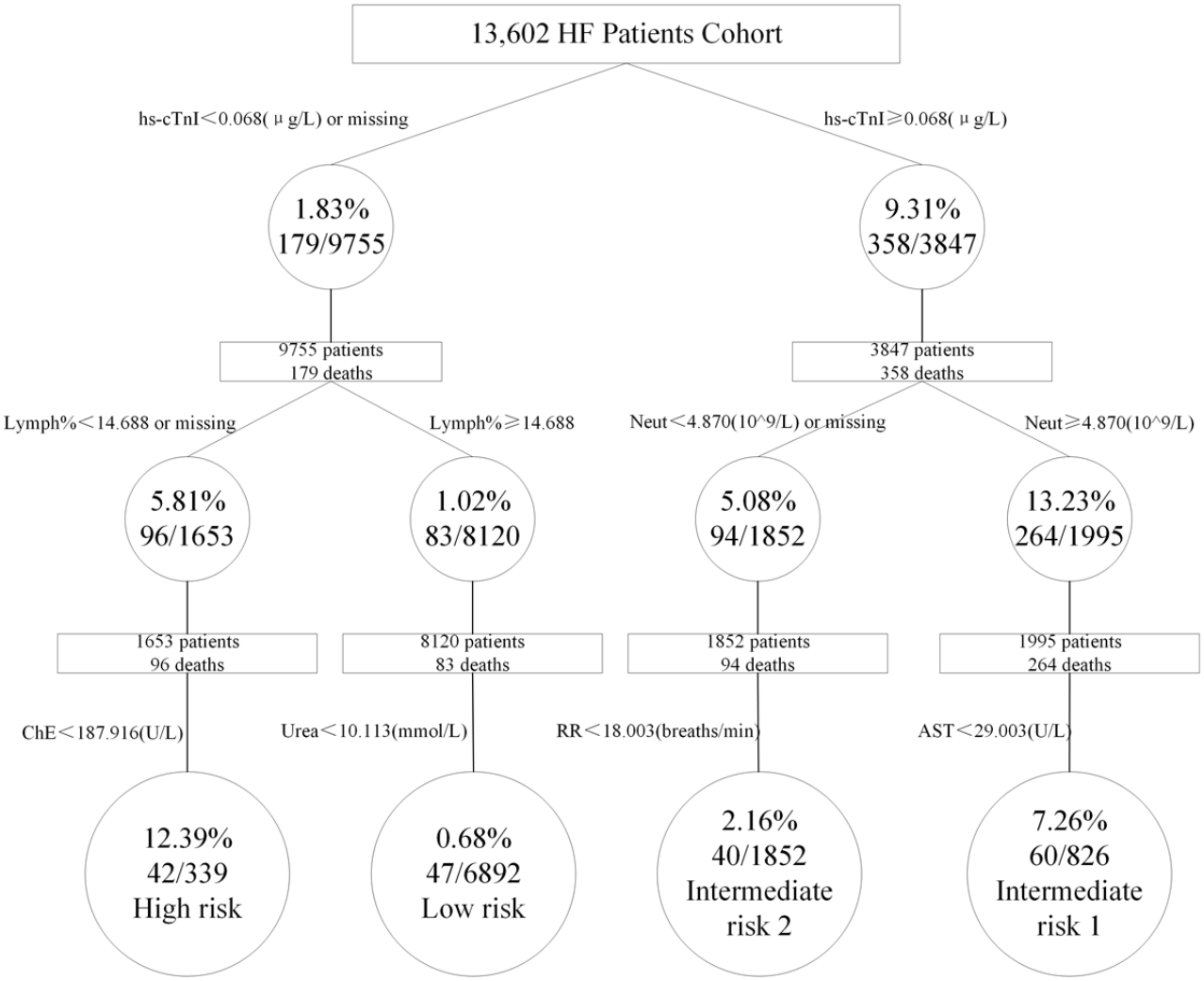
Of the 79 variables analyzed, the presence of diabetes was identified as the most important marker for prediction of 1-year all-cause readmission using the SVM method, whereas basophil count was a significant predictor of readmission using the XGBoost algorithm. There appeared to be a certain discrepancy in the ranking of other features derived by the two methods (Figure 3E and 3F).

Mortality Risk Assessment Model

The patients were divided into subgroups of high-, intermediate-, and low-risk in-hospital mortality according to the cutoff value of various variables obtained using the XGBoost algorithm (Figure 4). Of the 79 variables, the hs-cTnI level (<0.068 μg/L) was identified as the first single predictor to discriminate between deceased and surviving patients. The next discriminator

in the left node of hs-cTnI was the percentage of lymphocytes at a discrimination level of <14.688%; conversely, the next discriminator in the right node of hs-cTnI was neutrophil count at a cutoff value of less than 4.870×10^9/L. Subsequently, the patients with HF were stratified by these branch points: high risk (hs-cTnI <0.068 μg/L, percentage of lymphocytes <14.688%, and cholinesterase <187.916 U/L); low risk (hs-cTnI <0.068 μg/L, percentage of lymphocytes ≥14.688%, and urea <10.113 mmol/L); intermediate risk 1 (hs-cTnI ≥0.068 μg/L, neutrophil count ≥4.870×10^9/L, and AST <29.003 U/L); and intermediate risk 2 (hs-cTnI ≥0.068 μg/L, neutrophil count <4.870×10^9/L, and respiratory rate <18.003 breaths/min).

Figure 4. Predictors of 1-year in-hospital mortality and risk stratification using an extreme gradient boosting (XGBoost) algorithm. AST: aspartate transaminase; ChE: cholinesterase; HF: heart failure; hs-cTnI: high-sensitivity cardiac troponin I; Lymph%: percentage of lymphocytes; neut: neutrophil count; RR: respiration rate.



Discussion

Principal Findings

In our study, the EHR data–driven prognostic models based on 5 different ML algorithms were developed for predicting 1-year in-hospital mortality, use of positive inotropic agents, and all-cause readmissions within 1 year for patients with HF in a single Chinese class A tertiary comprehensive hospital. These ML models produced better predictive values for prognostic outcomes by growing interpretability than traditional linear methods. Besides, the novel ML techniques could take advantage of a large scale of complex, high-dimensional variables to widen the scope of HF predictive indicators concerning prognostic outcomes.

An integrated EHR system comprises various data resources, including patients' demographic data, diagnostic information, laboratory test results, and prescriptions. However, many experts revealed that the availability of EHR-derived data is a prerequisite for promoting real-world studies [31,32]. In our study, the structured EHR data from over 10,000 patients in different years can be used directly for analysis without bias of data collection and clinical definitions. A total of 79 variables were applied from EHRs, including demographic information (n=4), current and previous disease history (n=8), HF diagnosis (n=3), vital signs (n=5), laboratory measurements (n=48), interventions (n=4), and medications (n=7). In general, the final target of using rich data derived from mature EHR systems is also patient outcome predictions. With the advancement of data science, different kinds of ML techniques have been broadly employed for model training because of their deep data

processing and decision-making capabilities [33]. Compared with previous studies using the traditional LR method [11,20], our results provide better discrimination of risk for prognostic outcomes in a population of patients hospitalized for HF by using ML approaches. The strength of the algorithm could establish complex models and make accurate decisions in relevant big data sets. Besides, far more variables were allowed for modeling by using ML algorithms. Given the 5 methods in the present study, LR was considered the commonly used method in various fields of medicine [12,18]. The RF algorithm involves multiple decision tree creations that identify important predictive features with better accuracy in processing large numbers of highly nonlinear data [20,34]. ANN is a complex network method connected by a large number of simple neurons and simulates the human brain in parallel processing and nonlinear transformation [35]. The SVM approach is designed to build a good classifier using a nonlinear decision boundary between classes of variables that enables the labels from one or more feature vectors [36]. From the perspective of AUC values, it is important to mention that XGBoost performed better than the other 4 methods in predicting 1-year in-hospital mortality, use of positive inotropic agents, and 1-year all-cause readmission in patients with HF. The method's dominant advantages are its ability to deal with missing values and integrate the power of weaker classifiers by creating combined and weighted variables. Relevant parameters were set to particular values.

Several traditional prognostic models have been established to estimate mortality in hospitalized patients with HF. Our results are consistent with the previous studies of 1-year mortality prediction while demonstrating a better predictive capability. Our study, which predicted outcomes for 13,602 patients with HF using 79 predictive variables, indicated that the ML approaches are superior to the EFFECT method (AUC 0.77) in predicting mortality, with an average AUC of 0.81 [10]. The classic SHFM study showed mortality prediction results within 1 year with AUCs of 0.729-0.76 [37], although the AUC value can be enhanced to 0.78 by combining the SHFM with the BNP level [38]. Furthermore, variables known to be associated with mortality in patients with HF were demonstrated in various previous models. Age, SBP, and level of blood urea nitrogen (BUN) were confirmed as the strongest independent predictors by the EFFECT model [10]. In the landmark Randomized Evaluation of Mechanical Assistance for the Treatment of Congestive Heart Failure (REMATCH) trial model [39], platelet counts, albumin, INR, and AST were associated with 1-year mortality in the HF population with implantable devices. The New York Heart Association class; use of beta-blockers, ACEIs, or ARBs; and presence of heart valve disease or atrial fibrillation remarkably influenced all-cause mortality in the Cardiac and Comorbid Conditions Heart Failure (3C-HF) score [40]. Our study obtained similar findings, demonstrating that in addition to specific biomarkers of HF or myocardial damage, renal function, coagulation indicators, and inflammation indicators are also critical factors regarding prognosis. Therefore, perhaps adopting a large scale of EHR-derived data could enhance discrimination and the predictive range of prognostic outcomes.

Readmission rate is a common index used to assess the quality of health care services for patient populations. Currently, most hospitals and institutions still implement traditional readmission risk models and certain variables to infer readmission probability [41]. Like previous prediction models, we found that the accuracy of mortality prediction was moderate, but the model was relatively poor at predicting readmission rates. Golas et al [42] suggested that the overall 30-day readmission rate was not improved by various ML algorithms, with AUC values around 0.66. Even compared with the established models for HF readmission (AUCs 0.6-0.7), our best ML model is still encouraging [43]. The variables that influenced mortality in the ML models were different from those of readmission among patients with HF. We believe that conventional diagnostic biomarkers (BNP) and vital signs (respiration rate and temperature) could dominate readmission prediction. However, diabetes, coronary heart disease, and gender were the top 3 significant features for predicting readmission using the SVM method. At an interpretable level, a high model AUC does not necessarily indicate that it best fits the scenario.

Positive inotropic agents are a kind of drug that can increase myocardial contractility and cardiac output and are often used to treat patients with HF. Among the drugs included in our study, dopamine and dobutamine have mainly inotropic effects, whereas milrinone, lanatoside C, and levosimendan have extra vasodilation functions [44]. However, they have been gradually limited because of poorer outcomes of patients after accepting inotropic therapy. Thus, controversy still exists regarding their reasonable use in clinical practice. Our study was found to have better performance than two other prognostic models in predicting the use of positive inotropic agents (AUCs in the range of 0.87 to 0.96). Furthermore, this study's findings are consistent with Aljundi et al [45], who reported that the conventional cardiac biomarker BNP was predictive of inotropic agent use. Many studies have confirmed that comorbidities—such as dyslipidemia, chronic renal and liver impairment, and hyperglycemia—are associated with a higher likelihood of accepting inotropic treatment. Other critically associated variables (INR, AST, and basophil count) were obtained from our models and showed consistent results with common clinical thought. Our models provided good discrimination among patients at high risk for mortality, use of positive inotropic agents, and readmission. Incorporating more variables based on traditional models is of great significance to public health transformation for early identification and individualized intervention of people at high risk of HF outcomes.

The study's strength was that we also developed a stratified risk assessment tool for the prediction of 1-year in-hospital mortality using the XGBoost algorithm. Far more features—including hs-cTnl, percentage of lymphocytes, percentage of neutrophils, cholinesterase, urea, respiratory rate, and AST—were identified for the first time using an ML approach for mortality prediction. Neutrophilia was reported to be associated with an increased incidence of acute decompensated HF (ADHF) in patients with acute myocardial infarction, and lymphopenia is related to poor prognosis in patients with HF [46,47]. Results from Seo et al [48] showed that cholinesterase was a simple marker for predicting adverse outcomes in patients with ADHF and tended to provide more accurate prognostic information than other

objective nutritional features. These results can be attributed to the dimensionality and breadth of EHR-based data, facilitating the real-world study of HF in risk stratification, decision making, and disease management from multiple perspectives. Various HF risk models for predicting mortality have been developed abroad. The ADHERE risk tree by regression analysis [12] has demonstrated that patients with ADHF at low, intermediate, and high risk for in-hospital mortality can be easily identified using BUN, SBP, and creatinine obtained on hospital admission. The MUSIC risk model [49] showed that risk markers including atrial fibrillation, hyponatremia ≤138 mEq/L, N-terminal pro-brain natriuretic peptide >1.000 ng/L, and troponin positive were associated with cardiac mortality in a real-life setting. However, a limitation of the previous model is that it was based on a specific population of patients with HF (ie, symptomatic chronic HF). It should be noted that the etiology, clinical characteristics, and treatments of different phenotypes of HF are quite distinct. Therefore, we compiled all types of patients with HF in this study; it follows that the risk factors predicting mortality are likely to have been more comprehensive and provide new insights for further studies in specific subgroups. These markers possibly provide a more accurate risk evaluation of patients with HF, allowing early implementation of the appropriate intervention in daily public health practice, which leads to better outcomes in patients with HF.

There are several limitations to our study that are worth mentioning. First, this was an in-hospital outcome prediction study based on retrospective use of EHR-derived data. Although our models' performances were considerable on their own, the predictive power could be further adjusted and compared with the established reference tools. Second, the number of patients who died was small compared with the number of surviving subjects; although rich in terms of clinical variables, the imbalance problem remained. Third, validation in an external cohort was not done in the present study but is planned for subsequent analysis. Fourth, different phenotypes of patients with HF should be taken into account in further model developments.

Conclusion

EHR-driven models, using novel ML algorithms, were developed to predict 1-year in-hospital mortality, use of positive inotropic agents, and 1-year all-cause readmission in patients hospitalized with HF. The discrimination and performance of our models also outperformed the existing tools constructed using traditional techniques. Besides, identifying a greater range of variables can further improve decisions regarding risk assessment for patients with HF.

Acknowledgments

We sincerely appreciate Yidu Cloud (Beijing) Technology Co., Ltd. for providing technical support in data extracting and model developing.

This work was supported by the National Natural Science Foundation of China (No. 81700245, 81900439, and 81970286) and the Natural Science Foundation of Liaoning Province (No. 2019-MS-081).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Features overview of the complete data set.
[DOCX File , 35 KB - jmir_v23i4e24996_app1.docx]

References

1. Chaudhry S, Stewart G. Advanced Heart Failure: Prevalence, Natural History, and Prognosis. Heart Fail Clin 2016 Jul;12(3):323-333. [doi: 10.1016/j.hfc.2016.03.001] [Medline: 27371510]
2. van Riet EES, Hoes A, Wagenaar K, Limburg A, Landman M, Rutten F. Epidemiology of heart failure: the prevalence of heart failure and ventricular dysfunction in older adults over time. A systematic review. Eur J Heart Fail 2016 Mar;18(3):242-252 [FREE Full text] [doi: 10.1002/ejhf.483] [Medline: 26727047]
3. Yu Y, Gupta A, Wu C, Masoudi F, Du X, Zhang J, China PEACE Collaborative Group. Characteristics, Management, and Outcomes of Patients Hospitalized for Heart Failure in China: The China PEACE Retrospective Heart Failure Study. J Am Heart Assoc 2019 Sep 03;8(17):e012884 [FREE Full text] [doi: 10.1161/JAHA.119.012884] [Medline: 31431117]
4. Wolfson A, Fong M, Grazette L, Rahman J, Shavelle D. Chronic heart failure management and remote haemodynamic monitoring. Heart 2018 Dec;104(23):1910-1919. [doi: 10.1136/heartjnl-2018-313397] [Medline: 30121633]
5. Bozkurt B. What Is New in Heart Failure Management in 2017? Update on ACC/AHA Heart Failure Guidelines. Curr Cardiol Rep 2018 Apr 17;20(6):39. [doi: 10.1007/s11886-018-0978-7] [Medline: 29667019]
6. Damen JAAG, Hooft L, Schuit E, Debray TPA, Collins GS, Tzoulaki I, et al. Prediction models for cardiovascular disease risk in the general population: systematic review. BMJ 2016 May 16;353:i2416 [FREE Full text] [doi: 10.1136/bmj.i2416] [Medline: 27184143]

7. Taslimitehrani V, Dong G, Pereira N, Panahiazar M, Pathak J. Developing EHR-driven heart failure risk prediction models using CPXR(Log) with the probabilistic loss function. J Biomed Inform 2016 Apr;60:260-269 [FREE Full text] [doi: 10.1016/j.jbi.2016.01.009] [Medline: 26844760]
8. Li S, Marcus P, Núñez J, Núñez E, Sanchis J, Levy W. Validity of the Seattle Heart Failure Model after heart failure hospitalization. ESC Heart Fail 2019 Jun;6(3):509-515 [FREE Full text] [doi: 10.1002/ehf2.12427] [Medline: 30825268]
9. Stempfle H, Alt A, Stief J, Siebert U. The Munich score: a clinical index to predict survival in ambulatory patients with chronic heart failure in the era of new medical therapies. J Heart Lung Transplant 2008 Feb;27(2):222-228. [doi: 10.1016/j.healun.2007.10.016] [Medline: 18267231]
10. Lee D, Austin P, Rouleau J, Liu P, Naimark D, Tu J. Predicting mortality among patients hospitalized for heart failure: derivation and validation of a clinical model. JAMA 2003 Nov 19;290(19):2581-2587. [doi: 10.1001/jama.290.19.2581] [Medline: 14625335]
11. Austin P, Tu J, Lee D. Logistic regression had superior performance compared with regression trees for predicting in-hospital mortality in patients hospitalized with heart failure. J Clin Epidemiol 2010 Oct;63(10):1145-1155. [doi: 10.1016/j.jclinepi.2009.12.004] [Medline: 20304609]
12. Fonarow G, Adams K, Abraham W, Yancy C, Boscardin W, ADHERE Scientific Advisory Committee, Study Group, and Investigators. Risk stratification for in-hospital mortality in acutely decompensated heart failure: classification and regression tree analysis. JAMA 2005 Feb 02;293(5):572-580. [doi: 10.1001/jama.293.5.572] [Medline: 15687312]
13. O'Connor C, Hasselblad V, Mehta R, Tasissa G, Califf R, Fiuzat M, et al. Triage after hospitalization with advanced heart failure: the ESCAPE (Evaluation Study of Congestive Heart Failure and Pulmonary Artery Catheterization Effectiveness) risk model and discharge score. J Am Coll Cardiol 2010 Mar 02;55(9):872-878 [FREE Full text] [doi: 10.1016/j.jacc.2009.08.083] [Medline: 20185037]
14. Komajda M, Carson P, Hetzel S, McKelvie R, McMurray J, Ptaszynska A, et al. Factors associated with outcome in heart failure with preserved ejection fraction: findings from the Irbesartan in Heart Failure with Preserved Ejection Fraction Study (I-PRESERVE). Circ Heart Fail 2011 Jan;4(1):27-35. [doi: 10.1161/CIRCHEARTFAILURE.109.932996] [Medline: 21068341]
15. Fonarow G, Peacock W, Phillips C, Givertz M, Lopatin M, ADHERE Scientific Advisory Committee and Investigators. Admission B-type natriuretic peptide levels and in-hospital mortality in acute decompensated heart failure. J Am Coll Cardiol 2007 May 15;49(19):1943-1950 [FREE Full text] [doi: 10.1016/j.jacc.2007.02.037] [Medline: 17498579]
16. Rumsfeld JK, Joynt K, Maddox T. Big data analytics to improve cardiovascular care: promise and challenges. Nat Rev Cardiol 2016 Jun;13(6):350-359. [doi: 10.1038/nrcardio.2016.42] [Medline: 27009423]
17. Almog Y, Rai A, Zhang P, Moulaison A, Powell R, Mishra A, et al. Deep Learning With Electronic Health Records for Short-Term Fracture Risk Identification: Crystal Bone Algorithm Development and Validation. J Med Internet Res 2020 Oct 16;22(10):e22550 [FREE Full text] [doi: 10.2196/22550] [Medline: 32956069]
18. Nanayakkara S, Fogarty S, Tremeer M, Ross K, Richards B, Bergmeir C, et al. Characterising risk of in-hospital mortality following cardiac arrest using machine learning: A retrospective international registry study. PLoS Med 2018 Nov;15(11):e1002709 [FREE Full text] [doi: 10.1371/journal.pmed.1002709] [Medline: 30500816]
19. Zhang Y, Allem J, Unger J, Boley Cruz T. Automated Identification of Hookahs (Waterpipes) on Instagram: An Application in Feature Extraction Using Convolutional Neural Network and Support Vector Machine Classification. J Med Internet Res 2018 Nov 21;20(11):e10513 [FREE Full text] [doi: 10.2196/10513] [Medline: 30452385]
20. Frizzell J, Liang L, Schulte P, Yancy C, Heidenreich P, Hernandez A, et al. Prediction of 30-Day All-Cause Readmissions in Patients Hospitalized for Heart Failure: Comparison of Machine Learning and Other Statistical Approaches. JAMA Cardiol 2017 Feb 01;2(2):204-209. [doi: 10.1001/jamacardio.2016.3956] [Medline: 27784047]
21. Bayati M, Braverman M, Gillam M, Mack KM, Ruiz G, Smith MS, et al. Data-driven decisions for reducing readmissions for heart failure: general methodology and case study. PLoS One 2014;9(10):e109264 [FREE Full text] [doi: 10.1371/journal.pone.0109264] [Medline: 25295524]
22. World Health Organization. ICD-10: International Statistical Classification of Diseases and Related Health Problems. Geneva: World Health Organization; 2016.
23. Nakamura M, Kajiwara Y, Otsuka A, Kimura H. LVQ-SMOTE - Learning Vector Quantization based Synthetic Minority Over-sampling Technique for biomedical data. BioData Min 2013 Oct 02;6(1):16 [FREE Full text] [doi: 10.1186/1756-0381-6-16] [Medline: 24088532]
24. Pian W, Khoo C, Chi J. Automatic Classification of Users' Health Information Need Context: Logistic Regression Analysis of Mouse-Click and Eye-Tracker Data. J Med Internet Res 2017 Dec 21;19(12):e424 [FREE Full text] [doi: 10.2196/jmir.8354] [Medline: 29269342]
25. Casanova R, Saldana S, Simpson S, Lacy M, Subauste A, Blackshear C, et al. Prediction of Incident Diabetes in the Jackson Heart Study Using High-Dimensional Machine Learning. PLoS One 2016;11(10):e0163942 [FREE Full text] [doi: 10.1371/journal.pone.0163942] [Medline: 27727289]
26. Gultepe E, Green J, Nguyen H, Adams J, Albertson T, Tagkopoulos I. From vital signs to clinical outcomes for patients with sepsis: a machine learning basis for a clinical decision support system. J Am Med Inform Assoc 2014;21(2):315-325 [FREE Full text] [doi: 10.1136/amiajnl-2013-001815] [Medline: 23959843]

27. van Gerven M, Bohte S. Editorial: Artificial Neural Networks as Models of Neural Information Processing. Front Comput Neurosci 2017;11:114 [FREE Full text] [doi: 10.3389/fncom.2017.00114] [Medline: 29311884]
28. Patel J, Goyal R. Applications of artificial neural networks in medical science. Curr Clin Pharmacol 2007 Sep;2(3):217-226. [doi: 10.2174/157488407781668811] [Medline: 18690868]
29. Klug M, Barash Y, Bechler S, Resheff Y, Tron T, Ironi A, et al. A Gradient Boosting Machine Learning Model for Predicting Early Mortality in the Emergency Department Triage: Devising a Nine-Point Triage Score. J Gen Intern Med 2020 Jan;35(1):220-227 [FREE Full text] [doi: 10.1007/s11606-019-05512-7] [Medline: 31677104]
30. Stojić A, Stanić N, Vuković G, Stanišić S, Perišić M, Šoštarić A, et al. Explainable extreme gradient boosting tree-based prediction of toluene, ethylbenzene and xylene wet deposition. Sci Total Environ 2019 Feb 25;653:140-147. [doi: 10.1016/j.scitotenv.2018.10.368] [Medline: 30408662]
31. Chauhan Z, Samarah M, Unertl K, Jones M. Adoption of Electronic Dental Records: Examining the Influence of Practice Characteristics on Adoption in One State. Appl Clin Inform 2018 Jul;9(3):635-645 [FREE Full text] [doi: 10.1055/s-0038-1667331] [Medline: 30112742]
32. Poirier C, Lavenu A, Bertaud V, Campillo-Gimenez B, Chazard E, Cuggia M, et al. Real Time Influenza Monitoring Using Hospital Big Data in Combination with Machine Learning Methods: Comparison Study. JMIR Public Health Surveill 2018 Dec 21;4(4):e11361 [FREE Full text] [doi: 10.2196/11361] [Medline: 30578212]
33. Shickel B, Tighe P, Bihorac A, Rashidi P. Deep EHR: A Survey of Recent Advances in Deep Learning Techniques for Electronic Health Record (EHR) Analysis. IEEE J Biomed Health Inform 2018 Sep;22(5):1589-1604 [FREE Full text] [doi: 10.1109/JBHI.2017.2767063] [Medline: 29989977]
34. Panahiazar M, Taslimitehrani V, Pereira N, Pathak J. Using EHRs and Machine Learning for Heart Failure Survival Analysis. Stud Health Technol Inform 2015;216:40-44 [FREE Full text] [Medline: 26262006]
35. Janet J, Chan L, Kulik H. Accelerating Chemical Discovery with Machine Learning: Simulated Evolution of Spin Crossover Complexes with an Artificial Neural Network. J Phys Chem Lett 2018 Mar 01;9(5):1064-1071 [FREE Full text] [doi: 10.1021/acs.jpclett.8b00170] [Medline: 29425453]
36. Noble WS. What is a support vector machine? Nat Biotechnol 2006 Dec;24(12):1565-1567. [doi: 10.1038/nbt1206-1565] [Medline: 17160063]
37. Levy W, Aaronson K, Dardas T, Williams P, Haythe J, Mancini D. Prognostic impact of the addition of peak oxygen consumption to the Seattle Heart Failure Model in a transplant referral population. J Heart Lung Transplant 2012 Aug;31(8):817-824. [doi: 10.1016/j.healun.2012.04.006] [Medline: 22759798]
38. May H, Horne B, Levy W, Kfoury A, Rasmusson K, Linker D, et al. Validation of the Seattle Heart Failure Model in a community-based heart failure population and enhancement by adding B-type natriuretic peptide. Am J Cardiol 2007 Aug 15;100(4):697-700. [doi: 10.1016/j.amjcard.2007.03.083] [Medline: 17697831]
39. Lietz K, Long J, Kfoury A, Slaughter M, Silver M, Milano C, et al. Outcomes of left ventricular assist device implantation as destination therapy in the post-REMATCH era: implications for patient selection. Circulation 2007 Jul 31;116(5):497-505. [doi: 10.1161/CIRCULATIONAHA.107.691972] [Medline: 17638928]
40. Senni M, Parrella P, De Maria R, Cottini C, Böhm M, Ponikowski P, et al. Predicting heart failure outcome from cardiac and comorbid conditions: the 3C-HF score. Int J Cardiol 2013 Feb 20;163(2):206-211. [doi: 10.1016/j.ijcard.2011.10.071] [Medline: 22130225]
41. Shameer K, Johnson K, Yahi A, Miotto R, Li L, Ricks D, et al. Predictive Modeling of Hospital Readmission Rates Using Electronic Medical Record-Wide Machine Learning: A Case-Study Using Mount Sinai Heart Failure Cohort. In: Proceedings of the Pacific Symposium on Biocomputing. 20161617 Presented at: Pacific Symposium on Biocomputing; January 4-8, 2016; Big Island, HI, USA p. 276-287 URL: http://psb.stanford.edu/psb-online/proceedings/psb17/abstracts/2017_p276.html [doi: 10.1142/9789813207813_0027]
42. Golas S, Shibahara T, Agboola S, Otaki H, Sato J, Nakae T, et al. A machine learning model to predict the risk of 30-day readmissions in patients with heart failure: a retrospective analysis of electronic medical records data. BMC Med Inform Decis Mak 2018 Jun 22;18(1):44 [FREE Full text] [doi: 10.1186/s12911-018-0620-z] [Medline: 29929496]
43. Kakarmath S, Golas S, Felsted J, Kvedar J, Jethwani K, Agboola S. Validating a Machine Learning Algorithm to Predict 30-Day Re-Admissions in Patients With Heart Failure: Protocol for a Prospective Cohort Study. JMIR Res Protoc 2018 Sep 04;7(9):e176 [FREE Full text] [doi: 10.2196/resprot.9466] [Medline: 30181113]
44. Ural D, Çavuşoğlu Y, Eren M, Karaüzüm K, Temizhan A, Yılmaz MB, et al. Diagnosis and management of acute heart failure. Anatol J Cardiol 2015 Nov;15(11):860-889 [FREE Full text] [doi: 10.5152/AnatolJCardiol.2015.6567] [Medline: 26574757]
45. Aljundi A, Mohammed S, Patel A, Singh R, Arabi A, AlBinali H, et al. Inotropic agents use in patients hospitalized with acute decompensated heart failure: a retrospective analysis from a 22-year registry in a Middle-Eastern Country (1991-2013). BMC Cardiovasc Disord 2016 Feb 19;16:47 [FREE Full text] [doi: 10.1186/s12872-016-0223-5] [Medline: 26892533]
46. Cho J, Cho H, Lee H, Ki Y, Jeon E, Hwang K, et al. Neutrophil-Lymphocyte Ratio in Patients with Acute Heart Failure Predicts In-Hospital and Long-Term Mortality. J Clin Med 2020 Feb 18;9(2):557 [FREE Full text] [doi: 10.3390/jcm9020557] [Medline: 32085596]

47. Ommen S, Hammill S, Gibbons R. The relative lymphocyte count predicts death in patients receiving implantable cardioverter defibrillators. Pacing Clin Electrophysiol 2002 Oct;25(10):1424-1428. [doi: 10.1046/j.1460-9592.2002.01424.x] [Medline: 12418738]
48. Seo M, Yamada T, Tamaki S, Hikoso S, Yasumura Y, Higuchi Y, et al. Prognostic Significance of Serum Cholinesterase Level in Patients With Acute Decompensated Heart Failure With Preserved Ejection Fraction: Insights From the PURSUIT-HFpEF Registry. J Am Heart Assoc 2020 Jan 07;9(1):e014100 [FREE Full text] [doi: 10.1161/JAHA.119.014100] [Medline: 31847660]
49. Vazquez R, Bayes-Genis A, Cygankiewicz I, Pascual-Figal D, Grigorian-Shamagian L, Pavon R, MUSIC Investigators. The MUSIC Risk score: a simple method for predicting mortality in ambulatory patients with chronic heart failure. Eur Heart J 2009 May;30(9):1088-1096. [doi: 10.1093/eurheartj/ehp032] [Medline: 19240065]

Abbreviations

ACEI: angiotensin-converting enzyme inhibitor
ADHERE: Acute Decompensated Heart Failure National Registry
ADHF: acute decompensated heart failure
ANN: artificial neural network
ARB: aldosterone receptor antagonist
AST: aspartate aminotransferase
AUC: area under the curve
BNP: B-type natriuretic peptide
BUN: blood urea nitrogen
EFFECT: Enhanced Feedback for Effective Cardiac Treatment
EHR: electronic health record
HF: heart failure
hs-cTnI: high-sensitivity cardiac troponin I
ICD-10: International Classification of Diseases, 10th Revision
INR: international normalized ratio
LR: logistic regression
ML: machine learning
MLP: multilayer perceptron
REMATCH: Randomized Evaluation of Mechanical Assistance for the Treatment of Congestive Heart Failure
RF: random forest
ROC: receiver operating characteristic
SBP: systolic blood pressure
SHAP: Shapley additive explanation
SHFM: Seattle Heart Failure Model
SVM: support vector machine
XGBoost: extreme gradient boosting

Edited by R Kukafka; submitted 14.10.20; peer-reviewed by CR Telles, H Mufti; comments to author 07.11.20; revised version received 04.01.21; accepted 16.03.21; published 19.04.21.

Please cite as:
Lv H, Yang X, Wang B, Wang S, Du X, Tan Q, Hao Z, Liu Y, Yan J, Xia Y
Machine Learning–Driven Models to Predict Prognostic Outcomes in Patients Hospitalized With Heart Failure Using Electronic Health Records: Retrospective Study
J Med Internet Res 2021;23(4):e24996
URL: https://www.jmir.org/2021/4/e24996
doi:10.2196/24996
PMID:33871375

Original Paper

Multimodal Recruitment to Study Ovulation and Menstruation Health: Internet-Based Survey Pilot Study

Shruthi Mahalingaiah[1,2,3*], MD, MS; J Jojo Cheng[2,4*], MS; Michael R Winter[5*], MPH; Erika Rodriguez[1,2*], MS; Victoria Fruh[1*], PhD; Anna Williams[2*], MS; MyMy Nguyen[2*], MS; Rashmi Madhavan[2*], MS; Pascaline Karanja[2*], MS; Jill MacRae[2*], MS; Sai Charan Konanki[1*], MS; Kevin J Lane[6*], PhD; Ann Aschengrau[3*], SCD

[1]Department of Environmental Health, Harvard T.H. Chan School of Public Health, Boston, MA, United States

[2]Department of Obstetrics and Gynecology, Boston University School of Medicine, Boston, MA, United States

[3]Department of Epidemiology, Boston University School of Public Health, Boston, MA, United States

[4]Department of Biostatistics and Medical Informatics, University of Wisconsin, Madison, WI, United States

[5]Biostatistics and Epidemiology Data Analytics Center, Boston University School of Public Health, Boston, MA, United States

[6]Department of Environmental Health, Boston University School of Public Health, Boston, MA, United States

[*]all authors contributed equally

Corresponding Author:
Shruthi Mahalingaiah, MD, MS
Department of Environmental Health
Harvard T.H. Chan School of Public Health
665 Huntington Avenue
Building 1, Floor 14
Boston, MA, 02115
United States
Phone: 1 6174324381
Email: shruthi@hsph.harvard.edu

Abstract

Background: Multimodal recruitment strategies are a novel way to increase diversity in research populations. However, these methods have not been previously applied to understanding the prevalence of menstrual disorders such as polycystic ovary syndrome.

Objective: The purpose of this study was to test the feasibility of recruiting a diverse cohort to complete a web-based survey on ovulation and menstruation health.

Methods: We conducted the Ovulation and Menstruation Health Pilot Study using a REDCap web-based survey platform. We recruited 200 women from a clinical population, a community fair, and the internet.

Results: We recruited 438 women over 29 weeks between September 2017 and March 2018. After consent and eligibility determination, 345 enrolled, 278 started (clinic: n=43; community fair: n=61; internet: n=174), and 247 completed (clinic: n=28; community fair: n=60; internet: n=159) the survey. Among all participants, the median age was 25.0 (SD 6.0) years, mean BMI was 26.1 kg/m^2 (SD 6.6), 79.7% (216/271) had a college degree or higher, and 14.6% (37/254) reported a physician diagnosis of polycystic ovary syndrome. Race and ethnicity distributions were 64.7% (176/272) White, 11.8% (32/272) Black/African American, 7.7% (21/272) Latina/Hispanic, and 5.9% (16/272) Asian individuals; 9.9% (27/272) reported more than one race or ethnicity. The highest enrollment of Black/African American individuals was in clinic (17/42, 40.5%) compared to 1.6% (1/61) in the community fair and 8.3% (14/169) using the internet. Survey completion rates were highest among those who were recruited from the internet (159/174, 91.4%) and community fairs (60/61, 98.4%) compared to those recruited in clinic (28/43, 65.1%).

Conclusions: Multimodal recruitment achieved target recruitment in a short time period and established a racially diverse cohort to study ovulation and menstruation health. There were greater enrollment and completion rates among those recruited via the internet and community fair.

(J Med Internet Res 2021;23(4):e24716) doi:10.2196/24716

KEYWORDS

polycystic ovary syndrome; PCOS; menstrual cycle; multimodal recruitment strategy; epidemiology; recruitment; pilot; strategy; women; feasibility; online survey; ovulation; menstrual

Introduction

Polycystic ovary syndrome, initially described in 1935 [1], is now considered the most common endocrine disorder in reproductive-aged women [2]. Affecting 10% to 15% [2], prevalence estimates range from as low as 8% [3] to as high as 26% [4], depending on the definition used and population studied. The disorder is characterized by clinical or biochemical androgen excess, menstrual irregularity, and the presence of polycystic ovarian morphology on ultrasound visualization [5]. Clinical androgen excess typically presents as acne, hirsutism, or androgenic alopecia [6-8]. Women with polycystic ovary syndrome may also experience infertility, insulin resistance, and obesity [9-12].

Existing research on polycystic ovary syndrome is typically conducted in either small clinical cohorts or larger epidemiologic studies. The latter have variable ascertainment of the disease or disease features which may increase misclassification of disease state and bias estimates of risk [13]. Preliminary studies on polycystic ovary syndrome in existing population-based cohorts such as the Nurses' Health Study 2, the Framingham Heart Study, and the Cape Cod Health Study were limited by poor correlation of polycystic ovary syndrome self-reports compared to medical records [14], too few reported cases of menstrual irregularity [15], and inadequate determination of phenotype [16], respectively. Inaccurate classification using single yes or no questions to identify women with polycystic ovary syndrome may bias estimates of the disease.

Existing epidemiologic cohorts were also limited with respect to race and ethnic diversity; more than 90% of participants were White in these 3 cohorts [16-18], thereby limiting the ability to detect differences in prevalence and etiological associations across racial and ethnic groups. Racial and ethnic-specific differences in androgen excess and metabolic syndrome symptoms are well established but the reasons for this are not well understood [19-21].

The Ovulation and Menstruation Health Pilot Study was conducted to (1) determine the feasibility of enrolling participants from diverse backgrounds using varied recruitment modalities, and (2) understand how survey completion status differed by participant characteristics.

Methods

Design

The Ovulation and Menstruation Health study website consists of a short animated educational recruitment video, a web-based consent form, screening questions, and a survey instrument. The animated video was designed to appeal to a diverse audience, as its illustrations included women of all races and body types. The goal of the Ovulation and Menstruation Health Pilot Study was to recruit at least 200 reproductive-aged women over a 1-year period. Those who were menstruating or had the capacity to menstruate were eligible (trans-male and other). Individuals who were <18 years or >45 years of age; identified as male; were pregnant at the time of the survey; had a hysterectomy/oophorectomy, amenorrhea due to radiation, or chemotherapy; or were unwilling or unable to provide an email address were deemed ineligible. The survey instrument was publicly accessible from August 21, 2017 to February 26, 2018 on the study website and was used in each recruitment modality. The Boston Medical Center and Boston University Medical Campus Institutional Review Boards approved the study (IRB H-35075).

Incentive

The first 200 participants who completed the survey were entered into a lottery for a US $200 gift card. All who were approached for recruitment were offered free earphones.

Recruitment

The multimodal recruitment locations were in clinic, a Boston-based community event, and the internet. The recruitment approach was adapted to meet the needs of each recruitment location.

In-Clinic

In-clinic recruitment occurred from September to October 2017. Approximately 2000 informational recruitment letters including the study website were sent during this recruitment period to patients with an upcoming visit to the Department of Obstetrics and Gynecology (OB/GYN) at Boston Medical Center (BMC). For 20 days, research assistants showed the study website and the promotional video via an electronic tablet to interested persons at 2 OB/GYN waiting rooms.

Community Event

A recruitment table was set up at the Boston Women's Market in the Jamaica Plain neighborhood of Boston on September 17, 2017. This community event was an opportunity for those who identify as female or supporters of women's causes to sell products from their small businesses.

Internet

Internet-based recruitment methods included sending out email communications, creating study social media engagement accounts, and sharing recruitment materials on individual social networks of the study staff. The study website and social media pages were discoverable on any internet search. Boston University Medical Campus (BUMC)–wide electronic communications, which included a brief description of the study, eligibility requirements, incentive, and study contact information, began in September 2017 and ran for the duration of the enrollment period. The study website link was shared on personal networks of study personnel (RM, MN, AW) with the Boston University Masters of Medical Science Class of 2018 and the Bates Feminist Collective Facebook groups in September 2017.

Physical flyers and business cards (paper materials) were posted around the BUMC campus for the duration of the study, and at Boston Skin Solutions, a center that treats excess facial and body hair on September 27, 2017.

Informed Consent, Screening, and Survey Initiation

The web-based consent form included questions on a participant's interest in follow-up surveys, other studies, and contributing biospecimens. Once consented, the individual was presented with an eligibility screener. Eligible participants recruited in clinic were able to start the survey in the waiting room. For eligible participants from the community event and the internet, a unique link to the survey was emailed to the participant's email address. For all enrolled participants, up to 3 reminders were sent on consecutive days to those who did not complete the survey.

Survey Description

The Ovulation and Menstruation Health Pilot Study survey had a total of 10 sections, with 219 questions; however, most respondents had fewer questions due to skip patterns. The About You section included questions on current residence and whether they ever received care at BMC. The Demographics section included questions on race and ethnicity, education, birth country, and income. The Anthropometrics section captured data regarding height, weight, and body shape using a pictorial tool. The Menstrual Cycle section asked about menarche, regularity of cycles, and cycle tracking. The Hormone Usage section obtained data on hormonal contraceptive usage, history and reasons for hormone usage. The Health and Body section included questions for self-reported androgen excess using a pictorial tool for androgenic alopecia and hirsutism based on the modified Ferriman-Gallwey scale [22]; polycystic ovary syndrome status (questions specific to polycystic ovary syndrome diagnosis, symptoms for diagnosis, and medication and supplement use, and family history of polycystic ovary syndrome); reproductive health (history of infertility, uterine fibroids, endometriosis, and premature ovarian failure); general health (diagnosed medical conditions such as hypertension, diabetes, and other chronic conditions and their treatment); diet and lifestyle (limited questions about alcoholic, nonalcoholic beverage consumption, and smoking habits); obstetric history (ever pregnant, number of prior pregnancies, and each pregnancy outcome, including stillbirths and miscarriages, and live births). For every livebirth, an additional question was asked regarding gestational length and birth outcomes such as birth weight. Study data were recorded and stored on encrypted servers at the Boston University School of Public Health Biostatistics & Epidemiology Data Analytics Center. The Ovulation and Menstruation Health Pilot Study survey can be found on the Harvard Dataverse [23].

Data Analysis

Participant characteristics were evaluated by recruitment location and survey completion. Those who started the survey via the internet included those who input the link into their web-browser after learning of the study from any web-based or printed recruitment materials and completed the web-based consent and screener. Survey initiation was defined as anyone who started the "About You" section. Survey completion was defined as having completed the first question of the "Pregnancy and Birth History" upon entering the total number of pregnancies experienced. Those who entered 0 pregnancies had no further questions to complete. Those who entered 1 or more had additional pregnancy related questions to complete. For both nulliparous and multiparous responses, completing the survey was as entering the total number of pregnancies experienced. To assess the geographical catchment area associated with the survey, participants' self-reported home states were mapped.

Results

Recruitment and Enrollment

Recruitment exceeded our target of 200 women in 8 weeks. During the 6-month recruitment window, 438 women began the consent process, 384 were assessed for eligibility, and 345 women were screened and included (88.7% of those assessed for eligibility). Of the 345, 278 women started the survey, of whom 66 were recruited in clinic, 61 were recruited from the community event, and 174 were recruited from the internet. Figure 1 displays the study enrollment flow chart. Spikes in cumulative enrollment coincided with the active in-clinic recruitment and the community event (Figure 2). Out of the 384 participants who consented and completed screening, only 39 (10.2%) were excluded. Of the individuals who were excluded, 18 were no longer menstruating, 16 were unwilling or unable to provide email addresses, and 5 were over the age of 45 years. Among those who initiated the survey and responded to re-contact questions, agreement for re-contact was high—96.0% (266/277) consented to be re-contacted for additional information, 81.2% (225/277) for biological samples, 96.4% (267/277) for a follow-up survey, and 79.4% (220/277) for a different research study.

Figure 1. Flowchart of participants enrolled in the Ovulation and Menstruation Health Pilot Study.

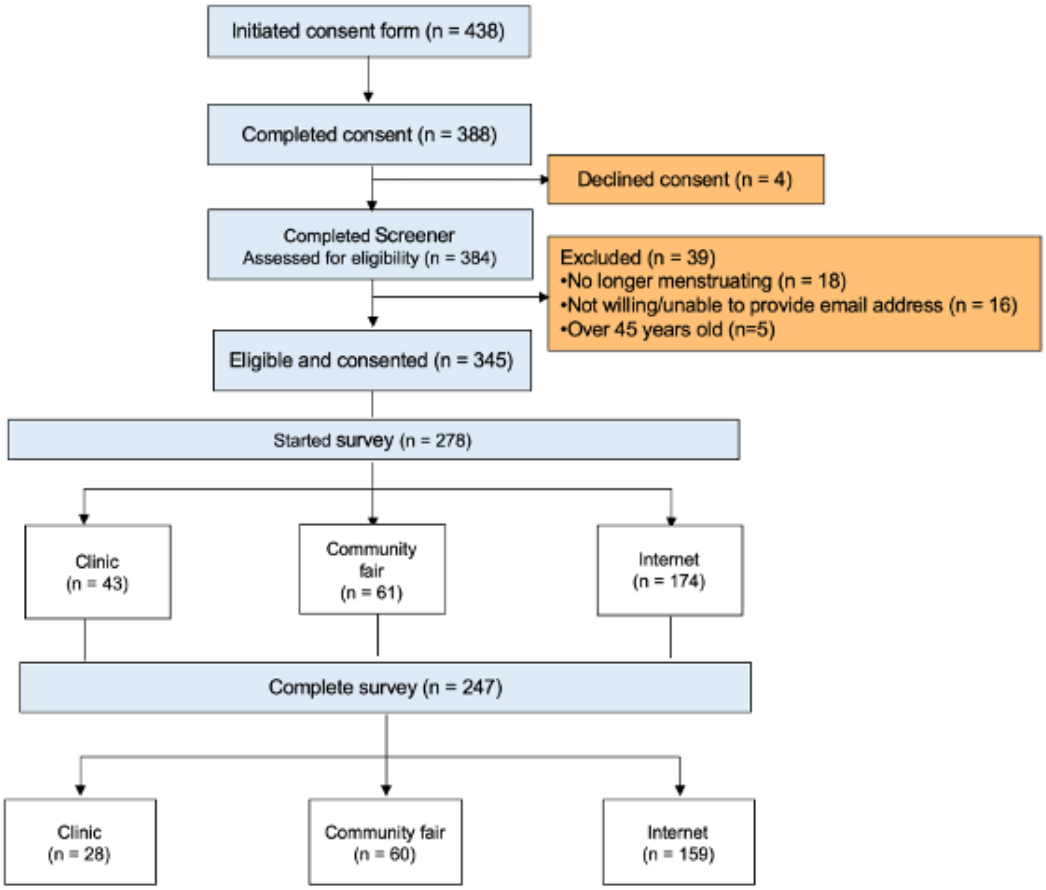
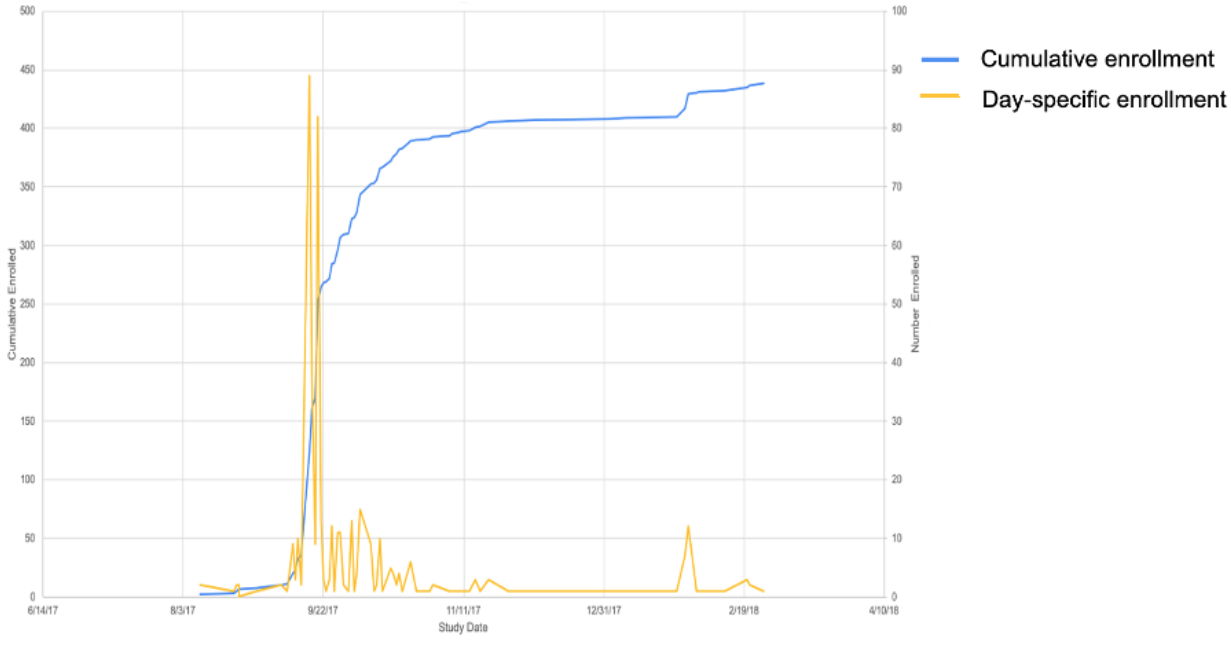


Figure 2. Ovulation and Menstruation Health Pilot Study enrollment graph.



Participant Characteristics

Among the 278 women who started the survey, the average age was 27.3 years and the median age was 25.0 (SD 6.0), with the majority (224/273, 82.1%) born in the United States (Table 1). Participants were White (176/272, 64.7%), Black/African American (32/272, 11.8%), Latina/Hispanic (21/272, 7.7%), and mixed race (27/272, 9.9%). Two individuals identified their

gender as other (nonbinary or gender queer). While the majority (216/271, 79.7%) had a 4-year college degree or more, 7.0% (19/271) had a high school degree or less education. Participants were distributed across all income categories, with 21.3% (58/272) in the less than $25,000 category. The mean BMI was 26.1 kg/m^2 (SD 6.6), with 15.4% (41/266) in the overweight category; 13.1% (32/245) had smoked at least 100 cigarettes over their lifetime. The reported prevalence of doctor-diagnosed polycystic ovary syndrome was 14.6% (37/254), and 82.3% (218/265) had ever used hormonal contraception. The geographic distribution of the cohort included 20 states (Figure 3). The majority of the participants were recruited from Massachusetts (n=173) and Missouri (n=50).

Figure 3. Ovulation and Menstruation Health Study participant population map.

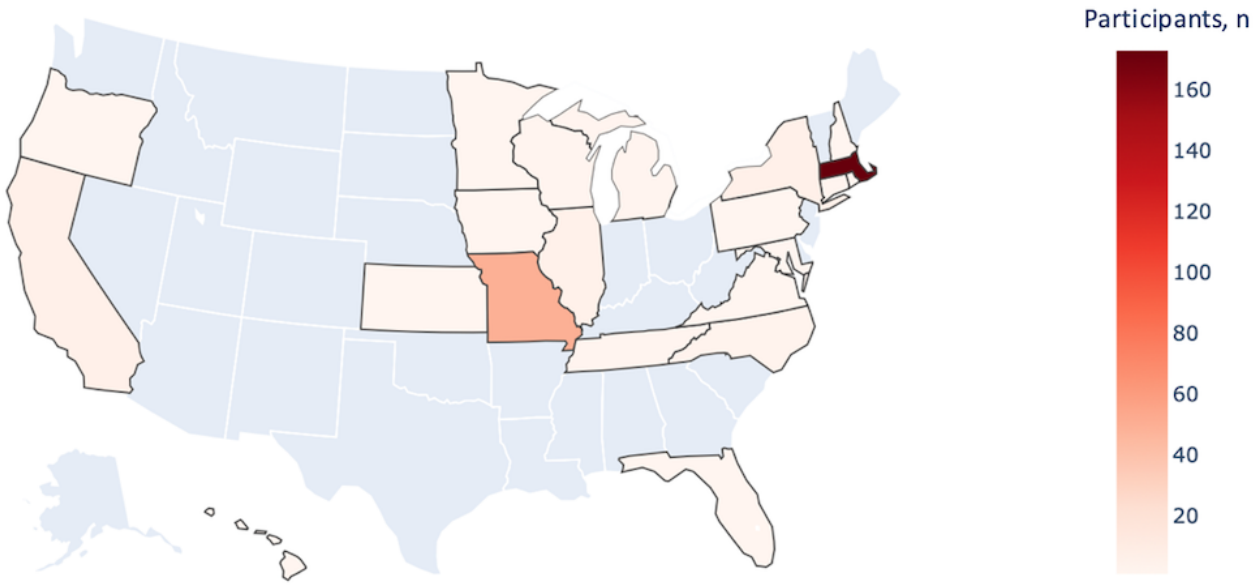


Table 1. Demographics in the Ovulation and Menstruation pilot study.

Characteristic	Value (n=278[a])
Age (years), median (SD)	25.0 (6.0)
Born in the United States, n (%)	224 (82.1)
Race, n (%)	
White	176 (64.7)
Latina/Hispanic	21 (7.7)
Black/African American	32 (11.8)
Asian	16 (5.9)
More than one race	27 (9.9)
Educational attainment, n (%)	
High school graduate/GED or less education	19 (7.0)
Some college or 2-year degree	36 (13.3)
4-year college graduate	101 (37.3)
More than 4-year college degree	115 (42.4)
Total annual household income, n (%)	
Below $25,000	58 (21.3)
$25,000 to $49,999	70 (25.7)
$50,000 to $74,999	44 (16.2)
$75,000 to $99,999	20 (7.4)
$100,000 or more	43 (15.8)
Prefer not to answer	17 (6.3)
Don't know	20 (7.4)
Smoked at least 100 cigarettes over lifetime, n (%)	32 (13.1)
BMI[b], n (%)	
Underweight (<18.5 kg/m^2)	8 (3.0)
Normal weight (18.5-24.9 kg/m^2)	153 (57.5)
Overweight (25.0-29.9 kg/m^2)	41 (15.4)
Obese (≥30.0 kg/m^2)	64 (24.1)
Polycystic ovary syndrome diagnosis by doctor, n (%)	37 (14.6)
Ever pregnant, n (%)	42 (17.0)
Gravidity[c], n (%)	
1 pregnancy	17 (40.5)
2 pregnancies	10 (23.8)
3 pregnancies	7 (16.7)
>4 pregnancies	8 (19.1)
Hormonal contraceptives ever use, n (%)	218 (82.3)
Self-rated current health, n (%)	
Excellent	37 (14.1)
Very good	108 (41.2)
Good	91 (34.7)
Fair	25 (9.5)
Poor	1 (0.4)

Characteristic	Value (n=278[a])
Survey completed, n (%)	247 (88.9)

[a]Missing—born in the United States n=5 (1.8%); race: n=6 (2.2%); education: n=7 (2.5%); income: n=6 (2.2%); smoking status: n=33 (11.9%); BMI: n=12 (4.3%); polycystic ovary syndrome diagnosis: n=24 (8.6%); gravidity: n=26 (9.4%); ever pregnant: n=31 (11.2%); hormonal contraceptive: n=13 (4.7%); self-rated current health: n=16 (5.8%).

[b]BMI: body mass index.

[c]Gravidity among women who were ever pregnant (n=42).

Population Demographic Characteristics by Recruitment Location

Differences were found in age, birth country, race, educational attainment, BMI, and physician-diagnosed polycystic ovary syndrome, by location of recruitment (Table 2). Women recruited in clinic tended to be older (median 34.0 years, SD 6.7), born outside the US (26/42, 61.9%), and with lower educational attainment, 19.0% (8/42) were high school graduate/GED or less education compared to women recruited at the community fair and via the internet. Women recruited in clinic were also more likely to be obese (18/38, 47.4%), have had a physician diagnosis of polycystic ovary syndrome (12/33, 36.4%), and less likely to have ever used oral contraceptives (23/37, 62.2%) compared to women recruited at the community fair and via the internet. Those from the community fair and internet were more likely to be born in the US (58/61, 95.1% and 150/170, 88.2%, respectively), White (47/61, 77.0% and 121/169, 71.6%, respectively), and have greater than a 4-year college degree (24/61, 39.3% and 83/168, 49.4%, respectively) than those recruited from the clinic. Women recruited from outside the clinic were younger (community fair: median 23.0, SD 4.0; internet: median 25.0, SD 5.4 years), leaner BMI (community fair: 23.3 kg/m^2 SD 4.8; internet: 23.5 kg/m^2 SD 7.0) and had a lower prevalence of physician diagnosed polycystic ovary syndrome (community fair: 3/61, 4.9%; internet: 22/160, 13.8%). Of note, 46.4% of women recruited in clinic (13/28) had been pregnant compared to 3.3% (2/61) and 17.0% (27/159) for the community fair and internet participants, respectively. Income and smoking status were similar among recruitment locations.

Table 2. Demographic characteristics by recruitment location.

Characteristic	Location		
	Clinic (n=43)[a]	Community fair (n=61)[b]	Internet (n=174)[c]
Age (years), median (SD)	34.0 (6.7)	23.0 (4.0)	25.0 (5.4)
Born in the United States, n (%)	16 (38.1)	58 (95.1)	150 (88.2)
Race, n (%)			
White	8 (19.0)	47 (77.0)	121 (71.6)
Latina/Hispanic	12 (28.6)	2 (3.3)	7 (4.1)
Black/African American	17 (40.5)	1 (1.6)	14 (8.3)
Asian	3 (7.1)	1 (1.6)	12 (7.1)
More than one race	2 (4.8)	10 (16.4)	15 (8.9)
Educational attainment, n (%)			
High school graduate/GED or less education	8 (19.0)	6 (9.8)	5 (3.0)
Some college or 2-year degree	14 (33.3)	8 (13.1)	14 (8.3)
4-year college graduate	12 (28.6)	23 (37.7)	66 (39.3)
More than 4-year college degree	8 (19.1)	24 (39.3)	83 (49.4)
Total annual household income, n (%)			
Below $25,000	4 (9.5)	13 (21.3)	41 (24.3)
$25,000 to $49,999	12 (28.6)	20 (32.8)	38 (22.5)
$50,000 to $74,999	11 (26.2)	9 (14.8)	24 (14.2)
$75,000 to $99,999	1 (2.4)	3 (4.9)	16 (9.5)
$100,000 or more	5 (11.9)	6 (9.8)	32 (18.9)
Prefer not to answer	8 (19.1)	2 (3.3)	7 (4.1)
Don't know	1 (2.4)	8 (13.1)	11 (6.5)
Smoked at least 100 cigarettes over lifetime, n (%)	5 (17.9)	5 (8.3)	22 (14.0)
BMI[d], n (%)			
Normal weight or underweight (≤25 kg/m^2)	14 (36.8)	40 (65.6)	107 (64.1)
Overweight (25-30 kg/m^2)	6 (15.8)	11 (18.0)	24 (14.4)
Obese (≥30 kg/m^2)	18 (47.4)	10 (16.4)	36 (21.6)
Polycystic ovary syndrome diagnosis by doctor, n (%)	12 (36.4)	3 (4.9)	22 (13.8)
Ever pregnant, n (%)	13 (46.4)	2 (3.3)	27 (17.0)
Gravidity[e], n (%)			
1 pregnancy	5 (38.5)	1 (50.0)	11 (40.7)
2 or more pregnancies	8 (61.5)	1 (50.0)	16 (59.3)
Hormonal contraceptives ever use, n (%)	23 (62.2)	54 (88.5)	141 (84.4)
Self-rated current health, n (%)			
Excellent	6 (16.2)	8 (13.1)	23 (14.0)
Very good	7 (18.9)	29 (47.5)	72 (43.9)
Good	18 (48.7)	21 (34.4)	52 (31.7)
Fair to poor	6 (16.2)	3 (4.9)	17 (10.4)
Survey completed, n (%)	28 (65.1)	60 (98.4)	159 (91.4)

[a]Missing—born in the US: n=1 (2.3%); race: n=1 (2.3%); education: n=1 (2.3%); income: n=1 (2.3%); smoking status: n=15 (34.9%); BMI: n=5 (11.6%); polycystic ovary syndrome diagnosis: n=10 (23.3%); ever pregnant: n=15 (34.9%); hormonal contraceptive: n=6 (14.0%); self-rated current health: n=6

(14.0%).

[b]Missing—smoking status: n=1 (1.6%).

[c]Missing—born in the US: n=4 (2.3%); race: n=5 (2.9%); education: n=6 (3.5%); income: n=5 (2.9%); smoking status n=17 (9.8%); BMI: n=7 (4.0%); polycystic ovary syndrome diagnosis: n=14 (8.1%); ever pregnant: n=15 (8.6%); hormonal contraceptive: n=7 (4.0%); self-rated current health: n=10 (5.7%).

[d]BMI: body mass index.

[e]Gravidity among women who were ever pregnant (n=42).

Survey Completion

Among the 278 women who started the survey, 247 (88.8%) completed it. The median time to completion was 12.6 minutes (range 1.03 minutes to 92 days); 7 participants took more than 24 hours to complete the survey. Women completing the survey were more likely to be US born (210/246, 85.4% vs 14/27, 51.9%), White (165/245, 67.3% vs 11/27, 40.7%) and have 4 years of college education or more (200/244, 81.9% vs 16/27, 59.2%) compared to those who did not complete the survey (Table 3).

We also found that the women who did not complete the survey were older and had higher annual household income compared to those who completed the survey. Among the 31 women who did not complete the survey, dropout seemed equally distributed across section categories (Figure 4).

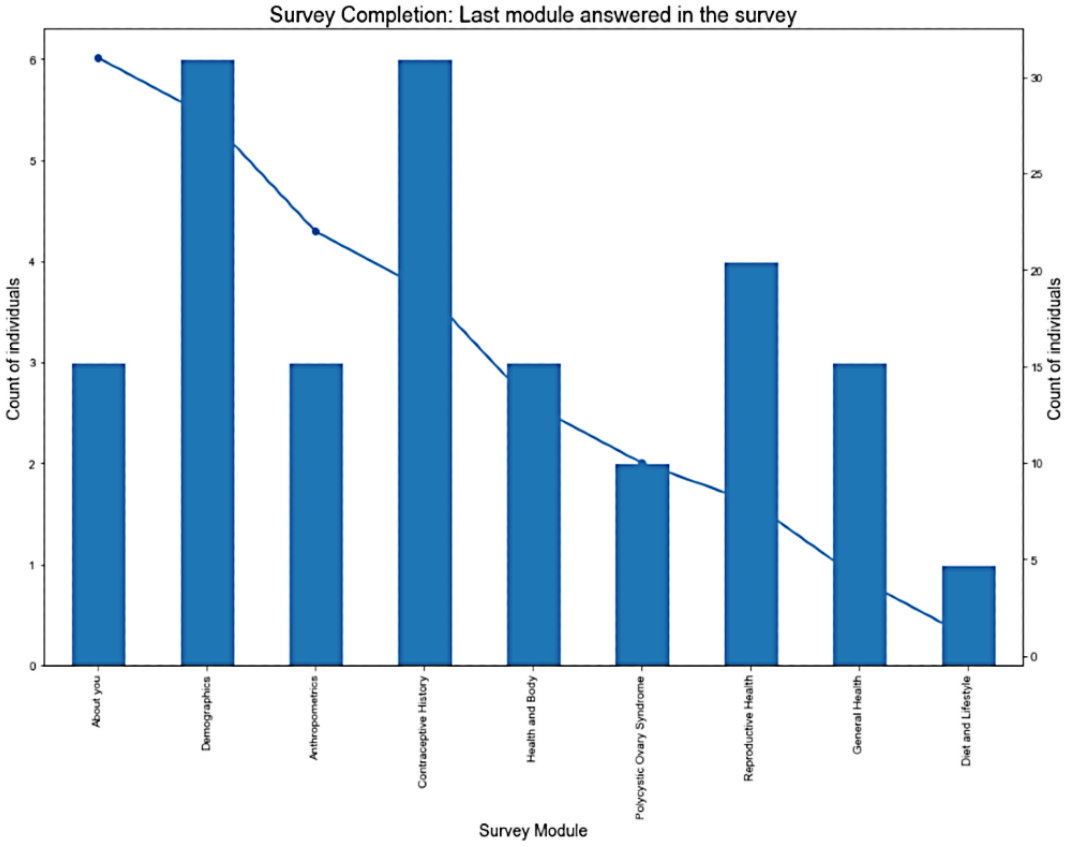
Table 3. Demographic characteristics by survey completion status (N=278).

Characteristic	Complete (n=247)[a]	Incomplete (n=31)[b]
Age (years), median (SD)	25.0 (5.7)	29.0 (7.4)
Born in the United States, n (%)	210 (85.7)	14 (51.9)
Race, n (%)		
White	165 (67.4)	11 (40.7)
Latina/Hispanic	15 (6.1)	6 (22.2)
Black/African American	25 (10.2)	7 (25.9)
Asian	15 (6.1)	1 (3.7)
More than one race	25 (10.2)	2 (7.4)
Educational attainment, n (%)		
High school graduate/GED or less education	14 (5.7)	5 (18.5)
Some college or 2-year degree	30 (12.3)	6 (22.2)
4-year college graduate	94 (38.5)	7 (25.9)
More than 4-year college degree	106 (43.4)	9 (33.3)
Total annual household income, n (%)		
Below $25,000	54 (22.0)	4 (14.8)
$25,000 to $49,999	63 (25.7)	7 (25.9)
$50,000 to $74,999	42 (17.1)	2 (7.4)
$75,000 to $99,999	19 (7.8)	1 (3.7)
$100,000 or more	39 (15.9)	4 (14.8)
Prefer not to answer	11 (4.5)	6 (22.2)
Don't know	17 (4.5)	3 (11.1)

[a]Missing—born in the US: n=1 (0.4%); race: n=2 (0.8%); education: n=3 (1.2%); income: n=2 (0.8%).

[b]Missing—born in the US: n=4 (12.9%); race: n=4 (12.9%); education: n=4 (12.9%); income: n=4 (12.9%).

Figure 4. Survey completion by last module answered in the survey.



Discussion

We determined the feasibility of a multimodal recruitment approach for a study of ovulation and menstruation health by comparing in-clinic, community-based, and internet-based recruitment activities. Among the recruitment modalities, the majority of women were recruited via the internet after encountering advertising materials informing them of the study website. The race and ethnicity of the cohort was notably similar to the US population [24]. Based on census data from 2018, the US population was comprised of 166,038,755 women which was 50.8% of the total population. Of those women, 60.3% were White, 17.8% were Latina or Hispanic, 12.9% were Black or African American, 5.9% were Asian, and 2.2% identified as being of more than one race [24].

The participant characteristics did differ by recruitment location, suggesting that multimodal recruitment is a feasible solution to increase the variation in participant characteristics including race, ethnicity, and health characteristics such as BMI, oral contraceptive use, and whether physician diagnosed. These different characteristics may reflect differences in the source population. For example, approximately 59% of the patients at BMC come from an underserved population [25], whereas the community event was held in Jamaica Plain, a neighborhood that has a 53.6% White population and a median household annual income of $55,861 [26]. While use of the internet usage demographics are reported as 92% of White, 86% of Hispanic, and 85% of Black individuals in the US [27], those accessing the study via the internet were recruited through flyers, email communications, and personal social media networks. The geographic distribution of the participants was notable for the majority in Massachusetts, with second-highest participation in Missouri, corresponding to the personal network of 1 research assistant.

Similarly, another study [28] was also successful in recruiting racially and ethnically diverse participants using a combination of 340 recruitment sites and digital campaigns. Among the core participants of that study [28], 51% were considered non-White, whereas participants in traditional reproductive health cohorts have been 83% Caucasian [29] and 89% White [14]. Furthermore, while participants from the original Framingham Heart Study were 100% White, racial diversity increased in subsequent studies with inclusion of 28% African American, 42% Hispanic American, and 24% Asian American participants within the Framingham-related Omni Cohort [30].

The completion rate among those who started our survey (247/278, 88.8%) was high. This completion rate is comparable to those of the Pregnancy Study Online [29], with 72% of enrolled participants completing the baseline survey, and the most recent Framingham Heart Study [31], with participants using an electronic survey (85% completion). It is possible that our survey was shorter than those used in those studies [29,31] and as such had a slightly higher completion rate. Furthermore, the Ovulation and Menstruation Health survey was designed for an 8th grade reading level and underwent cognitive and usability testing to facilitate question comprehension [32], potentially supporting a pleasant participant experience. While participants could skip any question, the rate of missing responses was low. Birthplace outside of the US was a key

variable noted to be missing among a small proportion of women who did not complete the survey.

This study has several strengths. We achieved target recruitment in half the expected time, which may have been due to the appealing web-based study platform that included an educational cartoon featuring women from multiple backgrounds or the monetary incentive. Use of a cognitively tested survey designed for an 8th grade reading level may also have facilitated the high completion rate. The broad inclusion criteria also allowed for those with and without polycystic ovary syndrome to be included. Most importantly, the multimodal recruitment approach enabled us to recruit a more diverse cohort, which is comparable to those in other studies using a similar multimodal recruitment strategy for other health conditions [33].

However, the study also has some key limitations. While racial and ethnic diversity were improved compared to existing large epidemiologic cohorts, the proportion of Latina or Hispanic women was slightly lower than that in the US population, possible due to language access issues. We recently translated the Ovulation and Menstruation Health Platform for availability in Spanish language to test in future studies. Also, completion rate by those born outside the US may be improved by having the survey in the primary language of the participant. In the future, using paid advertisements may not only lead to high click-through rates to the study website, but may also provide for a unique opportunity to engage with specific targeted demographics depending on advertisement dissemination. For example, the Nurses' Health Study 3 [17] sent recruitment postcards to minority-dense zip codes doubling the enrollment rate of African American and Hispanic women.

Furthermore, we considered whether bias in multimodal recruitment exists. We do not believe that multimodal recruitment increased the likelihood of selection bias since enrollment would need to be related to both the exposure and outcome under study for such a bias to occur [34-37]. The multimodal recruitment strategy increases the generalizability of the study and increases its ability to examine effect measure modification by racial and ethnic group by ensuring sufficient numbers in each category. Because differences exist in racial or ethnic-specific risks for a variety of health outcomes potentially due to structural racism and other inequities [38], a multiethnic cohort is needed to assess health outcomes subsequent to polycystic ovary syndrome diagnosis and to determine ideal windows for risk-reducing interventions.

Multimodal recruitment was feasible and established a more racially and ethnically diverse cohort for the study of ovulation and menstruation health than those of prior studies [14,28,29]. Samples from 3 different recruitment locations demonstrated variability in racial, ethnic, and other demographic and health-related features.

Acknowledgments

We would like to acknowledge the support of the Department of OB/GYN at Boston University Medical Campus, where this work was conducted, and Vasantha Padmanabhan for her inspirational input.

Conflicts of Interest

None declared.

References

1. Stein IF, Leventhal ML. Amenorrhea associated with bilateral polycystic ovaries. Am J Obstet Gynecol 1935;29(2):181-191. [doi: 10.1016/s0002-9378(15)30642-6]
2. Yildiz BO, Bozdag G, Yapici Z, Esinler I, Yarali H. Prevalence, phenotype and cardiometabolic risk of polycystic ovary syndrome under different diagnostic criteria. Hum Reprod 2012 Oct;27(10):3067-3073. [doi: 10.1093/humrep/des232] [Medline: 22777527]
3. Diamanti-Kandarakis E, Kouli CR, Bergiele AT, Filandra FA, Tsianateli TC, Spina GG, et al. A survey of the polycystic ovary syndrome in the Greek island of Lesbos: hormonal and metabolic profile. J Clin Endocrinol Metab 1999 Nov;84(11):4006-4011. [doi: 10.1210/jcem.84.11.6148] [Medline: 10566641]
4. Boyle JA, Cunningham J, Norman RJ, Dunbar T, O'Dea K. Polycystic ovary syndrome and metabolic syndrome in Indigenous Australian women. Intern Med J 2015 Dec;45(12):1247-1254. [doi: 10.1111/imj.12910] [Medline: 26387977]
5. Rotterdam ESHRE/ASRM-Sponsored PCOS Consensus Workshop Group. Revised 2003 consensus on diagnostic criteria and long-term health risks related to polycystic ovary syndrome. Fertil Steril 2004 Jan;81(1):19-25. [doi: 10.1016/j.fertnstert.2003.10.004] [Medline: 14711538]
6. Sheehan MT. Polycystic ovarian syndrome: diagnosis and management. Clin Med Res 2004 Feb;2(1):13-27 [FREE Full text] [doi: 10.3121/cmr.2.1.13] [Medline: 15931331]
7. Azziz R, Carmina E, Dewailly D, Diamanti-Kandarakis E, Escobar-Morreale HF, Futterweit W, Task Force on the Phenotype of the Polycystic Ovary Syndrome of The Androgen ExcessPCOS Society. The Androgen Excess and PCOS Society criteria for the polycystic ovary syndrome: the complete task force report. Fertil Steril 2009 Feb;91(2):456-488. [doi: 10.1016/j.fertnstert.2008.06.035] [Medline: 18950759]
8. Lizneva D, Gavrilova-Jordan L, Walker W, Azziz R. Androgen excess: investigations and management. Best Pract Res Clin Obstet Gynaecol 2016 Nov;37:98-118. [doi: 10.1016/j.bpobgyn.2016.05.003] [Medline: 27387253]

9. Wild S, Pierpoint T, McKeigue P, Jacobs H. Cardiovascular disease in women with polycystic ovary syndrome at long-term follow-up: a retrospective cohort study. Clin Endocrinol (Oxf) 2000 May;52(5):595-600. [doi: 10.1046/j.1365-2265.2000.01000.x] [Medline: 10792339]
10. Apridonidze T, Essah PA, Iuorno MJ, Nestler JE. Prevalence and characteristics of the metabolic syndrome in women with polycystic ovary syndrome. J Clin Endocrinol Metab 2005 Apr;90(4):1929-1935. [doi: 10.1210/jc.2004-1045] [Medline: 15623819]
11. Jacewicz-Święcka M, Kowalska I. Polycystic ovary syndrome and the risk of cardiometabolic complications in longitudinal studies. Diabetes Metab Res Rev 2018 Nov;34(8):e3054. [doi: 10.1002/dmrr.3054] [Medline: 30089337]
12. Aziz M, Naver KV, Wissing MLM, Mikkelsen AL, Nilas L, Skouby SO. Polycystic ovary syndrome: infertility, cardiovascular, metabolic and obstetrical risks, laboratory and clinical outcomes -- the PICOLO Study. Gynecol Endocrinol 2012 Apr;28(4):253-258. [doi: 10.3109/09513590.2011.613966] [Medline: 22217188]
13. Wolf WM, Wattick RA, Kinkade ON, Olfert MD. Geographical prevalence of polycystic ovary syndrome as determined by region and race/ethnicity. Int J Environ Res Public Health 2018 Nov 20;15(11):2589 [FREE Full text] [doi: 10.3390/ijerph15112589] [Medline: 30463276]
14. Chavarro JE, Rich-Edwards JW, Gaskins AJ, Farland LV, Terry KL, Zhang C, et al. Contributions of the Nurses' Health Studies to reproductive health research. Am J Public Health 2016 Sep;106(9):1669-1676. [doi: 10.2105/AJPH.2016.303350] [Medline: 27459445]
15. Mahalingaiah S, Sun F, Cheng JJ, Chow ET, Lunetta KL, Murabito JM. Cardiovascular risk factors among women with self-reported infertility. Fertil Res Pract 2017;3:1-7. [doi: 10.1186/s40738-017-0034-0] [Medline: 28620545]
16. Mahalingaiah S, Winter MR, Aschengrau A. Association of prenatal and early life exposure to tetrachloroethylene (PCE) with polycystic ovary syndrome and other reproductive disorders in the Cape Cod health study: A retrospective cohort study. Reprod Toxicol 2016 Oct;65:87-94 [FREE Full text] [doi: 10.1016/j.reprotox.2016.07.005] [Medline: 27412368]
17. Bao Y, Bertoia ML, Lenart EB, Stampfer MJ, Willett WC, Speizer FE, et al. Origin, methods, and evolution of the three Nurses' Health Studies. Am J Public Health 2016 Sep;106(9):1573-1581. [doi: 10.2105/AJPH.2016.303338] [Medline: 27459450]
18. Govindaraju D, Adrienne Cupples L, Kannel W, O'Donnell C, Atwood L, D'Agostino R, et al. Genetics of the Framingham Heart Study population. Adv Genet 2008;62:33-65 [FREE Full text] [doi: 10.1016/S0065-2660(08)00602-0] [Medline: 19010253]
19. DeUgarte CM, Woods KS, Bartolucci AA, Azziz R. Degree of facial and body terminal hair growth in unselected black and white women: toward a populational definition of hirsutism. J Clin Endocrinol Metab 2006 Apr;91(4):1345-1350. [doi: 10.1210/jc.2004-2301] [Medline: 16449347]
20. Cheewadhanaraks S, Peeyananjarassri K, Choksuchat C. Clinical diagnosis of hirsutism in Thai women. J Med Assoc Thai 2004 May;87(5):459-463. [Medline: 15222512]
21. Moore JX, Chaudhary N, Akinyemiju T. Metabolic syndrome prevalence by race/ethnicity and sex in the United States, National Health and Nutrition Examination Survey, 1988-2012. Prev Chronic Dis 2017 Mar 16;14:E24 [FREE Full text] [doi: 10.5888/pcd14.160287] [Medline: 28301314]
22. Bode D, Seehusen DA, Baird D. Hirsutism in women. Am Fam Physician 2012 Feb 15;85(4):373-380 [FREE Full text] [Medline: 22335316]
23. Mahalingaiah S. The Ovulation and Menstruation Health (OM) Pilot Study survey instrument. Harvard Dataverse. 2020. URL: https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/XACYQA [accessed 2021-03-31]
24. 2018 population estimates by age, sex, race and hispanic origin. United States Census Bureau. 2019 Jun 20. URL: https://www.census.gov/newsroom/press-kits/2019/detailed-estimates.html [accessed 2020-10-30]
25. Commitment to our community. Boston Medical Center.: Boston Medical Center; 2017. URL: https://tinyurl.com/3smmk59r [accessed 2021-03-31]
26. Jamaica Plain data profile population demographics. City of Boston.: Policy Development & Research Division; 2010. URL: https://www.cityofboston.gov/images_documents/Jamaica_Plain_Planning_District_Profile_tcm3-12993.pdf [accessed 2020-10-30]
27. Internet/broadband fact sheet. Pew Research Center. 2019 Jun 12. URL: https://www.pewresearch.org/internet/fact-sheet/internet-broadband/ [accessed 2020-10-30]
28. All of Us Research Program Investigators, Denny JC, Rutter JL, Goldstein DB, Philippakis A, Smoller JW, et al. The "All of Us" research program. N Engl J Med 2019 Aug 15;381(7):668-676. [doi: 10.1056/NEJMsr1809937] [Medline: 31412182]
29. Wise LA, Rothman KJ, Mikkelsen EM, Stanford JB, Wesselink AK, McKinnon C, et al. Design and conduct of an internet-based preconception cohort study in North America: pregnancy study online. Paediatr Perinat Epidemiol 2015 Jul;29(4):360-371 [FREE Full text] [doi: 10.1111/ppe.12201] [Medline: 26111445]
30. Tsao CW, Vasan RS. Cohort profile: the Framingham Heart Study (FHS): overview of milestones in cardiovascular epidemiology. Int J Epidemiol 2015 Dec;44(6):1800-1813 [FREE Full text] [doi: 10.1093/ije/dyv337] [Medline: 26705418]
31. McManus DD, Trinquart L, Benjamin EJ, Manders ES, Fusco K, Jung LS, et al. Design and preliminary findings from a new electronic cohort embedded in the Framingham Heart Study. J Med Internet Res 2019 Mar 01;21(3):e12143 [FREE Full text] [doi: 10.2196/12143] [Medline: 30821691]

32. Mahalingaiah S, Cosenza C, Cheng J, Rodriguez E, Aschengrau A. Cognitive testing of a survey instrument for self-assessed menstrual cycle characteristics and androgen excess. Fertil Res Pract 2020 Dec 09;6(1):1-10 [FREE Full text] [doi: 10.1186/s40738-020-00088-x] [Medline: 33292647]
33. DiBonaventura MD, Sadosky A, Concialdi K, Hopps M, Kudel I, Parsons B, et al. The prevalence of probable neuropathic pain in the US: results from a multimodal general-population health survey. J Pain Res 2017;10:2525-2538 [FREE Full text] [doi: 10.2147/JPR.S127014] [Medline: 29138590]
34. Hatch EE, Hahn KA, Wise LA, Mikkelsen EM, Kumar R, Fox MP, et al. Evaluation of selection bias in an internet-based study of pregnancy planners. Epidemiology 2016 Jan;27(1):98-104. [doi: 10.1097/EDE.0000000000000400] [Medline: 26484423]
35. Aschengrau A, Seage III GR. Essentials of Epidemiology in Public Health 4th edition. Burlington, MA: Jones & Bartlett Learning LLC; 2020:267-292.
36. Nohr EA, Liew Z. How to investigate and adjust for selection bias in cohort studies. Acta Obstet Gynecol Scand 2018 Apr;97(4):407-416. [doi: 10.1111/aogs.13319] [Medline: 29415329]
37. Rothman KJ, Greenland S, Lash TL. Modern Epidemiology 3rd edition. Philadelphia, PA: Lippincott Williams & Wilkins; 2008.
38. Singh GK, Daus GP, Allender M, Ramey CT, Martin EK, Perry C, et al. Social determinants of health in the United States: addressing major health inequality trends for the nation, 1935-2016. Int J MCH AIDS 2017;6(2):139-164 [FREE Full text] [doi: 10.21106/ijma.236] [Medline: 29367890]

Abbreviations

BMC: Boston Medical Center
BMI: body mass index
BUMC: Boston University Medical Campus
OB/GYN: Obstetrics and Gynecology

Edited by G Eysenbach, R Kukafka; submitted 01.10.20; peer-reviewed by S Butts, R Krukowski; comments to author 23.10.20; revised version received 30.10.20; accepted 11.03.21; published 16.04.21.

Please cite as:
Mahalingaiah S, Cheng JJ, Winter MR, Rodriguez E, Fruh V, Williams A, Nguyen M, Madhavan R, Karanja P, MacRae J, Konanki SC, Lane KJ, Aschengrau A
Multimodal Recruitment to Study Ovulation and Menstruation Health: Internet-Based Survey Pilot Study
J Med Internet Res 2021;23(4):e24716
URL: https://www.jmir.org/2021/4/e24716
doi:10.2196/24716
PMID:33861203

Original Paper

Factors Associated With Perceived Trust of False Abortion Websites: Cross-sectional Online Survey

Sarina Rebecca Chaiken[1,2], BA; Lisa Han[3], MA, PhD; Blair G Darney[1,2,4], MPH, PhD; Leo Han[1], MD, MPH

[1]Oregon Health & Science University, Portland, OR, United States

[2]OHSU-PSU School of Public Health, Portland, OR, United States

[3]Arizona State University, Tempe, AZ, United States

[4]Center for Population Health Research, National Institute of Public Health, Cuernavaca, Mexico

Corresponding Author:
Sarina Rebecca Chaiken, BA
Oregon Health & Science University
3181 SW Sam Jackson Park Road
Portland, OR, 97239
United States
Phone: 1 5034942999
Email: chaiken@ohsu.edu

Abstract

Background: Most patients use the internet to search for health information. While there is a vast repository of searchable information online, much of the content is unregulated and therefore potentially incorrect, conflicting, or confusing. Abortion information online is particularly prone to being inaccurate as antichoice websites publish purposefully misleading information in formats that appear as neutral resources. To understand how antichoice websites appear neutral, we need to understand the specific website features of antichoice websites that impart an impression of trustworthiness.

Objective: We sought to identify the characteristics of false or misleading abortion websites that make these websites appear trustworthy to the public.

Methods: We conducted a cross-sectional study using Amazon's Mechanical Turk platform. We used validated questionnaires to ask participants to rate 11 antichoice websites and one neutral website identified by experts, focusing on website content, creators, and design. We collected sociodemographic data and participant views on abortion. We used a composite measure of "mean overall trust" as our primary outcome. Using correlation matrices, we determined which website characteristics were most associated with mean overall trust. Finally, we used linear regression to identify participant characteristics associated with overall trust.

Results: Our analytic sample included 498 participants aged from 22 to 70 years, and 50.1% (247/493) identified as female. Across 11 antichoice websites, creator confidence ("I believe that the creators of this website are honest and trustworthy") had the highest correlation coefficient (strongest relationship) with mean overall trust (coefficient=0.70). Professional appearance (coefficient=0.59), look and feel (coefficient=0.59), perception that the information is created by experts (coefficient=0.59), association with a trustworthy organization (coefficient=0.58), valued features and functionalities (coefficient=0.54), and interactive capabilities (coefficient=0.52) all demonstrated strong relationships with mean overall trust. At the individual level, prochoice leaning was associated with higher overall trust of the neutral website (B=−0.43, 95% CI −0.87 to 0.01) and lower mean overall trust of the antichoice websites (B=0.52, 95% CI 0.05 to 0.99).

Conclusions: The mean overall trust of antichoice websites is most associated with design characteristics and perceived trustworthiness of website creators. Those who believe that access to abortion should be limited are more likely to have higher mean overall trust for antichoice websites.

(J Med Internet Res 2021;23(4):e25323) doi:10.2196/25323

KEYWORDS

abortion; website trust; internet use; reproductive health; misinformation

Introduction

Most patients use the internet to search for health information [1-3]. While there is a vast repository of searchable information online, much of the content is unregulated and therefore has potential to be incorrect, conflicting, or confusing. As such, the accuracy of online health information remains variable [4]. Despite warnings that patients should remain cautious when accessing health-related websites, patients often view inaccurate health information encountered online as trustworthy [5,6].

Abortion is one of the most common medical services in the world, with approximately 30 million safe abortions accessed annually [7]. Although the abortion rate in the United States (abortions per 1000 females aged 15 to 49 years) has declined over the last decade, internet searches for abortion-related information have steadily increased [8,9]. As the demand for online abortion information rises, many top search results contain incorrect and misleading information [10]. Many misleading websites are created by crisis pregnancy centers (CPCs) and other antichoice organizations, which seek to dissuade patients from accessing abortion and other reproductive health care services [11]. These websites provide deliberately incorrect information about reproductive health, including abortion, often by overstating the risks of abortion and contraceptive options [12,13]. In addition, these antichoice organization websites often appear neutral in order to intercept patients who are potentially seeking an abortion [14].

Previous research identified website design and layout, interactive features, tone or partiality of content, owner's authority, and the website's relationship with an organization or sponsor as key elements impacting the perceived trustworthiness of a website [15-25]. Abortion websites may utilize these characteristics to make the misinformation they seek to perpetuate seem trustworthy. It is therefore important to understand what features of abortion websites impact trustworthiness. In this study, we sought to identify the characteristics of false or misleading abortion websites that make these websites appear trustworthy to the public.

Methods

Recruitment

We conducted a cross-sectional study using an anonymous online survey. We recruited respondents between March 14, 2019, and April 8, 2019, via Amazon's Mechanical Turk (MTurk) platform. MTurk is an online crowdsourcing tool that recruits anonymous users to participate in a variety of computer-based tasks, including assessment of web content [26]. Participants were eligible to respond if they were over 18 years of age, English speaking, and given the "Masters Qualification" by MTurk. Amazon designates participants as "Masters" if they demonstrate reliability in completing a large number and variety of tasks posted by different requesters [27]. We advertised the survey task as "Public Opinion about Abortion Websites." We compensated participants US $2.50 to complete the assignment. After the first 425 responses, we increased compensation to US $3.00 for the final 75 responses in order to increase recruitment. Participants were limited to responding to the survey once based on their unique worker ID number that is assigned by Amazon and is attached to their taxpayer number [27]. The study was reviewed and approved by the Oregon Health & Science University institutional review board.

Website Selection

We selected our websites from a database of abortion websites used in a previous study [10]. For each website in that database, three investigators determined the slant (prochoice, neutral, or antichoice) and slant clarity (easy, moderate, or difficult to determine website slant) based on their impression of the website. All three researchers have published abortion-related research. We then selected 11 websites categorized as both "antichoice" and "difficult to determine" in slant (Multimedia Appendix 1). These 11 antichoice websites included one state government website (Alaska Department of Health and Social Services), four CPC pages associated with brick and mortar establishments, and six websites that detail abortion risks and side effects. We included one website that was rated as being "neutral" in slant and "easy to determine" in slant clarity (The Louisiana Department of Public Health website on "Methods & Medical Risks"). We selected the Louisiana website as a comparator for the antichoice websites.

We used previously validated studies measuring website trust [17-20,24,25] to develop our survey (Table 1). We included items across the following four domains: creators, design, content, and overall trust (Table 1). Participants viewed each website and responded to statements about website characteristics using a 7-point Likert scale ranging from −3 ("strongly disagree") to 3 ("strongly agree") [20,21,25,28]. We assigned each participant to assess three randomly selected antichoice websites and the Louisiana website. Therefore, all participants viewed the Louisiana website but only a subset of participants viewed each of the 11 antichoice websites. Our survey displayed all four websites in random order, and users had 18 minutes to complete the survey and rate all four websites.

Table 1. Website questions.

Category	Question number	Question tag	Statement about the website characteristic
Creators	q0	Creator confidence	I believe that the creators of this website are honest and trustworthy.
Creators	q1	Expert quote	There are experts quoted or referenced on this website.
Creators	q2	Organization	This website notes an association with a trustworthy organization.
Creators	q3	Expert info	Experts wrote information on this website.
Creators	q4	Nonprofit	This website is run by a nonprofit organization.
Design	q5	Interactive	I like the interactive aspects of this website.
Design	q6	Images	I like the use of images on this website.
Design	q7	Up to date	This website appears up to date.
Design	q8	Ads	I can distinguish ads from website content.
Design	q9	Look	I like the look and feel of this website.
Design	q10	Features	This website includes all relevant features/functionalities that I value on websites.
Design	q11	Design	The design of this website is similar to websites I like to use.
Design	q12	Professional	This website appears professional.
Design	q13	Navigate	This website is easy to navigate.
Content	q14	Errors	This website contains errors (spelling or grammar, not accuracy of content).
Content	q15	Address	This website provides an address or phone number.
Content	q16	Contact	This website makes it easy to contact its creators.
Content	q17	Promotional	This website has limited advertising or other promotional material.
Overall trust	q18	Info trust	I trust information from this website.
Overall trust	q19	Quality	Overall, this website is a quality resource.
Overall trust	q20[a]	Biased	This website is biased.
Overall trust	q21	Recommend	I would recommend this page to a friend who is searching for information on this topic.

[a]Question 20 was removed from the overall trust domain owing to poor correlation with other overall trust questions.

We compiled our primary outcome (“mean overall trust”) by averaging three of the four items comprising the overall trust domain. During analysis, we excluded one item “this website is biased” from the primary outcome owing to low correlation with the other questions in the trust domain. We reverse coded responses to items 14 “this website contains errors” and 20 “this website is biased,” so that all positive Likert scale responses become markers of a favorable characteristic.

After participants evaluated all four websites (responses were not required to move on to the next question), we asked sociodemographic questions, including questions on age, race/ethnicity, gender, educational background, state of residence, residential density (rural, urban, and suburban), and internet use (hours per week online) [29]. Unfortunately, an error in the MTurk survey assigned Alabama to those who did not respond to the item regarding state of residence. Finally, we asked about views on abortion, including whether abortion should be legal (no law limiting abortion, some limits, severe limits, or should be illegal in all circumstances) and self-rated abortion knowledge (1-10, with 1 being no knowledge and 10 being expert knowledge).

Analysis

We used RStudio Version 1.2.1335 for all analyses [30]. We used descriptive statistics to characterize the overall demographics of our participants. Due to small numbers, we collapsed race and ethnicity into a binary category (white vs not white). We then conducted analyses for individual items and domains at the website level and multivariate analysis to determine relationships between demographic variables and overall trust.

For each website, we calculated mean scores for responses to the individual items as well as composite means for the four domains, including the primary outcome of mean overall trust. We tested bivariate associations between mean overall trust and each survey item from Table 1. We used correlation coefficients to measure the strength of the relationship of each individual item in the survey with the items comprising the “overall trust” category (info trust, quality, and recommend) as well as mean overall trust itself. Correlation coefficient scores range from “1” (perfect positive correlation) to “−1” (perfect negative correlation), with “0” meaning there is no correlation between two items. In order to test the robustness of our findings, we assessed these correlations in the following three ways: for all 12 websites combined, for the antichoice websites combined,

and for the prochoice Louisiana website alone. We set 0.5 as the cutoff for a "marked" correlation and considered coefficients between 0.4 and 0.5 as indicating "medium" correlation [31].

Last, we used linear regression to identify participant characteristics associated with mean overall trust, where mean overall trust was treated as a continuous variable. We excluded participants with missing items in the trust domain from multivariate modeling (36/498). We created two models. One model was created for the antichoice websites, and another for the Louisiana website alone in order to compare model findings stratified by slant. Variables in our model included race, age, self-rated abortion knowledge, abortion views, hours spent on the internet per week, educational attainment, and residential density. We selected these covariates based on bivariate relationships and prior literature suggesting that these variables impact website trust [24].

Results

Sample

We obtained surveys from 500 participants and excluded only two respondents who did not complete 50% of the website survey questions. We included all other partial responses. Of the 498 respondents, 403 (80.9%) completed all questions for all four websites. Of the 95 incomplete surveys, 53 (56%) missed only one item on one website and fully completed the other three website evaluations. An additional 20 (21%) missed two or three items total across the four websites. A total of 6 (6%) participants missed all 22 items from one website but completed all items from the other three websites.

Participant Characteristics

Full demographic characteristics are shown in Table 2. Respondent age ranged from 22 to 70 years, and 50.1% (247/493) identified as female. Just over half (279/495, 56.3%) of the respondents attained at least a bachelor's degree. The majority of our respondents were white (327/495, 66.1%), and an additional 20% identified as Asian. Participants were distributed across the US census region, with 16.7% (83/498) from the Northeast, 34.7% (173/498) from the South, 19.5% (97/498) from the Midwest, 20.1% (100/498) from the West, and 9.1% (45/498) not from the United States (Table 2). The MTurk error makes it likely that 34.7% is higher than the true proportion of respondents who are from the South. On average, respondents reported spending 35.6 hours per week on the internet (95% CI 1.0-72.8). The majority of participants self-identified as prochoice, with 80.3% (399/497) indicating that there should only be some limits or no law limiting access to abortion. We found self-reported abortion knowledge in our participants to be normally distributed with a mean of 5.9/10 (95% CI 2.4-9.4).

Table 2. Sample characteristics.

Demographics	Value, n (%) or mean (SD; min-max)
Gender (N=493)	
Female	247 (50.1)
Male	245 (49.7)
Other	1 (0.2)
Education (N=495)	
Less than high school and high school	60 (12.1)
Some college/associates	156 (35.12)
Bachelor's degree	213 (43.0)
More than bachelor's degree	66 (13.3)
Location (N=489)	
Rural	100 (20.5)
Suburban	215 (44.0)
Urban	174 (35.6)
Region (N=498)	
Northeast	83 (16.7)
South[a]	173 (34.7)
Midwest	97 (19.5)
West	100 (20.1)
Not United States	45 (9.1)
Race (N=495)	
White	327 (66.1)
Non-white	168 (33.9)
Access (N=497)	
No law	182 (36.6)
Some limits	217 (43.7)
Severe limits	68 (13.7)
Illegal	30 (6.0)
Other	
Age (years)	37.9 (9.9; 22-70)
Abortion knowledge score	5.9 (1.8; 1-10)
Internet hours	35.6 (19.0, 1-107)

[a]Those who left the region question blank were coded as from Alabama and subsequently the South.

Website Outcomes

Mean scores were calculated for each of the four domains (content, creators, design, and overall trust). The mean score of the overall trust domain is the primary outcome of mean overall trust. Domain scores ranged from −0.66 to 1.89, where −3 is highly unfavorable and 3 is highly favorable (Table 3). For the primary outcome of mean overall trust, the Louisiana State Health website had the highest score (1.89) and smallest standard deviation (1.27) (Table 3). In addition, the Louisiana State Health website had the lowest standard deviations in the content (1.27) and design (1.65) domains. By comparison, the Alaska State Health website received the highest score among all websites for creators (1.41), a high score for mean overall trust (1.63), and lower scores for content (1.05) and design (0.97) (Table 3).

Table 3. Responses by domain.

Website	Domain score[a], mean (SD)			
	Content	Creators	Design	Mean overall trust
Louisiana[b]	1.77 (1.44)	1.31 (1.64)	1.30 (1.65)	1.89 (1.27)
Abortion Facts	0.23 (1.80)	0.32 (1.88)	0.11 (1.88)	−0.24 (1.79)
Abortion Pill Risks	0.30 (1.70)	0.39 (1.91)	0.48 (1.91)	−0.04 (1.85)
Abortion Risks	0.41 (1.71)	0.99 (1.71)	0.32 (1.91)	0.49 (1.79)
Alaska	1.05 (1.70)	1.41 (1.58)	0.97 (1.77)	1.63 (1.38)
AmerPreg	1.26 (1.54)	0.97 (1.76)	1.04 (1.67)	0.92 (1.72)
Baby Gaga	0.02 (1.71)	−0.15 (1.64)	0.56 (1.80)	0.09 (1.80)
CareNet	1.61 (1.43)	0.55 (1.67)	1.16 (1.64)	0.92 (1.61)
PregCenter	1.77 (1.42)	0.82 (1.72)	1.61 (1.31)	1.45 (1.49)
RamaInternat	0.59 (1.75)	0.37 (1.80)	0.67 (1.80)	0.37 (1.89)
U Pregnancy	−0.14 (1.67)	−0.54 (1.62)	−0.23 (1.83)	−0.66 (1.59)
WomenRes	1.70 (1.46)	0.70 (1.89)	1.05 (1.71)	1.11 (1.73)

[a]For each item, the maximum rating is +3 and minimum rating is −3.

[b]The Louisiana website is displayed as a neutral website.

Eight of the 11 antichoice websites were associated with a positive mean overall trust score (Table 3). Abortion Facts, Abortion Pill Risks, and U Pregnancy received negative scores for mean overall trust, ranging from −0.66 to −0.04. In addition, Baby Gaga received a score of −0.15 in the creators domain and U Pregnancy received scores of −0.14 for the content domain, −0.54 for the creators domain, and −0.23 for the design domain.

Pooled correlation coefficient matrices were used to compare each item to the three items included in the primary outcome of mean overall trust as well as to the primary outcome of mean overall trust (Table 4, Table 5, and Table 6). We assessed if the correlation between website characteristics and mean overall trust was different for the antichoice websites and the neutral Louisiana website. For both the antichoice websites and the Louisiana website, the creator confidence item ("I believe that the creators of this website are honest and trustworthy") correlated most highly with mean overall trust (Table 5 and Table 6). For the Louisiana website, we found that the creator confidence item was the only item markedly correlated with mean overall trust (coefficient=0.66) (Table 6). In contrast, for the antichoice websites, professional appearance (coefficient=0.59), look and feel (coefficient=0.59), perception that the information is created by experts (coefficient=0.59), association with a trustworthy organization (coefficient=0.58), valued features and functionalities (coefficient=0.54), and interactive capabilities (coefficient=0.52) all demonstrated marked relationships with mean overall trust. Up-to-date appearance (coefficient=0.50), overall design (coefficient=0.49), ability to navigate the website (coefficient=0.48), and presence of images (coefficient=0.46) had medium correlation with mean overall trust. We found no correlation between presence of spelling or grammar errors and mean overall trust (coefficient=−0.07) (Table 5).

Table 4. Correlation coefficients between items in the primary outcome of mean overall trust and all individual items for all websites.

All individual items	Items in the primary outcome			
	Info trust	Quality	Recommend	Mean overall trust
Info trust	1.00	0.86	0.79	0.88
Quality	0.86	1.00	0.79	0.88
Recommend	0.79	0.79	1.00	0.86
Creator confidence	0.76	0.74	0.65	0.72
Professional	0.61	0.64	0.59	0.61
Expert info	0.63	0.61	0.58	0.61
Organization	0.60	0.61	0.57	0.60
Look	0.57	0.59	0.58	0.58
Features	0.53	0.56	0.53	0.54
Up to date	0.52	0.55	0.50	0.52
Navigate	0.49	0.52	0.46	0.49
Design	0.47	0.48	0.48	0.48
Interactive	0.47	0.47	0.48	0.47
Address	0.43	0.43	0.44	0.43
Contact	0.40	0.43	0.42	0.42
Promotional	0.43	0.42	0.38	0.41
Images	0.37	0.38	0.38	0.38
Biased	0.39	0.37	0.29	0.35
Expert quote	0.34	0.34	0.31	0.33
Nonprofit	0.34	0.31	0.29	0.31
Ads	0.27	0.29	0.27	0.28
Errors	−0.06	−0.05	−0.07	−0.06

Table 5. Correlation coefficients between items in the primary outcome of mean overall trust and all individual items for all websites except Louisiana.

All individual items	Items in the primary outcome			
	Info trust	Quality	Recommend	Mean overall trust
Info trust	1.00	0.85	0.78	0.88
Quality	0.85	1.00	0.77	0.87
Recommend	0.78	0.77	1.00	0.85
Creator confidence	0.74	0.73	0.64	0.70
Professional	0.58	0.62	0.58	0.59
Look	0.58	0.60	0.58	0.59
Expert info	0.62	0.59	0.56	0.59
Organization	0.59	0.60	0.55	0.58
Features	0.53	0.57	0.53	0.54
Interactive	0.51	0.53	0.52	0.52
Up to date	0.49	0.52	0.48	0.50
Design	0.48	0.50	0.49	0.49
Navigate	0.48	0.51	0.44	0.48
Images	0.46	0.46	0.46	0.46
Contact	0.37	0.40	0.38	0.38
Address	0.37	0.36	0.39	0.37
Promotional	0.39	0.38	0.35	0.37
Expert quote	0.37	0.37	0.34	0.36
Biased	0.38	0.36	0.28	0.34
Nonprofit	0.37	0.33	0.31	0.34
Ads	0.26	0.29	0.27	0.27
Errors	−0.07	−0.06	−0.08	−0.07

Table 6. Correlation coefficients between items in the primary outcome of mean overall trust and all individual items for Louisiana alone.

All individual items	Items in the primary outcome			
	Info trust	Quality	Recommend	Mean overall trust
Quality	0.82	1.00	0.70	0.84
Info trust	1.00	0.82	0.69	0.84
Recommend	0.69	0.70	1.00	0.80
Creator confidence	0.72	0.69	0.56	0.66
Expert info	0.49	0.48	0.42	0.46
Professional	0.45	0.51	0.41	0.46
Up to date	0.46	0.47	0.41	0.45
Organization	0.44	0.44	0.42	0.43
Features	0.43	0.43	0.43	0.43
Look	0.39	0.38	0.43	0.40
Navigate	0.41	0.42	0.37	0.40
Design	0.31	0.32	0.34	0.32
Promotional	0.31	0.35	0.25	0.30
Interactive	0.28	0.26	0.33	0.29
Expert quote	0.28	0.26	0.23	0.26
Address	0.22	0.27	0.29	0.26
Contact	0.18	0.25	0.28	0.24
Ads	0.20	0.25	0.19	0.21
Images	0.14	0.16	0.17	0.16
Nonprofit	0.16	0.14	0.15	0.15
Biased	0.20	0.17	0.09	0.15
Errors	0.00	0.01	−0.05	−0.01

Multivariate Analysis

We performed multivariate analysis to identify associations between participant demographics and the primary outcome of mean overall trust. We found that preference for fewer abortion restrictions was associated with higher mean overall trust of the Louisiana website (*B*=−0.43, 95% CI −0.87 to 0.01) and with lower mean overall trust of the antichoice websites (*B*=0.52, 95% CI 0.05-0.99) (Table 7 and Table 8). In addition, white race was associated with lower mean overall trust of the antichoice websites (*B*=−1.41, 95% CI −2.26 to −0.55) (Table 8). Other participant characteristics were not associated with mean overall trust.

Table 7. Association between individual demographic factors and the primary outcome of mean overall trust (N=462) for the Louisiana website.

Variable	Estimate	95% CI
Intercept	5.16	2.75 to 7.57
White race	0.36	−0.45 to 1.16
Age	0.004	−0.03 to 0.04
Abortion knowledge	0.14	−0.05 to 0.32
Internet hours	0.0003	−0.02 to 0.02
Abortion access	−0.43	−0.87 to 0.01
Education	0.09	−0.32 to 0.50
Urban residence	0.05	−0.90 to 1.00
Suburban residence	−0.43	−1.35 to 0.49

Table 8. Association between individual demographic factors and the primary outcome of mean overall trust (N=462) for all websites except the Louisiana website.

Variable	Estimate	95% CI
Intercept	−0.49	−3.09 to 2.11
White race	−1.41	−2.26 to −0.55
Age	−0.008	−0.05 to 0.03
Abortion knowledge	0.33	0.13 to 0.53
Internet hours	0.01	−0.01 to 0.03
Abortion access	0.52	0.05 to 0.99
Education	0.08	−0.35 to 0.52
Urban residence	0.10	−0.92 to 1.13
Suburban residence	−0.58	−1.56 to 0.40

Discussion

In this study, we analyzed which characteristics of antichoice websites are highly correlated with mean overall trust. User opinion of website creators had the strongest correlation with mean overall trust, but esthetic and website interface factors were also important. Moreover, even with our predominantly prochoice respondent pool, eight out of the 11 antichoice websites received positive mean overall trust scores, consistent with their selection as websites with "difficult to determine" slant. We also found that a viewer's stance on abortion is inversely related to the mean overall trust of antichoice websites. If participants believed that abortion should be limited or banned, they were more likely to have higher mean overall trust in websites with incorrect information.

To examine the importance of website creators, we compared the high-quality neutral-stance Louisiana state health website with the poor-quality antichoice Alaska state health website. While the antichoice Alaska website received lower scores in the content, design, and overall trust domains than the neutral Louisiana website, the Alaska website received a higher score in the creators domain (1.41 vs 1.31) and was rated as the website with the second highest mean overall trust. Furthermore, we found that design items were less correlated with mean overall trust for the Alaska website than they were for the other antichoice websites. It is reasonable that users would be inclined to assume that an official state health website is trustworthy. Our results suggest that the belief that a website creator is trustworthy can override other aspects of the website with decreased vigilance to lower quality content.

To attract users, antichoice websites use a full array of internet tools that draw from all trust domains to give the impression of credibility [11]. In our study, nine out of 11 antichoice websites received a positive score for the creators domain. These websites use strategies similar to those that brick and mortar CPCs use, such as deceptive advertising, to attract patients [32]. While brick and mortar CPCs often present as though they offer services similar to abortion clinics (ultrasound, "options" counseling, and procedures), the online websites rely on website cues that resemble high-quality resource websites. For instance, many antichoice websites use generic URL titles like "americanpregnancy.org" and include features like "FAQs" that resemble accurate informational websites, but contain misleading information that misrepresents rare or impossible adverse outcomes and fail to refer patients to clinics that provide abortion care. Our findings support that these methods are effective by demonstrating that a website is more likely to have high mean overall trust scores if its appearance and design are highly rated.

Our trends are consistent with prior literature including research examining user trust in other online health contexts [17-20,24,25]. Shanley et al conducted a qualitative study on the perceptions of abortion websites and found that creator expertise and affiliation with an organization were important criteria when assessing website trustworthiness [25]. In addition, several studies describing the impact of health website characteristics on credibility noted that creator "authority" and esthetic attributes, such as design, are associated with trustworthiness [19,21,24]. Of note, several studies indicated that page ranking of search results affects user trust scores [15,20]. On Google web search, many CPC websites pay to become promoted search items and appear at the top of the page [33]. In response to CPC website tactics, reputable organizations should tackle search engine optimization, clearly identify website creators, and prioritize website design.

Our findings should be interpreted with several limitations. We only reviewed a limited sample of antichoice websites and compared them to a single standard "neutral website" in order to learn how these websites are perceived. However, our websites were selected from those most commonly found on a web search, and thus, they are more likely to be seen by the general public. Moreover, compared to prior studies that used qualitative interviews and contained small convenience samples, each one of our websites received over 130 unique evaluations from across the entire country [25]. Our participant pool leaned toward white, higher-educated, and prochoice individuals, consistent with previous studies on how MTurk workers may be demographically skewed [34-36]. However, there is evidence to suggest that MTurk studies are as reliable as studies involving other survey platforms [37,38]. Our study population was diverse as it included a nearly equal number of male and female participants, a wide age range, and participants from all US census regions. Furthermore our multivariate analysis allowed

us to control for many demographic characteristics. Finally, our study did not allow us to determine if causality exists between different website factors and mean overall trust scores. For example, study participants may have chosen to give high scores to all items for websites they regarded as trustworthy. However, many of our websites had low mean overall trust scores but still received high ratings for items in other domains.

This study highlights the characteristics of antichoice websites that are most associated with user trust and demonstrates that many antichoice websites are viewed as trustworthy by a lay audience. People who seek online medical information about abortion may be susceptible to deceptive websites and misinformation. For organizations and individuals seeking to disseminate accurate information about abortion, this study underscores that attention should be directed toward highlighting the credentials of website creators in addition to providing evidence supporting website content.

Acknowledgments

We would like to thank Quinn Schiller for his technological support with MTurk and R Studio.

Authors' Contributions

All authors contributed to the study design, content, and manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
All websites and URLs.
[DOCX File , 14 KB - jmir_v23i4e25323_app1.docx]

References

1. Varga Z, Horváth T. [Patients' preferences for health-related use of Internet]. Orv Hetil 2018 Dec;159(51):2175-2182. [doi: 10.1556/650.2018.31210] [Medline: 30556412]
2. Sayakhot P, Carolan-Olah M. Internet use by pregnant women seeking pregnancy-related information: a systematic review. BMC Pregnancy Childbirth 2016 Mar 28;16(1):65-65 [FREE Full text] [doi: 10.1186/s12884-016-0856-5] [Medline: 27021727]
3. Prestin A, Vieux SN, Chou WS. Is Online Health Activity Alive and Well or Flatlining? Findings From 10 Years of the Health Information National Trends Survey. J Health Commun 2015 Jun 04;20(7):790-798. [doi: 10.1080/10810730.2015.1018590] [Medline: 26042588]
4. Zhang Y, Sun Y, Xie B. Quality of health information for consumers on the web: A systematic review of indicators, criteria, tools, and evaluation results. J Assn Inf Sci Tec 2015 Apr 29;66(10):2071-2084. [doi: 10.1002/asi.23311]
5. Health Information on the Web. US Food & Drug Administration. URL: https://www.fda.gov/drugs/quick-tips-buying-medicines-over-internet/health-information-web [accessed 2021-04-09]
6. Miller LMS, Bell RA. Online health information seeking: the influence of age, information trustworthiness, and search challenges. J Aging Health 2012 Apr 20;24(3):525-541. [doi: 10.1177/0898264311428167] [Medline: 22187092]
7. Ganatra B, Gerdts C, Rossier C, Johnson BR, Tunçalp, Assifi A, et al. Global, regional, and subregional classification of abortions by safety, 2010–14: estimates from a Bayesian hierarchical model. The Lancet 2017 Nov;390(10110):2372-2381. [doi: 10.1016/s0140-6736(17)31794-4]
8. Foster AM, Wynn L, Trussell J. Evidence of global demand for medication abortion information: an analysis of www.medicationabortion.com. Contraception 2014 Mar;89(3):174-180 [FREE Full text] [doi: 10.1016/j.contraception.2013.05.005] [Medline: 24360644]
9. Nash E, Dreweke J. The U.S. Abortion Rate Continues to Drop: Once Again, State Abortion Restrictions Are Not the Main Driver. Guttmacher Policy Review. 2019 Sep 18. URL: https://www.guttmacher.org/gpr/2019/09/us-abortion-rate-continues-drop-once-again-state-abortion-restrictions-are-not-main [accessed 2019-10-09]
10. Han L, Boniface E, Han L, Albright J, Doty N, Darney B. The Abortion Web Ecosystem: Cross-Sectional Analysis of Trustworthiness and Bias. J Med Internet Res 2020 Oct 26;22(10):e20619 [FREE Full text] [doi: 10.2196/20619] [Medline: 33104002]
11. Bryant A, Swartz J. Why Crisis Pregnancy Centers Are Legal but Unethical. AMA J Ethics 2018 Mar 01;20(1):269-277 [FREE Full text] [doi: 10.1001/journalofethics.2018.20.3.pfor1-1803] [Medline: 29542437]
12. Bryant-Comstock K, Bryant AG, Narasimhan S, Levi EE. Information about Sexual Health on Crisis Pregnancy Center Web Sites: Accurate for Adolescents? J Pediatr Adolesc Gynecol 2016 Feb;29(1):22-25. [doi: 10.1016/j.jpag.2015.05.008] [Medline: 26493590]

13. Swartzendruber A, Newton-Levinson A, Feuchs AE, Phillips AL, Hickey J, Steiner RJ. Sexual and Reproductive Health Services and Related Health Information on Pregnancy Resource Center Websites: A Statewide Content Analysis. Womens Health Issues 2018 Jan;28(1):14-20. [doi: 10.1016/j.whi.2017.10.007] [Medline: 29158038]
14. Bryant AG, Narasimhan S, Bryant-Comstock K, Levi EE. Crisis pregnancy center websites: Information, misinformation and disinformation. Contraception 2014 Dec;90(6):601-605. [doi: 10.1016/j.contraception.2014.07.003] [Medline: 25091391]
15. Kammerer Y, Gerjets P. The Role of Search Result Position and Source Trustworthiness in the Selection of Web Search Results When Using a List or a Grid Interface. International Journal of Human-Computer Interaction 2014 Jan 31;30(3):177-191. [doi: 10.1080/10447318.2013.846790]
16. Liao QV, Fu W. Age differences in credibility judgments of online health information. ACM Trans. Comput. Hum. Interact 2014 Feb;21(1):1-23. [doi: 10.1145/2534410]
17. Fogg BJ, Kameda T, Boyd J, Marshall J, Sethi R, Sockol M, et al. Stanford-Makovsky web credibility study 2002: investigating what makes web sites credible today. A Research Report by the Stanford Persuasive Technology Lab & Makovsky & Company. Stanford University. 2002. URL: http://credibility.stanford.edu/pdf/Stanford-MakovskyWebCredStudy2002-prelim.pdf [accessed 2021-04-09]
18. Fogg BJ, Marshall J, Kameda T, Solomon J, Rangnekar A, Boyd J, et al. Web credibility research: a method for online experiments and early study results. In: CHI EA '01: CHI '01 Extended Abstracts on Human Factors in Computing Systems. 2001 Presented at: CHI '01 Extended Abstracts on Human Factors in Computing Systems; March 2001; Seattle, Washington p. 295-296. [doi: 10.1145/634067.634242]
19. Sillence E, Briggs P, Harris PR, Fishwick L. How do patients evaluate and make use of online health information? Soc Sci Med 2007 May;64(9):1853-1862. [doi: 10.1016/j.socscimed.2007.01.012] [Medline: 17328998]
20. Westerwick A. Effects of Sponsorship, Web Site Design, and Google Ranking on the Credibility of Online Information. J Comput-Mediat Comm 2013 Jan 28;18(2):80-97. [doi: 10.1111/jcc4.12006]
21. Robins D, Holmes J, Stansbury M. Consumer health information on the Web: The relationship of visual design and perceptions of credibility. J. Am. Soc. Inf. Sci 2009 Oct 02;61(1):13-29. [doi: 10.1002/asi.21224]
22. Allam A, Sak G, Diviani N, Schulz PJ. Do quality markers for health websites affect the perception of vaccination webpages? Computers in Human Behavior 2017 Feb;67:273-281. [doi: 10.1016/j.chb.2016.11.003]
23. Harris PR, Sillence E, Briggs P. Perceived threat and corroboration: key factors that improve a predictive model of trust in internet-based health information and advice. J Med Internet Res 2011 Jul 27;13(3):e51 [FREE Full text] [doi: 10.2196/jmir.1821] [Medline: 21795237]
24. Sbaffi L, Rowley J. Trust and Credibility in Web-Based Health Information: A Review and Agenda for Future Research. J Med Internet Res 2017 Jun 19;19(6):e218 [FREE Full text] [doi: 10.2196/jmir.7579] [Medline: 28630033]
25. Shanley CM. Assessing credibility in online abortion information. University of North Carolina Digital Repository. 2009. URL: https://ils.unc.edu/MSpapers/3523.pdf [accessed 2021-04-09]
26. Overview of Mechanical Turk. Amazon. 2019. URL: https://docs.aws.amazon.com/AWSMechTurk/latest/RequesterUI/OverviewofMturk.html [accessed 2021-04-09]
27. MTurk FAQs. Amazon. URL: https://www.mturk.com/worker/help [accessed 2021-04-09]
28. Eastin MS. Credibility assessments of online health information: The effects of source expertise and knowledge of content. Journal of Computer-Mediated Communication 2001;6(4):JCMC643. [doi: 10.1111/j.1083-6101.2001.tb00126.x]
29. Rattray J, Jones MC. Essential elements of questionnaire design and development. J Clin Nurs 2007 Feb;16(2):234-243. [doi: 10.1111/j.1365-2702.2006.01573.x] [Medline: 17239058]
30. RStudio Team. RStudio: Integrated Development Environment for R. RStudio. URL: http://www.rstudio.com/ [accessed 2021-04-09]
31. Hazra A, Gogtay N. Biostatistics Series Module 6: Correlation and Linear Regression. Indian J Dermatol 2016;61(6):593-601 [FREE Full text] [doi: 10.4103/0019-5154.193662] [Medline: 27904175]
32. Daniels J. From Crisis Pregnancy Centers to TeenBreaks.com: Anti-Abortion Activism's Use of Cloaked Websites. In: McCaughey M, editor. Cyberactivism on the Participatory Web. United Kingdom: Routledge; 2014:152-166.
33. Holtzman B. Have crisis pregnancy centers finally met their match: California's Reproductive Fact Act. Northwestern Journal of Law & Social Policy 2017;12:78 [FREE Full text]
34. Mason W, Suri S. Conducting behavioral research on Amazon's Mechanical Turk. Behav Res Methods 2012 Mar 30;44(1):1-23. [doi: 10.3758/s13428-011-0124-6] [Medline: 21717266]
35. Huff C, Tingley D. "Who are these people?" Evaluating the demographic characteristics and political preferences of MTurk survey respondents. Research & Politics 2015 Sep 10;2(3):205316801560464. [doi: 10.1177/2053168015604648]
36. Difallah D, Filatova E, Ipeirotis P. Demographics and dynamics of mechanical turk workers. In: WSDM '18: Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining. 2018 Presented at: Eleventh ACM International Conference on Web Search and Data Mining; February 2018; Marina Del Rey, CA, USA p. 135-143. [doi: 10.1145/3159652.3159661]
37. Buhrmester M, Kwang T, Gosling S. Amazon's Mechanical Turk: A New Source of Inexpensive, Yet High-Quality, Data? Perspect Psychol Sci 2011 Jan;6(1):3-5. [doi: 10.1177/1745691610393980] [Medline: 26162106]

38. Levay KE, Freese J, Druckman JN. The Demographic and Political Composition of Mechanical Turk Samples. SAGE Open 2016 Mar 15;6(1):215824401663643. [doi: 10.1177/2158244016636433]

Abbreviations

CPC: crisis pregnancy center
MTurk: Mechanical Turk

Edited by R Kukafka; submitted 30.10.20; peer-reviewed by J Swartz, D Levine; comments to author 08.12.20; revised version received 05.02.21; accepted 16.03.21; published 19.04.21.

Please cite as:
Chaiken SR, Han L, Darney BG, Han L
Factors Associated With Perceived Trust of False Abortion Websites: Cross-sectional Online Survey
J Med Internet Res 2021;23(4):e25323
URL: https://www.jmir.org/2021/4/e25323
doi:10.2196/25323
PMID:33871378

Original Paper

Requirements and Operational Guidelines for Secure and Sustainable Digital Phenotyping: Design and Development Study

Raj R Jagesar[1], MSc; Jacob A Vorstman[2,3*], PhD, MD; Martien J Kas[1*], Prof Dr

[1]Groningen Institute for Evolutionary Life Sciences, University of Groningen, Groningen, Netherlands

[2]Program in Genetics and Genome Biology, Research Institute, The Hospital for Sick Children, Toronto, ON, Canada

[3]Department of Psychiatry, University of Toronto, Toronto, ON, Canada

[*]these authors contributed equally

Corresponding Author:
Martien J Kas, Prof Dr
Groningen Institute for Evolutionary Life Sciences
University of Groningen
Nijenborgh 7 9747 AG
PO Box 11103
Groningen, 9700 CC
Netherlands
Phone: 31 503632381
Email: m.j.h.kas@rug.nl

Abstract

Background: Digital phenotyping, the measurement of human behavioral phenotypes using personal devices, is rapidly gaining popularity. Novel initiatives, ranging from software prototypes to user-ready research platforms, are innovating the field of biomedical research and health care apps. One example is the BEHAPP project, which offers a fully managed digital phenotyping platform as a service. The innovative potential of digital phenotyping strategies resides among others in their capacity to objectively capture measurable and quantitative components of human behavior, such as diurnal rhythm, movement patterns, and communication, in a real-world setting. The rapid development of this field underscores the importance of reliability and safety of the platforms on which these novel tools are operated. Large-scale studies and regulated research spaces (eg, the pharmaceutical industry) have strict requirements for the software-based solutions they use. Security and sustainability are key to ensuring continuity and trust. However, the majority of behavioral monitoring initiatives have not originated primarily in these regulated research spaces, which may be why these components have been somewhat overlooked, impeding the further development and implementation of such platforms in a secure and sustainable way.

Objective: This study aims to provide a primer on the requirements and operational guidelines for the development and operation of a secure behavioral monitoring platform.

Methods: We draw from disciplines such as privacy law, information, and computer science to identify a set of requirements and operational guidelines focused on security and sustainability. Taken together, the requirements and guidelines form the foundation of the design and implementation of the BEHAPP behavioral monitoring platform.

Results: We present the base BEHAPP data collection and analysis flow and explain how the various concepts from security and sustainability are addressed in the design.

Conclusions: Digital phenotyping initiatives are steadily maturing. This study helps the field and surrounding stakeholders to reflect upon and progress toward secure and sustainable operation of digital phenotyping–driven research.

(J Med Internet Res 2021;23(4):e20996) doi:10.2196/20996

KEYWORDS

digital phenotyping; mobile behavioral monitoring; passive behavioral monitoring; smartphone-based behavioral monitoring; research data management; psychoinformatics; mobile phone

Introduction

Background

Digital phenotyping is the practice of collecting and analyzing objective, longitudinal, and possibly high-resolution data streams from personal devices, such as smartphones and wearables, which are descriptive of a person's real life and real-time behavior [1]. Digital phenotyping provides a new and much more detailed perspective on human behavior, with the potential to innovate both research and clinical outcome measures in the health care space.

The field is currently still experimental, featuring many studies reporting on methodologies and pilot data, but it has yet to deliver replicable findings that may prove useful for clinical translation [2]. At present, much of the current efforts concentrate on correlating smartphone-derived data with clinical diagnoses and symptoms domains, whereas the next steps will examine to what extent such data can be exploited to bring about real positive changes in clinical care [3]. In addition, the ethical aspects of digital phenotyping are examined [4]. At any rate, the pace of innovation is fast, the field is expanding rapidly, and the sample size of studies is steadily increasing, raising the question of whether the applications and services at the backend of such programs are up to the task of continuing to support the operations. More concretely, participating in large cohort studies and operating in regulated research spaces require us to ensure that our platforms are sustainable (ie, maintainable and scalable [5]) and, most importantly, secure [6]. Human research invariably requires the participation of individuals willing to participate in studies after obtaining appropriate informed consent. Establishing trust and continuity is key, given the sensitivity of the data that we collect and the context in which we operate [7].

The current state of the art shows that security and sustainability are not at the forefront of design and execution of digital phenotyping initiatives, despite some awareness of security and privacy implications [8] and the high cost of sustainability [9] in the currently available tools. However, the uptake in the field dictates a necessary shift from the implementation of a limited set of security measures to a more mature and holistic conception of all aspects of sustainability and security as integral parts of software engineering, starting at the very first stages of design. In many current reports on behavioral monitoring platforms, these factors are only discussed in a limited fashion [10-12]. We do not want to suggest that these initiatives are necessarily insufficient in this respect, but in general, there appears to be a focus on the scope of phenotypic outcomes captured by the platform, whereas attention to aspects related to security and sustainability is relatively peripheral.

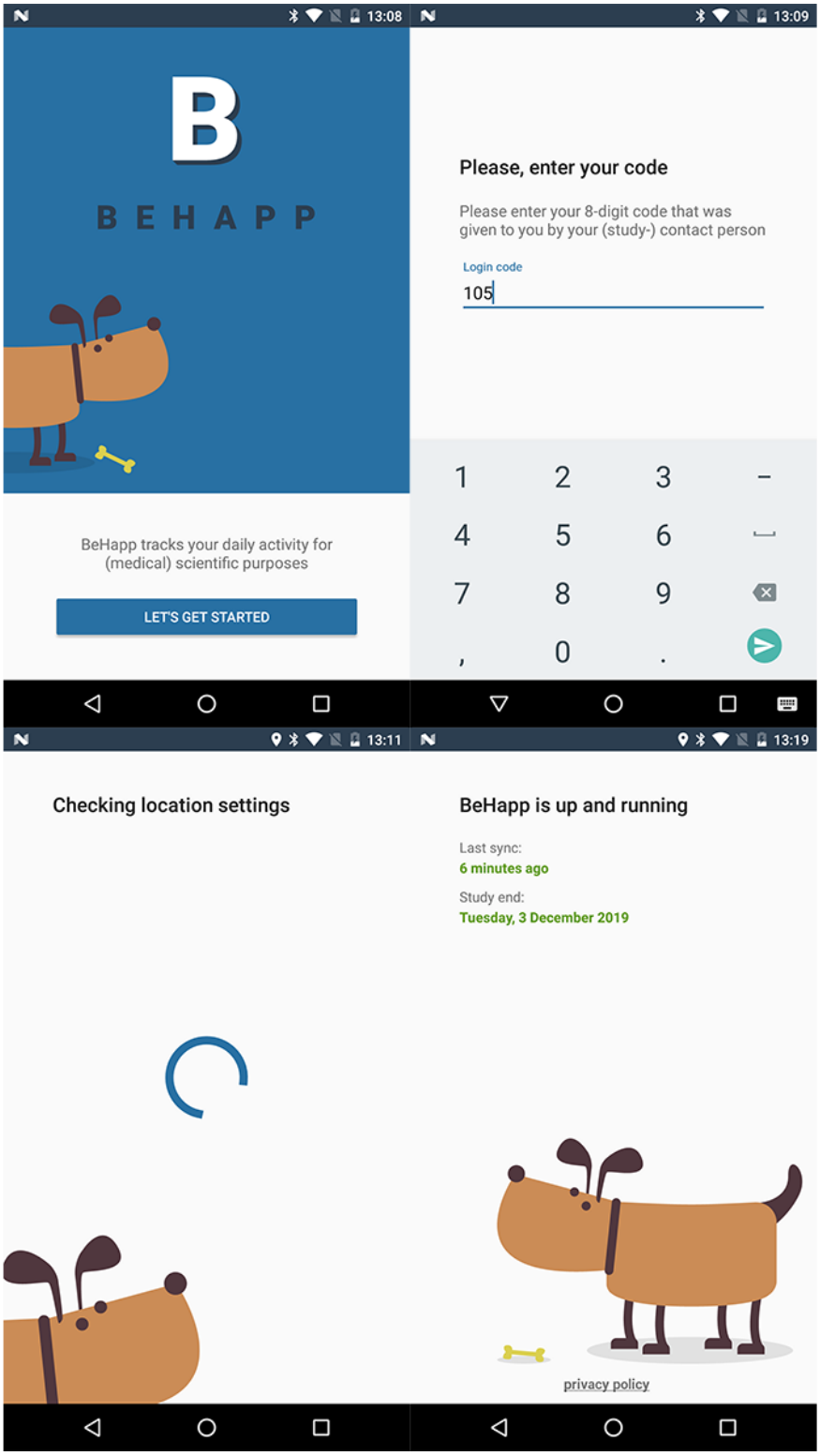
To address this gap, we present here a detailed description of the requirements and guidelines to responsibly develop and operate a behavioral monitoring platform, with specific consideration of the aforementioned concerns.

About BEHAPP

This research is part of the BEHAPP project [13]. BEHAPP is a research platform that features the following 2 components:

1. The front end (Figure 1): a mobile app originally conceptualized for mobile passive monitoring (MPM) of human subjects. As a subfield of digital phenotyping, MPM refers to the practice of naturalistic observation through personal mobile devices exclusively relying on the collection of data that do not require any active input from the participant (an example of active input would be queries probing for emotions or situations such as in the experience sampling methods [14]).
2. The backend, which is the focus of this study: the backend is designed following the software as a service paradigm supporting multicenter studies for international research consortia, academic institutes, and the pharmaceutical industry. The work presented stems directly from our experience in accommodating the needs of industry partners and research groups representing large-scale study cohorts.

Figure 1. Screenshots of the installation and activation flow of the BEHAPP smartphone app.

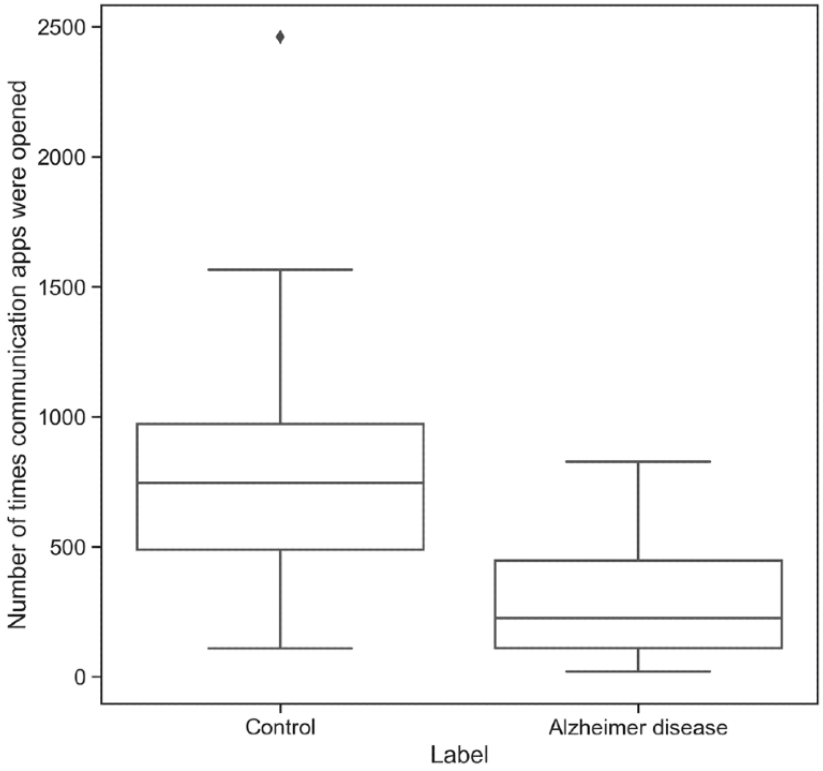


BEHAPP is comparable with initiatives such as MindLamp (Division of Digital Psychiatry, Beth Israel Deaconess Medical Center) [11], Beiwe (Onnela Laboratory, Harvard T H Chan School of Public Health) [15], Funf Open Sensing Framework [16], and the AWARE Framework [17]. All these platforms have in common that they collect various data streams through individual smartphones with the purpose of facilitating the study of human behavior in real time and in a natural (real-world) setting. BEHAPP also draws from a wide array of data collection sources that can be tailored to the needs of specific studies.

From the backend perspective, however, BEHAPP is different from the aforementioned initiatives. The alternatives require research teams to set up, manage, monitor, and secure the basic technical infrastructure themselves. Rather, BEHAPP is designed to be offered as a fully managed service aimed at low effort integration in (existing) studies. BEHAPP is currently used as an exploratory research instrument in general behavioral

and clinical intervention studies. From these studies, we have provided a proof of principle demonstrating that these tools, the resulting data, and the clinical measures that we extract can be used to distinguish between neuropsychiatric patient and control groups [18], courtesy of the Psychiatric Ratings Using Intermediate Stratified Markers (PRISM) program [19]. For example, as depicted in Figure 2, we highlight a feature of communication app use. In this example, we observe that the overall frequency of communication app use is lower for the group with Alzheimer disease when compared with age-matched healthy controls. The service is currently exclusively used as a research instrument and is not employed as a tool to assist in the diagnosis, prevention, monitoring, treatment, or alleviation of disease, that is, at this moment in time, BEHAPP is not a medical device [20].

Figure 2. A group comparison data example on communication app use between Alzheimer disease patients and age-matched healthy controls (n=30) demonstrating the ability to observe differences in behavior through the use of BEHAPP's behavioral monitoring platform. We measured a statistically significant difference in the number of times that communication apps were opened (median 746.5 and 226.0 times communication apps were opened; *P*=.003). The median age was 66.5 years (60% male and 40% female).



Methods

Overview

We focus on security and sustainability as a starting point for the development and operation of a behavioral monitoring platform. Within the concept of security, we address measures such as data isolation and encryption and highlight the importance of organizational security. We continue to discuss sustainability, explaining what is required to ensure that our platforms remain in service in a secure and stable way.

Security

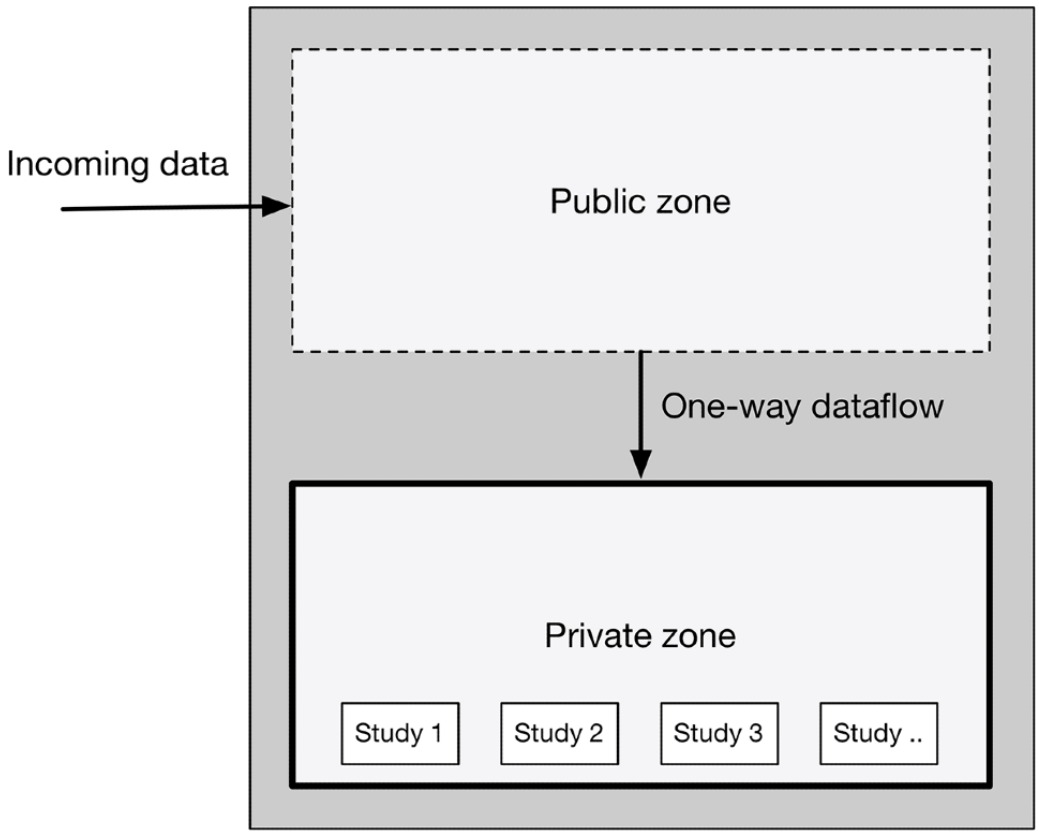
Security refers to the "absence of unauthorized access to information systems" [21]. Given the sensitivity of the data collected, the risk of (accidental) data leaks is a key consideration in setting up the platform. The current state of affairs on information security teaches us to step back and reflect critically on the design of our systems [22]. Regulatory frameworks such as the European General Data Protection Regulation (GDPR), in effect since 2018, must also be considered, given their important impact on technology development [23]. The GDPR expects initiatives involved with personal data to carry out a data protection impact assessment (DPIA) [24]. In a DPIA, we evaluate data sensitivity and how these data must be handled to prevent unauthorized access to and loss of personal data. Formulating a defense in-depth strategy is one such approach to address these concerns [25]. Defense in-depth strategies aim to ensure security by layering security measures at both the technical and organizational levels.

Data Isolation

Data isolation aims to minimize the exposure of data to data consumers. Data consumers can be both technical elements, such as (web) servers, and researchers interacting with data. The intent is to keep high-risk technical elements away from sensitive data. Technical elements connected to the public internet are particularly high risk because anyone can interact with and potentially exploit possible vulnerabilities of these elements. The most threatening are zero-day exploits, which are novel vulnerabilities that, by definition, are unpublished and thus difficult to protect against [26]. To address these risks, isolation of all sensitive data from publicly exposed technical elements is essential. This was achieved through network segmentation [27]. For BEHAPP, we established a public and

private zone of operation, which are strictly separated (Figure 3). The public zone, which carries a higher risk profile, is responsible for receiving participant data from the outside world and immediately transmitting these data to the private zone. The private zone, which has limited outside world connectivity, receives and permanently stores participant data.

Figure 3. Public and private zone isolation overview featuring separated data storage.



Furthermore, we defined a set of rules specifying the flow of data within the BEHAPP platform:

- Participant data may not be stored in a publicly exposed zone: The platform *needs* access to the public internet to receive data from smartphones and wearables. However, as explained, exposure to public internet access carries the risk of (part of) the platform (unknowingly) becoming compromised. Therefore, as a rule, any technical element that lives in BEHAPP's public zone may not (temporarily) store any sensitive participant data, regardless of its encrypted state.
- Participant data may only flow in one direction, from the public to the private zone: Publicly exposed elements may only serve as an ingestion point for sensitive data to flow through to isolated private zones. Technical elements in BEHAPP's public zone do not have any capability to interact with or retrieve participant data from the private zone and are completely unaware of the existence of these data.
- Participant data must be stored in a separate database for each study: The principle of least privileges dictates that automated operations and researchers only need to be assigned the absolute minimal set of authorizations to perform their task [28]. Therefore, any data collected by BEHAPP is segmented into separate data sets for each study, thereby further isolating the data allowing for granular access permissions. By separating study databases, our researchers can be authorized to access only specific studies. Furthermore, researchers are only given permission to *read* the data, which guarantees data integrity. This limits the potential fallout from data being lost or corrupted because of accidental leaks or compromised user accounts.

Note that data isolation by itself does not necessarily limit the processing of data. A private zone of operation can feature technical elements that perform automated data analysis tasks such as data enrichment and annotation, extraction of clinical endpoints, and data compliance and quality control checks. The results can then be written back to the public zone to enable direct reporting to study partners on the condition that the resulting information is fully anonymized. Finally, the resulting information can also be used for signaling purposes, for example, to address potential data quality and compliance issues that are found during automated data analysis runs. At BEHAPP, we are currently in the process of developing such capabilities.

Encryption

Data encryption is the practice of obfuscating data by "converting information from an intelligible form into an unintelligible form" [29], thus rendering data unusable in case of a data leak. *The main challenge lies not in the application of encryption itself but in adequately managing encryption keys.* Protecting encryption keys is equally as important as protecting raw data; otherwise, the added level of protection will be de

facto limited or nonexistent. A layered approach allows for responsible key and, subsequently, data management.

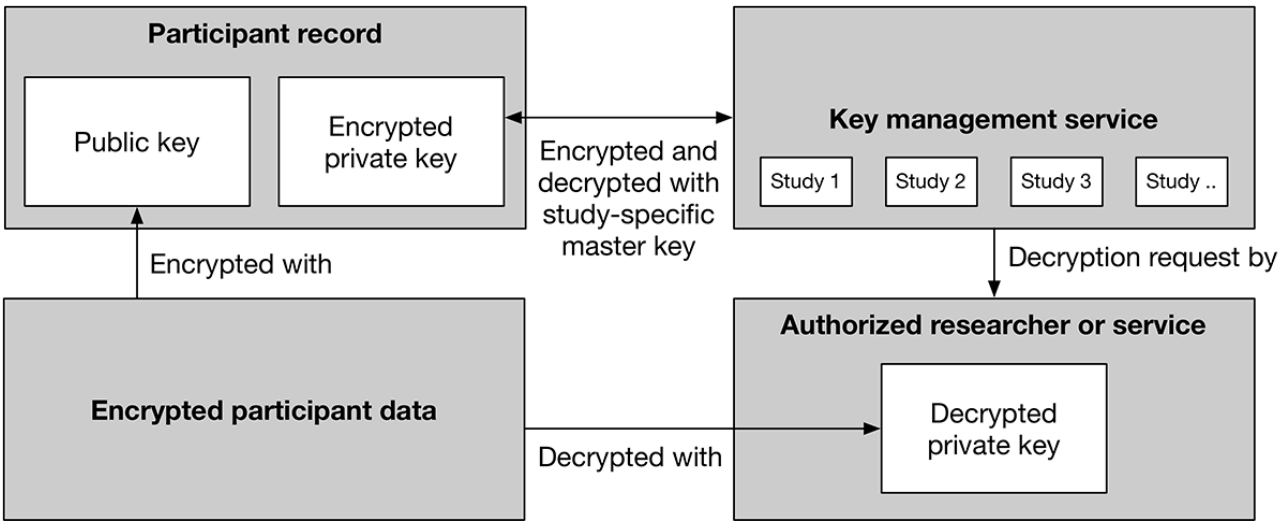
For BEHAPP, we employ a combination of symmetric and asymmetric encryption techniques to establish a closed encryption hierarchy. The main difference between both encryption techniques is in the key material used for the encryption and decryption of data. Symmetric encryption uses one key that is used for both encryption and decryption purposes. Asymmetric encryption uses 2 keys (a key pair), consisting of a public key and a private key. The public key is meant to encrypt data and therefore can be freely shared. The private key is meant for decrypting data and therefore needs to be stored safely [30].

The aforementioned encryption hierarchy consists of 2 levels:

1. Each study participant on the platform is assigned an asymmetric key pair. The key pair is generated in a private zone to prevent leakage of private key information. This key pair is used for encryption and decryption operations on sensitive participant data, such as the raw data that are collected.
2. Each private key is then encrypted itself before being stored in a central database. Private keys are encrypted using a key management service (KMS) [31] using symmetric encryption. A core characteristic of KMSs is that the key material itself never leaves the service. Instead, cryptographic operations are performed based on user authorization. This closes the encryption hierarchy because (unauthorized) database access by itself is not sufficient to decrypt data. Instead, access to both the database and KMS is required. For BEHAPP, we provide study-specific master keys within our KMS, segmenting the encryption hierarchy. The concept of isolation applies here as well, resulting in limited access to cryptographic operations on a per-study basis.

This results in a 3-step encryption model, where (1) the researcher first loads the encrypted private key for a specific participant, (2) the researcher requests a decrypted version of this private key from the KMS, and (3) the researcher loads encrypted participant data from the database and decrypts the data using the decrypted private key obtained from the KMS (Figure 4). Authorizations such as data access permissions and permissions for cryptographic operations are checked at all stages of this process.

Figure 4. Three-step encryption model: (1) load a participant-bound encrypted private key, (2) request a decrypted version of the private key from the key management service, and (3) load participant data and decrypt with the decrypted private key.



Finally, we adhere to the following rule on the use of encryption on the platform: data may only be collected, transported, and stored in an encrypted state. Encryption measures are continuously required to remain in place at all stages of the data lifecycle. In other words, data are only decrypted when necessary and discarded immediately after use. This means that data must be encrypted directly when they are collected through a smartphone or wearable device. For BEHAPP, we achieved this goal relying on the use of asymmetric encryption, as discussed earlier. We send a participant's public key, which is safe to share, to their privately owned devices to be used by our app to encrypt any data that are collected. The encrypted data then remain in this state until they arrive at their destination in the private zone. Analytical workloads naturally require a decrypted view of the data. They are performed by keeping sensitive data in volatile memory only for the purpose of exploration, experimentation, and extraction of outcome measures. The data are then purged from memory.

Information Security Culture

Technical measures alone are not sufficient to protect sensitive data collected by behavioral monitoring platforms. Users of information systems, such as researchers, play an important role in the secure use of information systems and, subsequently, any data under study. Raising awareness about security issues and safe practices is of paramount importance. Indeed, security incidents as a consequence of lack of awareness or negligence are common occurrences [32]. Researchers and newly entering Doctor of Philosophy (PhD) students, in particular, represent a class of users who, by the nature of their work, require access to sensitive data but who may not have any prior experience in the safe handling of these data. The challenge is to balance

between maintaining security while not inhibiting (experimental) workflows of researchers [33]. We recommend the following guidelines to promote information security: (1) provide training and messaging to raise general security awareness and (2) provide seamless security to end users only minimally confronting them with security and compliance decisions [34].

First, raise awareness by establishing an information security policy (ISP) and ensure that all research staff are given an information security briefing. This briefing can be part of a general onboarding process, and its contents must be based on the ISP. An ISP is a document that defines "the rights and responsibilities of information resource users" [35]. The aim of this document is to explain how users responsibly handle sensitive data on a daily basis as part of their work. An effective document is understandable, practical, and inclusive of the needs of researchers. The ISP of the BEHAPP project is targeted at an audience with above-average computer literacy (PhD students, postdocs, and principal investigators). Given this audience and the sensitive data that we use, we specify concrete rules for high-risk areas of attention, such as data flow. For example, we concretely specify that data may only reside on the central server and that local copies on individuals' devices may temporarily exist for analytical purposes only. We also clearly specify that the data may not be uploaded or transferred to any other service or device (eg, for data enrichment purposes) without discussing the purpose and scope of this intention with the team.

Second, seamless security is provided through software development effort. We recommend investing in the development of custom toolkits that are responsible for the heavy lifting around safely loading, decrypting, exploring, and analyzing data. Furthermore, user authentication should be based on multifactor authentication strategies adhering to modern password security guidelines, as defined in National Institute of Standards and Technology 800-63-3 [36]. This considerably eases the burden of security compliance–conscious behavior. At BEHAPP, we developed an internal tool based on these exact principles, called the behapp-data-kit. The kit is a Python package or library aimed at ease of use and security, offering a simple programming interface for data exploration and analysis. Meanwhile, security compliance is handled hidden from the researchers' perspective. For example, the data kit automatically manages local copies of data, ensuring that these copies are encrypted and removed when they are older than 14 days. Furthermore, when loading the data into active working memory for analysis, the kit ensures that the decryption keys necessary for decrypting the data are only held in memory for the shortest amount of time, explicitly deleting these keys when they are no longer required. Finally, the kit relies on a single multifactor authentication strategy, which results in a high level of trust without researchers having to deal with multiple sets of credentials.

Third, schedule and hold weekly team meetings to discuss any (potential) use of sensitive data. Reflect on whether the use complies with the ISP and if the ISP still holds relevance or if an adjustment is required. Be especially mindful of *shadow security.* Shadow security refers to ad hoc practices devised by security-conscious employees that are not compliant with formally prescribed security policies in an effort to achieve a more optimal balance between getting work done and protecting information security [37]. Kirlappos et al [37] recommend learning from these practices, arguing that without engaging with users on these practices, one cannot claim that a specified security infrastructure exists as intended. For example, at BEHAPP, over time, the security policy proved to be difficult to accommodate to researchers who were not directly affiliated with the team, such as graduate students working on temporary assignments. The logistics of account and hardware security key provisioning did not fit the short and temporary character of these projects. Thus, the following shadow security practice emerged: manual data exports were generated for graduate students, but these exports were limited to fully anonymized clinical measure overviews. Although initially any form of manual data exports was formally prohibited, the adjustment of providing anonymized exports offered a workable middle ground and has been adopted as a standard practice.

Sustainability

Sustainability refers to the ability to ensure availability, support, and improvement of the software products and services that we create [38]. We highlight 3 qualities that are closely connected to the concept of sustainability: maintainability, reliability, and scalability [5].

Maintainability is the degree of effectiveness at which software products can be modified [39-41], which depends on multiple factors such as documentation, design, and the consistent application of clean coding standards. Maintenance is an essential part of operating a software service. *The foundations that we built upon, such as mobile operating systems and web application frameworks, change continuously and thus require frequent modification of our own code.* Failure to do so results in diminished service performance, possible loss of functionality, and increased exposure to security threats. Maintenance is especially relevant with regard to mobile apps that we employ as our measurement instruments. Each yearly upgrade of mobile platforms brings about changes that often directly impact the quality of the data that we intend to collect. In addition, noncompliance with continuously changing Apple's App Store and Google's Play Store policies may result in apps being removed altogether. Thus, keeping up with maintenance-related tasks is essential in that it avoids or at the very least mitigates the potential negative impact of platform upgrades and policy changes. Unfortunately, negative changes cannot always be avoided. For example, over the current 4-year development span with BEHAPP, we have experienced many changes in Google's terms of service for the distribution of our Android (Google) app through the Play Store. One change, in particular, revolved around Google's strides to curb malicious apps invading the privacy of their users. This change, which came into effect in March 2019, limited access to call and text messaging logs for the majority of apps distributed through Google's Play Store [42]. This change was unfortunate, as call logs are very expressive of the communicative behavior of participants. Fortunately, we could work around this problem by directly distributing our app to our end users and thus circumvent the Play Store through sideloading.

Reliability refers to the probability of failure-free operation of a software product [43]. The goal is to ensure that our platforms are highly available, keeping service downtime at a minimum. However, it must be noted that experimental research initiatives may typically have wider tolerances for service uptime requirements. This changes when our behavioral monitoring platforms evolve to a stage where they are involved with mission critical workloads, for example, when their continuity is essential to large-scale research endeavors or when they provide important information to clinical care processes.

Scalability is the ability of a software product to adapt to changing circumstances in demand [44]. The rapid adoption of behavioral monitoring platforms and increasing cohort sizes require us to ensure that our software products can sustain the stresses of increases in demand.

Finally, note that the aforementioned quality attributes and the corresponding responsibilities not only apply to the programs and code behind our platforms but also to every supporting information technology (IT) infrastructure element that is required to bring our platform into service. This can span the full range of elements such as networking, storage, physical servers, virtualization, operating systems, and web server software. All these elements require setup and maintenance for stable and continued secure operation. The proverb that a chain is only as strong as its weakest link applies here: a vulnerability or weakness in any supporting IT element will affect the other elements as well, including our behavioral monitoring platforms. *Thus, we are not only required to develop a secure and sustainable behavioral monitoring platform but we also need to ensure that the supporting IT infrastructure is equally secure and sustainable.* Unfortunately, both the knowledge and resources required to do so are extensive and form a barrier to entry.

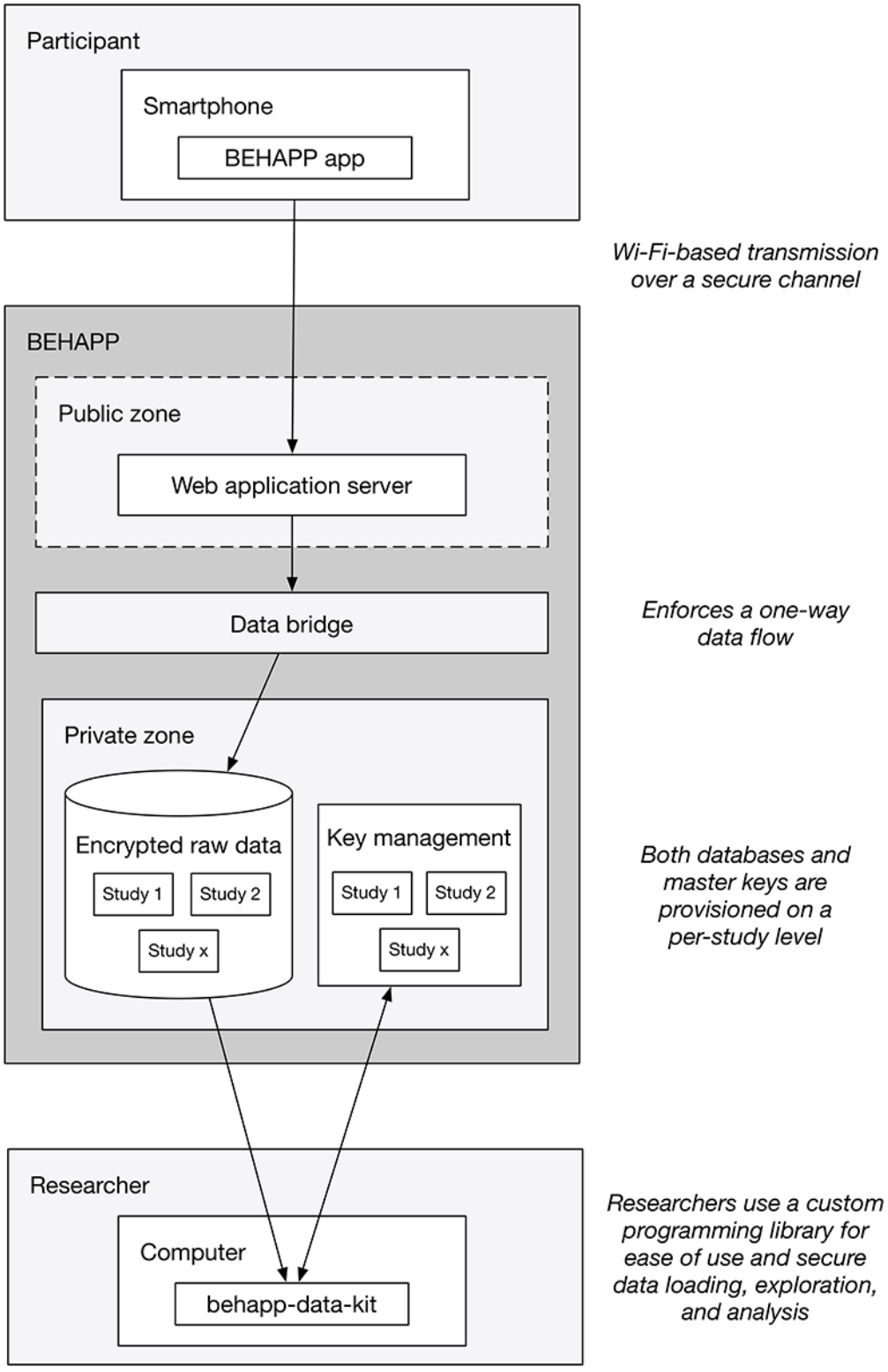
At BEHAPP, we have built a service relying on fully managed IT infrastructure components offered by large cloud providers, focusing in particular on an upcoming class of products known as serverless computing. With serverless computing, the cloud provider is responsible for maintaining and securing the majority of the required IT infrastructure, leaving us to focus on our platform only [45]. This includes the ability to (rapidly) scale, which means that we can flexibly meet any changes in demand.

Results

Here, we introduce the base BEHAPP data flow for the collection and analysis of raw data (Figure 5). The flow implements the requirements and guidelines, as discussed in the aforementioned sections:

1. Data collection starts at the personal devices of participants that run the BEHAPP smartphone app. The app unobtrusively collects various types of data descriptive of the participants' behavior. Any data collected were encrypted immediately and temporarily stored on the device.
2. The data are sent to the public facing the web application server. The smartphone app is programmed to upload these data on a fixed interval and attempts to do so on the condition that a Wi-Fi (or unmetered network) is available. Given that some modes of data collection can yield substantial amounts of data, we do not want to risk accidentally consuming the data plans of our participants. Data transmission always occurs over a secured connection. Once received, the web application server immediately passes the data on to the data bridge.
3. The *data bridge* is not a formal concept but represents a technical construct responsible for enforcing a one-way data flow, which can be achieved in multiple ways. BEHAPP employs message queues configured with minimized permissions to realize unidirectional flow of data.
4. The private zone receives data from the bridge and is responsible for placing the raw data in the corresponding study database. In this way, researchers can access only a select number of databases instead of being able to access the full data set.
5. Finally, researchers use a custom programming library, the behapp-data-kit, designed for ease of use and security. The kit supports data loading, exploration, and analysis and is aimed at security and ease of use. Security compliance, such as maintaining data encryption and enforcing short-lived life cycles of decrypted and locally stored data, is hidden from view. This allows researchers to focus on the task of data analysis while maintaining a secure level of use.

Figure 5. BEHAPP data flow for the collection and analysis of data using smartphone-based participant data.



Discussion

Principal Findings

We present a data flow design that includes a set of requirements and operational guidelines for the realization and responsible operation of a digital behavioral monitoring platform. Having done so, we recognize the following limitations in the proposed work.

First, the proposed data isolation measures prevent the public elements of the platform from accessing (raw) sensitive data. In other words, by design, researchers will not be able to retrieve sensitive data in an automated way. Instead, data are provided (both raw and extracted features) manually on request. This may not be an alluring policy, as it implies that researchers cannot have immediate data access. However, given the risk profile of publicly connected technologies, we opted to go for a risk-averse approach and stand by the decision to strictly

separate sensitive data to protect the participants under observation by the platform. In line with the aforementioned data isolation rules, once received, any form of sensitive data may not leave the service through a publicly connected element. However, there are ways to automate the retrieval of sensitive data, for example, through the use of secure file transfer mechanisms initiated in private zones, which should be the focus of further development and future studies.

Second, by implementing both technical and organizational measures to protect the data of our participants, we realize that there is room for improvement on the data protection front by the application of anonymization methods. BEHAPP's data, while fully encrypted, are not fully anonymous and have, considering the location data that are collected from, the potential for direct identification of individuals. However, although strategies exist for anonymizing location data, its effectiveness is still under debate [46]. As a consequence, analytics endeavoring down the line could be limited by the application of anonymization strategies. For example, we would lose the ability to retrospectively annotate locations by adding meaning to specific points (eg, schools, hospitals, and sports). However, note that although raw location data are considered as identifiable information, the behavioral features that we extract and subsequently report are not. Features such as *total time spent at home* and *total number of unique places visited* are highly interpretable and expressive of one's mobility without having to refer to any geographical type of data and as such can reside in publicly connected zones.

Third, considering our aforementioned proof of principle, we demonstrate the capability to accurately distinguish between patient and control groups. We feel it is time to look ahead and ensure that the models that we intend to build, be it for predictive or classification purposes, are not only properly validated but also compliant and sustainable from an ethical perspective. Many unintuitive and unverifiable inferences can be drawn from personal data, which can potentially result in negative consequences for the individuals for whom the inferences are drawn. Although the GDPR demands model transparency, the subject of inferential analytics is not well regulated [47]. Given our medical scientific operating context, we should tread carefully and actively work toward the creation of transparent models with limited application scopes to avoid negatively affecting our subjects under study.

Fourth, despite our efforts to simplify the design and minimize the operational overhead of running a behavioral monitoring platform, we realize that the level of complexity and the demands imposed on a research team may still leave this type of instrumentation out of reach for many independently operating research groups. However, we strongly feel that this is the minimum standard for responsibly operating such a platform. Self-hosted open-source models may be vulnerable in this regard and, therefore, are not the most secure and sustainable way forward. As argued in this paper, the responsible operation of such platforms extends beyond installing a client and server application, and we have to consider the underlying IT infrastructure as well. In our experience, the broad level of responsibilities tied to operating a behavioral monitoring platform warrants the inclusion of a dedicated team responsible for development, maintenance, monitoring, and security tasks. A more workable model would be to concentrate the required operational effort in a limited number of initiatives capable of supporting multiple studies. With BEHAPP, we aim to be one of these initiatives.

However, the open-source model is not without merit. Indeed, Torous et al [1,10] regularly raise a valid and important point about the lack of interoperability and interpretability of results across the whole spectrum of digital phenotyping initiatives. The open-source model and, consequently, the free distribution of these platforms is meant to address that problem by putting the technology in the hands of many, thereby ultimately contributing to a more uniform approach toward data collection and analysis. However, we think that the solution to interoperability and comparability does not rely on a single developer, mainly because the underlying problem is an overall lack of transparency in the methodology and specifications descriptive of data collection and data processing flows. Importantly, we need scientific reports to be accompanied with metadata descriptive of the format of data (attributes, shape, size, and semantics) at every stage of collection and analysis, starting from raw data to the clinical measures that we extract. These aspects, just as security and sustainability, are largely overlooked in most scientific reports. With phenotypic outcomes currently taking center stage, we unfortunately limit ourselves in building collective knowledge toward enabling reproducible science.

Conclusions

In search of strategies for secure and sustainable digital phenotyping, we identified a gap in the available knowledge related to the establishment of secure and sustainable platforms that drive such research initiatives. Here, we address it by providing a foundation including requirements and operational guidelines focusing on key elements such as the application of encryption, data isolation, and organizational security culture. Members of ethical research boards should consider using the security principles outlined in this manuscript in their evaluation of the privacy of study participants in research proposals. Principal investigators should account for these essential components while budgeting their grant proposals, keeping in mind that security and maintenance must be adequately addressed in any research plan that includes the use of a digital phenotyping platform. Taken together, this work contributes to the foundations on which digital phenotyping strategies can be operated in a safe and sustainable way, allowing for the collection of real-time, quantitative, and longitudinal behavioral data that are expected to generate novel insights and possibly support concrete innovations in clinical care.

Conflicts of Interest

None declared.

References

1. Torous J, Onnela J, Keshavan M. New dimensions and new tools to realize the potential of RDoC: digital phenotyping via smartphones and connected devices. Transl Psychiatry 2017 Dec 07;7(3):e1053 [FREE Full text] [doi: 10.1038/tp.2017.25] [Medline: 28267146]
2. Ebner-Priemer U, Santangelo P. Digital phenotyping: hype or hope? The Lancet Psychiatry 2020 Apr;7(4):297-299. [doi: 10.1016/s2215-0366(19)30380-3]
3. Insel TR. Digital Phenotyping: Technology for a New Science of Behavior. JAMA 2017 Oct 03;318(13):1215-1216. [doi: 10.1001/jama.2017.11295] [Medline: 28973224]
4. Kargl F, van DHR, Erb B, Bösch C. Privacy in Mobile Sensing. Studies in Neuroscience, Psychology and Behavioral Economics 2019:12. [doi: 10.1007/978-3-030-31620-4_1]
5. Venters C, Lau L, Griffiths M, Holmes V, Ward R, Jay C, et al. The Blind Men and the Elephant: Towards an Empirical Evaluation Framework for Software Sustainability. Journal of Open Research Software 2014:1-6. [doi: 10.5334/jors.ao]
6. Roberts SJ. The Necessity of Information Security in the Vulnerable Pharmaceutical Industry. JIS 2014;05(04):147-153. [doi: 10.4236/jis.2014.54014]
7. Insel TR. Digital Phenotyping: Technology for a New Science of Behavior. JAMA 2017 Oct 03;318(13):1215-1216. [doi: 10.1001/jama.2017.11295] [Medline: 28973224]
8. Aledavood T, Triana HAM, Alakörkkö T, Kaski K, Saramäki J, Isometsä E, et al. Data Collection for Mental Health Studies Through Digital Platforms: Requirements and Design of a Prototype. JMIR Res Protoc 2017 Jun 09;6(6):e110 [FREE Full text] [doi: 10.2196/resprot.6919] [Medline: 28600276]
9. de Grood C, Raissi A, Kwon Y, Santana MJ. Adoption of e-health technology by physicians: a scoping review. J Multidiscip Healthc 2016;9:335-344 [FREE Full text] [doi: 10.2147/JMDH.S103881] [Medline: 27536128]
10. Torous J, Kiang MV, Lorme J, Onnela J. New Tools for New Research in Psychiatry: A Scalable and Customizable Platform to Empower Data Driven Smartphone Research. JMIR Ment Health 2016;3(2):e16 [FREE Full text] [doi: 10.2196/mental.5165] [Medline: 27150677]
11. Torous J, Wisniewski H, Bird B, Carpenter E, David G, Elejalde E, et al. Creating a Digital Health Smartphone App and Digital Phenotyping Platform for Mental Health and Diverse Healthcare Needs: an Interdisciplinary and Collaborative Approach. J. technol. behav. sci 2019 Apr 27;4(2):73-85. [doi: 10.1007/s41347-019-00095-w]
12. Berrouiguet S, Ramírez D, Barrigón ML, Moreno-Muñoz P, Carmona Camacho R, Baca-García E, et al. Combining Continuous Smartphone Native Sensors Data Capture and Unsupervised Data Mining Techniques for Behavioral Changes Detection: A Case Series of the Evidence-Based Behavior (eB2) Study. JMIR Mhealth Uhealth 2018 Dec 10;6(12):197 [FREE Full text] [doi: 10.2196/mhealth.9472] [Medline: 30530465]
13. BeHapp. URL: https://behapp.org [accessed 2021-03-17]
14. Kubiak T, Smyth J. Connecting Domains?Ecological Momentary Assessment in a Mobile Sensing Framework. Studies in Neuroscience, Psychology and Behavioral Economics.? 2019:207. [doi: 10.1007/978-3-030-31620-4_12]
15. Torous J, Kiang MV, Lorme J, Onnela J. New Tools for New Research in Psychiatry: A Scalable and Customizable Platform to Empower Data Driven Smartphone Research. JMIR Ment Health 2016;3(2):16 [FREE Full text] [doi: 10.2196/mental.5165] [Medline: 27150677]
16. Aharony N, Pan W, Ip C, Khayal I, Pentland A. Social fMRI: Investigating and shaping social mechanisms in the real world. Pervasive and Mobile Computing 2011 Dec;7(6):643-659. [doi: 10.1016/j.pmcj.2011.09.004]
17. Ferreira D, Kostakos V, Dey AK. AWARE: Mobile Context Instrumentation Framework. Front. ICT 2015 Apr 20;2:2-6. [doi: 10.3389/fict.2015.00006]
18. Jongs N, Jagesar R, van Haren NEM, Penninx BWJH, Reus L, Visser PJ, et al. A framework for assessing neuropsychiatric phenotypes by using smartphone-based location data. Transl Psychiatry 2020 Jul 01;10(1):211 [FREE Full text] [doi: 10.1038/s41398-020-00893-4] [Medline: 32612118]
19. Bilderbeck AC, Penninx BWJH, Arango C, van der Wee N, Kahn R, Winter-van Rossum I, et al. Overview of the clinical implementation of a study exploring social withdrawal in patients with schizophrenia and Alzheimer's disease. Neurosci Biobehav Rev 2019 Feb;97:87-93 [FREE Full text] [doi: 10.1016/j.neubiorev.2018.06.019] [Medline: 29940238]
20. 2017. European Medical Device Regulation Internet. URL: https://eur-lex.europa.eu/eli/reg/2017/745/2017-05-05 [accessed 2020-08-29]
21. Hasselbring W, Reussner R. Toward Trustworthy Software Systems. Computer 2006 Apr;39(4):91-92. [doi: 10.1109/mc.2006.142]
22. K. Pool J, Akhlaghpour S. Causes and Impacts of Personal Health Information (PHI) Breaches: A Scoping Review and Thematic Analysis. In: SSRN Journal. 2019 Jun Presented at: Twenty-Third Pacific Asia Conference on Information Systems; 2019; China. [doi: 10.2139/ssrn.3584865]
23. Li H, Yu L, He W. The Impact of GDPR on Global Technology Development. Journal of Global Information Technology Management 2019 Jan 24;22(1):1-6. [doi: 10.1080/1097198x.2019.1569186]
24. Mulder T, Jagesar RR, Klingenberg AM, P Mifsud Bonnici J, Kas MJ. New European privacy regulation: Assessing the impact for digital medicine innovations. Eur Psychiatry 2018 Oct;54:57-58. [doi: 10.1016/j.eurpsy.2018.07.003] [Medline: 30121506]

25. Stytz M. Considering defense in depth for software applications. IEEE Secur. Privacy Mag 2004 Jan;2(1):72-75. [doi: 10.1109/msecp.2004.1264860]
26. Bilge L, Dumitras T. Before we knew it: an empiral study of zero-day attacks in the real world. 2012 Presented at: Proceedings of the ACM conference on Computer and communications security; 2012; Raleigh, North Carolina, USA. [doi: 10.1145/2382196.2382284]
27. Wagner N, Sahin C, Winterrose M, Riordan J, Pena J, Hanson D, et al. Towards automated cyber decision support: A case study on network segmentation for security. 2016 Presented at: IEEE Symposium Series on Computational Intelligence (SSCI); 2016; Athens, Greece. [doi: 10.1109/ssci.2016.7849908]
28. Schneider F. Least privilege and more [computer security]. IEEE Secur. Privacy 2003 Sep;1(5):55-59. [doi: 10.1109/msecp.2003.1236236]
29. Smid M, Branstad D. Data Encryption Standard: past and future. Proc. IEEE 1988;76(5):550-559. [doi: 10.1109/5.4441]
30. Simmons GJ. Symmetric and Asymmetric Encryption. ACM Comput. Surv 1979 Dec;11(4):305-330. [doi: 10.1145/356789.356793]
31. Fumy W, Landrock P. Principles of key management. IEEE J. Select. Areas Commun 1993;11(5):785-793. [doi: 10.1109/49.223881]
32. Safa NS, Solms RV, Futcher L. Human aspects of information security in organisations. Computer Fraud & Security 2016 Feb;2016(2):15-18. [doi: 10.1016/s1361-3723(16)30017-3]
33. Johnson ME, Goetz E. Embedding Information Security into the Organization. IEEE Secur. Privacy Mag 2007 May;5(3):16-24. [doi: 10.1109/msp.2007.59]
34. Pfleeger S, Angela SM, Furnham A. From Weakest Link to Security Hero: Transforming Staff Security Behavior. Journal of Homeland Security and Emergency Management 2014:489-510. [doi: 10.1515/jhsem-2014-0035]
35. Höne K, Eloff J. What Makes an Effective Information Security Policy? Network Security 2002 Jun;2002(6):14-16. [doi: 10.1016/s1353-4858(02)06011-7]
36. Grassi P, Garcia M, Fenton J. Digital identity guidelines: revision 3. NIST Digital Identity Guidelines. 2017. URL: https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-63-3.pdf [accessed 2020-08-28]
37. Kirlappos I, Parkin S, Angela SM. Learning from Shadow Security: Why Understanding Non-Compliant Behaviors Provides the Basis for Effective Security. 2014 Presented at: Proceedings Workshop on Usable Security; 2014; San Diego, CA, USA. [doi: 10.14722/usec.2014.23007]
38. Approaches to software sustainability. Software Sustainability Institute. URL: https://software.ac.uk/resources/approaches-software-sustainability [accessed 2020-05-14]
39. Coding rules for maintainable software. CISQ - Consortium for Information & Software Quality. URL: https://www.it-cisq.org/standards/code-quality-standards/maintainability/index.htm [accessed 2020-05-14]
40. Lientz BP, Swanson EB, Tompkins GE. Characteristics of application software maintenance. Commun. ACM 1978;21(6):466-471. [doi: 10.1145/359511.359522]
41. Aljawarneh SA, Alawneh A, Jaradat R. Cloud security engineering: Early stages of SDLC. Future Generation Computer Systems 2017 Sep;74:385-392. [doi: 10.1016/j.future.2016.10.005]
42. Bankhead P. Reminder SMS/Call Log Policy Changes. Android Developers Blog. URL: https://android-developers.googleblog.com/2019/01/reminder-smscall-log-policy-changes.html [accessed 2020-08-05]
43. Iannino A, Musa J. Software Reliability. Advances in Computers 1990:170. [doi: 10.1016/s0065-2458(08)60299-5]
44. Fayad M, Hamza H, Sanchez H. Towards scalable and adaptable software architectures. 2005 Presented at: IRI - IEEE International Conference on Information Reuse and Integration, Conf; 2005; Las Vegas, NV, USA. [doi: 10.1109/iri-05.2005.1506457]
45. Baldini I, Castro P, Chang K, Cheng P, Fink S, Ishakian V, et al. Serverless computing: Current trends and open problems. Research Advances in Cloud Computing 2017:1-20. [doi: 10.1007/978-981-10-5026-8_1]
46. Zang H, Bolot J. Anonymization of location data does not work: A large-scale measurement study. 2011 Presented at: Proceedings of the 17th annual international conference on Mobile computing and networking; 2011; Las Vegas, NV, USA p. 145-156. [doi: 10.1145/2030613.2030630]
47. Wachter S, Mittelstadt B. A Right to Reasonable Inferences: Re-Thinking Data Protection Law in the Age of Big Data and AI. Colum. Bus. L 2019:494. [doi: 10.31228/osf.io/mu2kf]

Abbreviations

DPIA: Data Protection Impact Assessment
GDPR: General Data Protection Regulation
ISP: Information Security Policy
IT: information technology
KMS: Key Management Service
MPM: Mobile Passive Monitoring

Edited by G Eysenbach; submitted 03.06.20; peer-reviewed by U Ebner-Priemer, B Davidson; comments to author 19.07.20; revised version received 31.08.20; accepted 25.01.21; published 07.04.21.

Please cite as:
Jagesar RR, Vorstman JA, Kas MJ
Requirements and Operational Guidelines for Secure and Sustainable Digital Phenotyping: Design and Development Study
J Med Internet Res 2021;23(4):e20996
URL: https://www.jmir.org/2021/4/e20996
doi:10.2196/20996
PMID:33825695

Original Paper

Leveraging Social Media Activity and Machine Learning for HIV and Substance Abuse Risk Assessment: Development and Validation Study

Anaelia Ovalle[1], BSc; Orpaz Goldstein[1], MSc; Mohammad Kachuee[1], MSc; Elizabeth S C Wu[2], MPH; Chenglin Hong[2], BSW, MSW, MPH; Ian W Holloway[2], LCSW, MPH, PhD; Majid Sarrafzadeh[1], PhD

[1]Department of Computer Science, University of California Los Angeles, Los Angeles, CA, United States

[2]Department of Social Welfare, University of California Los Angeles, Los Angeles, CA, United States

Corresponding Author:
Anaelia Ovalle, BSc
Department of Computer Science
University of California Los Angeles
Engineering VI
404 Westwood Plaza
Los Angeles, CA, 90095
United States
Phone: 1 3108253886
Email: anaelia@cs.ucla.edu

Abstract

Background: Social media networks provide an abundance of diverse information that can be leveraged for data-driven applications across various social and physical sciences. One opportunity to utilize such data exists in the public health domain, where data collection is often constrained by organizational funding and limited user adoption. Furthermore, the efficacy of health interventions is often based on self-reported data, which are not always reliable. Health-promotion strategies for communities facing multiple vulnerabilities, such as men who have sex with men, can benefit from an automated system that not only determines health behavior risk but also suggests appropriate intervention targets.

Objective: This study aims to determine the value of leveraging social media messages to identify health risk behavior for men who have sex with men.

Methods: The Gay Social Networking Analysis Program was created as a preliminary framework for intelligent web-based health-promotion intervention. The program consisted of a data collection system that automatically gathered social media data, health questionnaires, and clinical results for sexually transmitted diseases and drug tests across 51 participants over 3 months. Machine learning techniques were utilized to assess the relationship between social media messages and participants' offline sexual health and substance use biological outcomes. The F1 score, a weighted average of precision and recall, was used to evaluate each algorithm. Natural language processing techniques were employed to create health behavior risk scores from participant messages.

Results: Offline HIV, amphetamine, and methamphetamine use were correctly identified using only social media data, with machine learning models obtaining F1 scores of 82.6%, 85.9%, and 85.3%, respectively. Additionally, constructed risk scores were found to be reasonably comparable to risk scores adapted from the Center for Disease Control.

Conclusions: To our knowledge, our study is the first empirical evaluation of a social media–based public health intervention framework for men who have sex with men. We found that social media data were correlated with offline sexual health and substance use, verified through biological testing. The proof of concept and initial results validate that public health interventions can indeed use social media–based systems to successfully determine offline health risk behaviors. The findings demonstrate the promise of deploying a social media–based just-in-time adaptive intervention to target substance use and HIV risk behavior.

(J Med Internet Res 2021;23(4):e22042) doi:10.2196/22042

KEYWORDS

online social networks; machine learning; behavioral intervention; data mining; msm; public health

Introduction

Men who have sex with men are disproportionately affected by HIV and other sexually transmitted infections. In the United States, men who have sex with men accounted for two-thirds of incident HIV infections and more than half of new syphilis diagnoses in 2018 [1-3]. Substance use has been a persistent driver of the ongoing HIV epidemic in men who have sex with men. Research suggests substance use is highly associated with high-risk sexual behaviors such as condomless anal sex, multiple sex partners, and sex trading for drugs [4-6].

Web-based communication tools such as social networking sites (eg, "hookup apps," dating websites) have been used among men who have sex with men to seek sexual partners and share information and resources about substance use [7-10]. In the early 2010s, 85% of men who have sex with men used the internet to find sexual partners [9], and this figure grew to 96% in 2019 [11]. However, the rising popularity of these technologies has also raised concerns about their role in facilitating sexual risk behaviors. Studies [12] have shown that men who have sex with men are more likely to engage in condomless anal sex with sex partners met online compared to partners met offline and have demonstrated that men who have sex with men who seek partners online have greater numbers of sexual partners compared to those who do not seek partners online. Furthermore, men who have sex with men who identify sexual partners online have a greater likelihood of substance use [13], although the evidence is equivocal [8,14,15]. Further studies are needed to provide empirical evidence of the association between online social networking technologies and offline sexual and substance use behaviors.

With more than 40% of health care consumers utilizing social media for their health-related decision making, social networks have indeed caught the attention of the public health domain [16]. Population-based analyses and in-person interventions are costly, both in time and resources [17]. Health spending is projected to grow at an average rate of 5.5% per year, totaling $6.0 trillion by 2027, nearly one-fifth of the United States gross domestic product [18]. Given these costs, the opportunity for high user engagement, and accessibility of social media data, social networks provide a new opportunity to public health. In recent years, social media have been employed in behavioral and public health research and has demonstrated its effectiveness in prevention, education, and treatment [19,20]. For instance, analyzing user activity on social media platforms has been an effective way to estimate the risk and time of HIV infection [21]. A separate study [22] found that the strength of associations in a social network, its network shape, and size are predictors of HIV and sexually transmitted infection risk.

Public health studies have also begun using social media platforms to understand and intervene in sexual health and substance use risk behaviors among men who have sex with men; however, these methodologies remain nascent in that they still rely on self-report and costly data collection as a means of developing and testing interventions [23-25]. Research demonstrates the need to utilize big data in social media and machine learning to understand communication and patterns about substance use and observe and predict real-time risk behaviors [26-28]. These strategies can inform just-in-time adaptive interventions [29] that are responsive to the individual technology use patterns of research participants; however, before deploying any social media–based intervention, the feasibility and efficacy behind employing such a modality for public health initiatives in risk reduction must be determined.

In practice, adaptive intervention systems are driven by an ability to determine health risks. Placed in the context of social media data, it is encouraging to learn that assessing health risk using textual sources has shown promising results in disease-specific risk evaluation and in identifying individuals at higher risk of depression and self-harm [30-32]. While text collection from electronic health records is extremely effective in determining a diagnosis, it is not a readily available resource for continuous risk assessment. Meanwhile, social media text data have the advantage of being abundantly available and cost-effective. While these data are not as domain constrained as clinical notes, they remain promising channels to explore for risk assessment.

Additionally, system interventions should be able to accurately evaluate when an individual is about to engage in a targeted health risk behavior with high probability followed by successfully reducing such behaviors. Maher et al [33] reviewed the effectiveness of past social network interventions, concluding with a call for stronger evidence in interventions that incorporate online social networks. Our paper responds to this call by evaluating the efficacy of social media data in determining health risk behavior. We are guided by the following questions: (1) Can we further substantiate the association between online social networking technologies and offline sexual and substance use behaviors? (2) Can we extract health risk scores from social media data that align with public health expert evaluation?

In this paper, the practicality of social media as an intervention modality is evaluated through social media data to identify health risk behavior in a sample of men who have sex with men from Los Angeles, California. The contributions of this paper are the following: (1) an end-to-end platform that continuously collects data from common social media platforms and specialized social networks tailored to the men who have sex with men community, and in tandem, biological data and personal health questionnaires were collected at baseline, 1-month, and 3-months from intake; (2) health behavior risk scores that are comparable to adapted risk scores created by the Centers for Disease Control and Prevention (CDC) using natural language processing techniques; and (3) the application of machine learning techniques to determine the extent to which social media messages can be used to directly predict verified biological outcomes of substance use and sexual risk, reflected as sexually transmitted disease diagnoses.

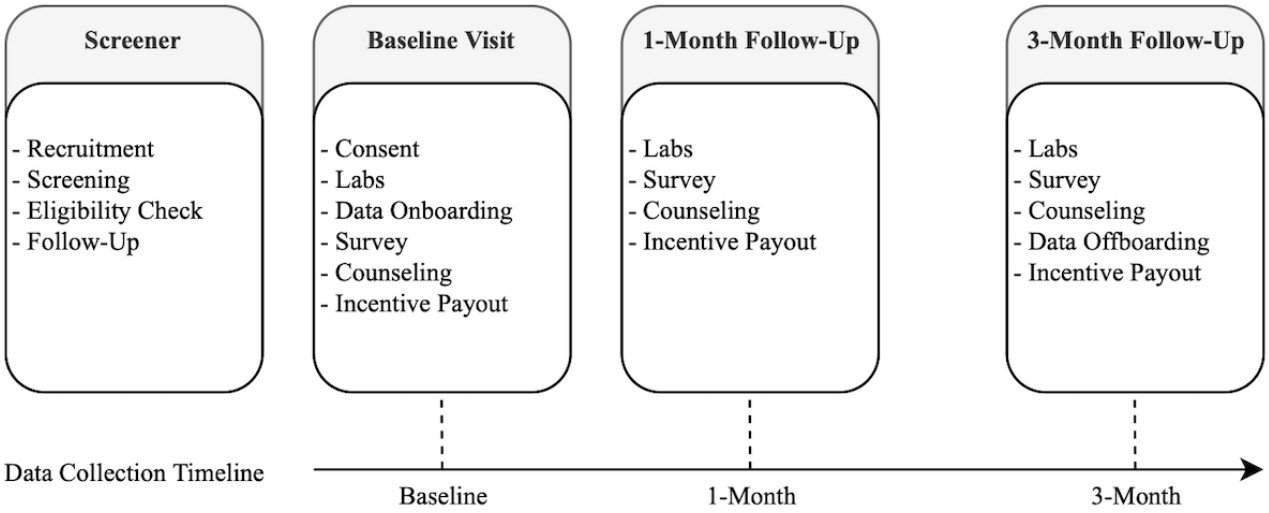
Methods

Study Protocol

The study protocol consisted of 4 milestones: (1) screener, (2) baseline visit, (3) 1-month follow-up, and (4) 3-month follow-up

(Figure 1). In the screener, flyers, online advertisements, and referrals were used to recruit potential participants. Further screening took place over the phone or through an web-based survey. Criteria such as age, sexual orientation, substance use, online dating, and social media activity were used to determine each participant's eligibility for the study.

Figure 1. Different phases of the study protocol and data collection.



Qualified participants were invited to an initial clinical visit to review the study in detail, ask for informed consent, and answer any of their remaining questions. Afterward, a series of lab tests were conducted to determine their substance use and the presence of sexually transmitted diseases. Site testing was conducted for *Chlamydia trachomatis* and *Neisseria gonorrhoeae* with pharyngeal, urethral, and anal swabs. Further tests included a rapid plasma reagin blood test for syphilis, a rapid oral test for HIV, and a urine drug screen. Additionally, a survey was completed by the participants which asked a series of questions regarding demographic characteristics, sexual risk behavior, illicit substance use, and online behavior. Finally, participants provided their log-in credentials for a set of social media sites on which they had been. The user credentials were registered in a custom data collection platform for each website and the participants authorized the data collection system to pull their daily online activity. We collected participant social media data for up to 3 months after onboarding. We found this to be a reasonable duration considering the need to follow participants long enough to observe any changes in social media use and behaviors over time that can be measured by follow-up surveys and their biomarkers.

The system began collecting participants' daily social media activity immediately after the baseline visit. One month into the study, they were scheduled to revisit the clinic and redo lab tests and surveys. A final follow-up was set for 3 months after the baseline visit to recollect lab and survey data in addition to conducting required off-boarding procedures, including the discontinuation of participant data collection.

This study protocol was approved by the University of California, Los Angeles institutional review board (IRB 17-000408). Informed consent was obtained from all individual participants included in the study. Each participant was provided with up to US $150 in cash incentives based on their participation. Certain medical conditions identified within this study were also reported to appropriate agencies as required by federal and state laws. Proper consultation and referrals were provided to each participant before and after reporting each screening test.

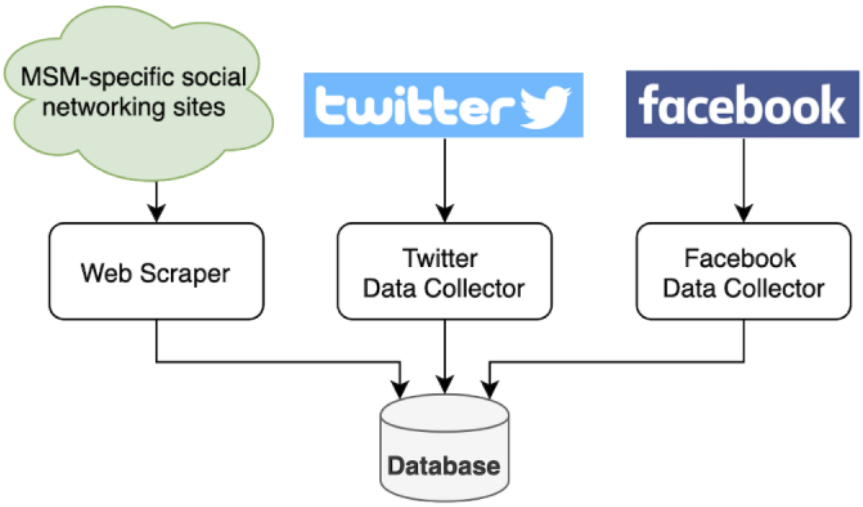
System Architecture

Data from Facebook and Twitter was automatically collected through official application programming interfaces (API). The APIs were used to query the content (eg, messages, posts) generated by each participant; however, it is important to note that Facebook changed its API policy during the study. Standard permission requests could no longer be used to access user data. To mitigate this issue and re-enable the permission to read user content, the participants enrolled as testers of a custom Facebook app developed during the study.

In addition to these major social media platforms, we collected similar data fields for men who have sex with men–specific social networks. The software implementation slightly differed from that of Facebook and Twitter in that a unique web scraper was built to collect data. Due to the study's privacy protocol, the name of each site was omitted. The websites, labeled Website A and Website B, are one of the most established websites in the men who have sex with men community, with over 10 million users overall. Qualitative interviews and input from a community advisory board further informed the decision to include these websites. Each scraper used a combination of browsing automation tools and web parsers to mimic a user going online and accessing their profile. URL requests were made using Selenium, a browsing tool that can conduct automatic authentication, search, and navigate each site of interest [34]. Participants explicitly provided consent and their usernames and passwords per institutional review board protocol approvals. These credentials were used by Selenium to automate logging into a website and navigating to pages where the system was able to collect profiles and messages between users. The data were extracted from each website of interest using the BeautifulSoup Python Package and later saved to the database [35]. Figure 2 provides a visualization of the study's overall

database schema. MySQL, a free and open-source database server, was used to store and query user messages and metadata including usernames and respective access tokens from each website. To maintain high data-quality, a hashing of the data and timestamp was created within the database to prevent duplication.

Figure 2. A visualization of the database schema used in this study. MSM: men who have sex with men.



We validated that the system collected participant messages consistently and reliably by maintaining a group of test users. Before and throughout the study, various mixed-media data were sent across these platforms to make sure all types of generated data were correctly collected and accurately represented. The data consisted of text, videos, and emojis.

Due to the sensitive nature of collecting online social media activity, providing security and privacy was of paramount importance. Therefore, the data collection system was designed to offer protection on multiple levels: at the system-level and the data-level. At the system-level, the user data were stored on a dedicated private server behind a firewall, protecting against outside cyber-attacks and malicious software. At the data-level, the user data went through a data sanitation process. Participants' identifiers in the data collection system were anonymized and could only be matched with a reference datasheet stored externally on a cloud storage service compliant with the Health Insurance Portability and Accountability Act (HIPAA). The cloud storage service was also used to store all other data, such as participant questionnaires and drug screening results.

Defining Health Behavior Risk

Given the ambiguous nature of identifying health risk behavior in social media messages, a map between commonly used colloquial terminology and risk behavior topics was curated. To gain a more nuanced understanding of online strategies and behaviors around seeking drugs and sexual partners, 24 men who have sex with men community members in Los Angeles who self-disclosed risky sex, illicit drug use, and usage of dating apps were qualitatively interviewed. The response to the following question became the foundation for the risk behavior dictionary: “What are some of the terms that you use on websites, chatrooms, message boards, and apps to find drugs, drug use partners, and/or sex partners?” The words were organized into topics (Table 1). All terms were vetted by a community advisory board throughout the study. The resulting risk behavior dictionary was leveraged to facilitate the modeling of the relationship between social media messages and health behavior risk.

Sexual health and illicit drug use risk was also evaluated for each participant using the CDC Risk Assessment Tool [36]. Each assessment was adjusted based on participants' weekly text message diary responses on illicit substance use. Scores for each participant were available at baseline intake, 1-month, and 3-month checkpoints.

Table 1. Sexual and substance use topics associated with colloquial terms found in text conversations based on qualitative interviews with men who have sex with men community members.

Topic	Words
Alcohol	drunk, drinking, booze, party, liquor
Marijuana	toke, pipe, weed, pot, mary jane
Cocaine	crack, blow, snow, yayo, powder
Methamphetamine	meth, speed, ice, crank, crystal
Amyl nitrate	poppers, rush, pops, amyl
Heroin	dope, smack, junk, tar
Ecstasy	X, Ex, molly, rolling, mdma
GHB	G, roll, water
Ketamine	K
Type of substance other/general	pills, favors, DMT, party, party favors
Snorting	snort, sniff, rack, rail, lines
Inhaling	smoke, blow clouds, hit, puff
Injection	straight to the point, straight to the poinT, slam, slamming, shoot, shooting
Anal insertion	booty bumping, butt rocket, plug, plugging
Substance use behavior other/general	gen, generous, friends with benefits
Buying drugs	Do you have a connect, I can contribute, can you do me a big/little/huge favor
Masturbation	JO, jack, stroke, HJ, jack off
Oral sex	blow, head, gloryhole, suck, BJ
Anal intercourse	top, bottom, fuck, power top, power bottom
Group sex	3some, 3way, gang bang, orgy, bukkake
Sex work	$, roses, generous, pro, GEN
Anonymous	anon, discreet, discrete, anonymous, random play
Sex with condoms	condoms, rubber, safe sex, play safe, safe
Condomless sex	Bareback, bare, raw, seed, seeding
Substance use and sex	Party and play, smoke and stroke, pnp, party, partying
Sexual behavior general/other	69

Data Set Processing and Auditing

Data were cleaned to maximize consistency and accuracy. The messages were first screened using Python regular expressions to pattern match texts identified as spam or in-app advertisements. Pattern personal identifiable information such as addresses and phone numbers were tokenized. For instance, if a phone number was provided in a message, it was replaced with “phonenumbertoken.” Data were deduplicated and the privacy of those that communicated with our participants was protected. This was done by only considering the participants' sent messages, resulting in data for 48 individuals. Some messages also consisted of notifications such as when a participant's profile was seen, clicked on, or had a request to unlock their photos—all of which were tokenized with the suffix “token.” One data source in the study was excluded because it failed to produce data due to changes in the site's data collection policy. Furthermore, we focus our analysis on biomarkers that were present in at least 10% of participants.

After data cleaning, the messages went through an automated pattern matching pipeline using regular expressions to identify terms in the risk behavior dictionary defined in Table 1. The mapping was utilized to flag words and colloquial terms associated with health risk behaviors in each message. As an example, if the word in a given message was “rail,“ it was matched to the topic of snorting.

Biomarker Prediction With Social Media Messages

The relationship between participants' social media behavior and illicit substance use and sexual risk behaviors was examined; data collected across social media accounts were leveraged to predict participant’s respective offline substance use and sexual health biomarkers.

Standardized counts of each tokenized word and text-summary features, such as message length, were used as simple features to predict drug use and sexually transmitted diseases. Each outcome was treated as its own binary classification task. Logistic regression, linear support vector machine, naïve Bayes,

and random forest models were employed to predict the outcomes at 1-month and 3-month follow-up. Given the relatively small data set and the challenge of class imbalance, stratified 5-fold cross-validation was used to assess the generalizability of the predictive models. The performance of each model was assessed using the precision, defined as *true positive* / (*true positive* + *false positive*); recall, defined as *true positive* / (*true positive* + *false negative*); and F1 score, defined as 2 × *precision* × *recall* / (*precision* + *recall*).

Message-Based Risk Scores

Health risk behavior was assessed on a per-message level using available social media text correspondence. A risk score was given for each message based on how likely its words were associated with those in the risk behavior dictionary.

Identifying Health Risk Behaviors Using Social Media Messages

Natural language processing techniques were employed to create a risk score using social media data. We employed a Skip-Gram Word2Vec model which allowed us to extract word representations to determine the association between words in social media messages and words in the risk behavior dictionary [37]. The Skip-Gram Word2vec model constructs a word representation, or word embedding, based on how well it predicts the words that surround it within a given radius. Given all the words in the message corpus, w_1, w_2, w_3,..., w_T, the model tries to predict the probability of observing the context word w_{t+j} given a target word w_t.

The resulting model constructs a vector space that identifies context-dependent similarity between words used by the participants in the selected social networks monitored.

Health Behavior Risk Scoring

Once contextually similar instances of risky text messages are found, a risk score is constructed with cosine similarity, similar to the approach used by Kiros et al [38]. Given user *i*'s *t*th message, the risk score is defined as the average cosine similarity between the word in the message,  , and the risky word *w* from each risk topic *d* set of risky words.

The distance between the words found in messages and those from the risk behavior dictionary was used to decide when a user was displaying a risky textual correspondence. Afterward, similarity between the risk scores generated by the model and those provided by the CDC risk assessment tool was assessed for each participant.

Results

Data

A total of 15,695 sent messages were collected in the 3-month timeline for 48 participants across 4 different platforms: Twitter, Facebook, Website A, and Website B after preprocessing. 6.5% (1026) of the messages were advertisements filtered out by our data processor. As shown in Table 2, the majority of messages, 75.7% (11,877), came from Website A, followed by Website B with 21.2% (3327), Facebook at 2.2% (352), and Twitter with 0.89% (139) of messages. Participant activity across all platforms is displayed as a heatmap in Figure 3, with participants 28, 40, and 42 showing the highest activity after initial onboarding.

The distribution of each topic is visualized in Figure 4. The topics with the most discussion across the social media sites were anal intercourse and methamphetamines use. In the messages themselves, tokens such as "clickedprofiletoken," "unlockedphotostoken," and "phonenumbertoken" were the highest occurring tokens across all messages. This makes sense, as activity in dating sites is heavily based on interacting with photos and other information exchange, potentially leading up to an offline connection.

In terms of clinical data, the distribution of each outcome can be seen in Table 3, where the majority of selected outcomes suffer from a 10% to 15% imbalance.

Table 2. Messages sent across social media platforms by participants.

Source	Messages sent, n (%[a])	Cumulative messages sent, n (%[a])
Twitter	139 (0.9)	139 (0.9)
Facebook	352 (2.2)	491 (3.1)
Website[b] A	3327 (21.2)	3818 (24.3)
Website[b] B	11,877 (75.7)	15,695 (100)

[a]Percentage of total messages.

[b]Specific to men who have sex with men.

Figure 3. A heatmap of user activity across all sources of data during the study.

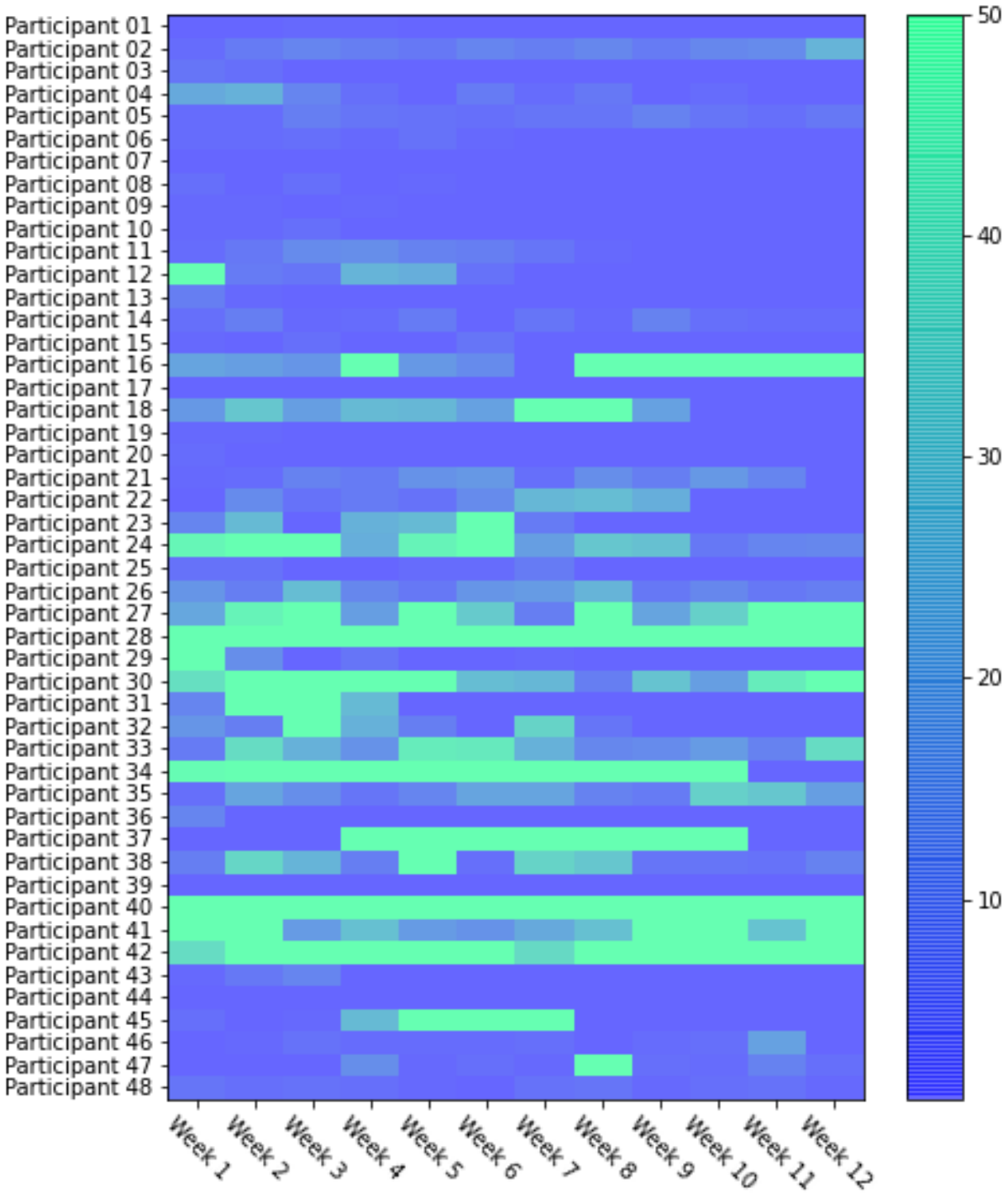


Figure 4. Topics found in messages using the risk behavior word dictionary provided by public health experts.

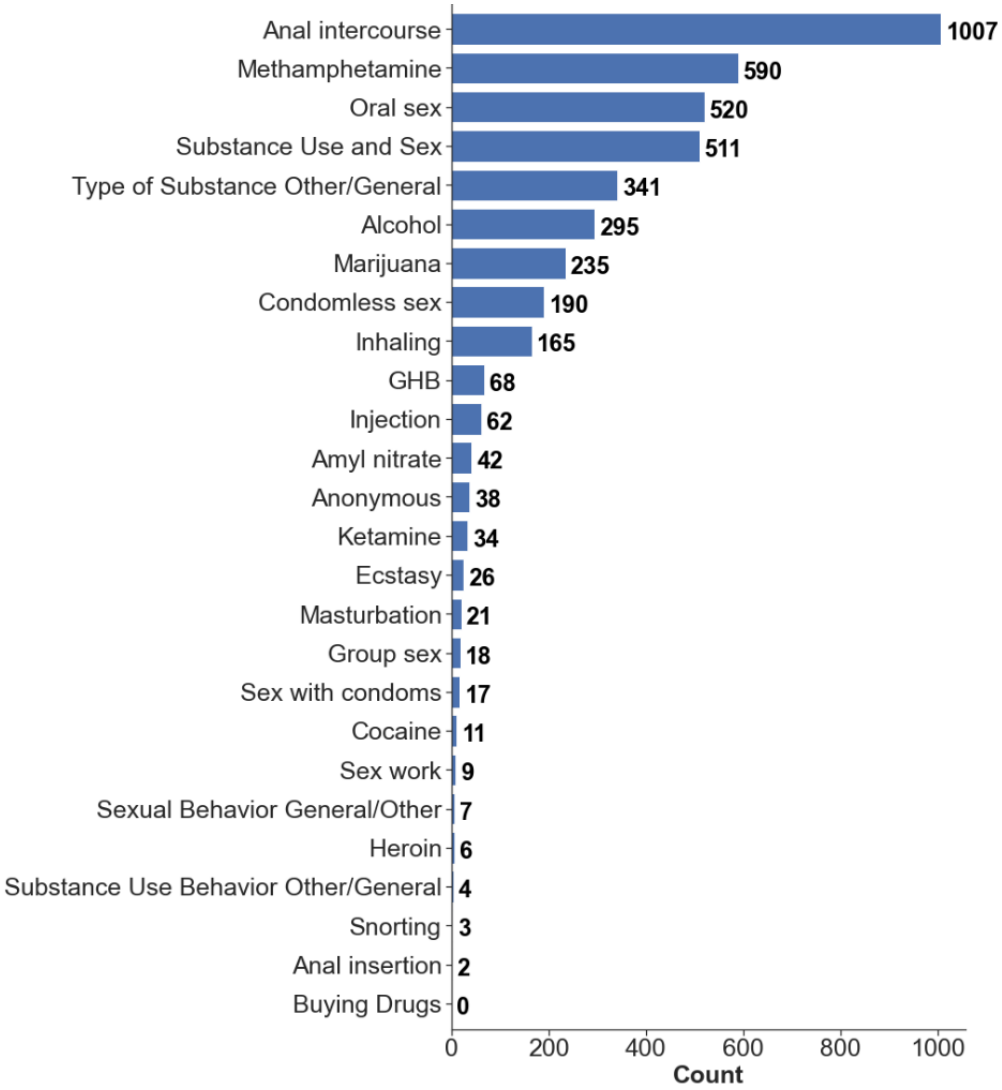


Table 3. Distribution of biomarker outcomes in a 1-month and 3-month follow-up after participants' initial onboarding.

Test and outcome	1-month follow-up (n=48), n (%)	3-month follow-up (n=48), n (%)
Sexually transmitted disease		
HIV[a]		
Positive	20 (41.7)	23 (47.9)
Negative	28 (58.3)	25 (52.1)
Substance use		
Amphetamine		
Positive	24 (50.0)	21 (43.8)
Negative	24 (50.0)	27 (56.2)
Methamphetamine		
Positive	25 (52.1)	22 (45.8)
Negative	23 (47.9)	26 (54.2)
THC[b]		
Positive	20 (41.7)	18 (37.5)
Negative	28 (58.3)	30 (62.5)

[a]HIV: human immunodeficiency virus.

[b]THC: tetrahydrocannabinol.

Biomarker Prediction

Of the 4 biomarkers, only 3 reflect an F1 score greater than 80%—HIV, amphetamine, and methamphetamine (Table 4). Although the tetrahydrocannabinol outcome did not suffer from severe class imbalance, one reason that may explain the significantly poorer performance is that the topic was heavily impacted by polysemy. As an example, participants alluded to marijuana usage with words and phrases that had multiple meanings. The phrase "blowing clouds" could refer to smoking marijuana or smoking methamphetamines.

For these 3 offline biomarker outcomes, the random forest model resulted in the highest F1 scores. Amphetamine usage was the best predicted outcome for all 4 models.

Table 4. F1 scores for each respective model and outcome in a 1-month and 3-month follow-up period after participants' initial onboarding.

Outcome and model	F1 Score, %	
	1-month follow-up	3-month follow-up
HIV[a]		
Support vector machine	70.5	81.6
Random forest	73.4	82.6
Naïve Bayes	69.6	82.1
Logistic regression	68.5	81.4
Amphetamines		
Support vector machine	88.2	85.9
Random forest	88.3	85.9
Naïve Bayes	88.0	85.5
Logistic regression	88.1	85.8
Methamphetamines		
Support vector machine	88.3	85.1
Random forest	88.3	85.3
Naïve Bayes	88.1	84.8
Logistic regression	88.2	85.0
THC[b]		
Support vector machine	11.1	7.4
Random forest	5.0	0.5
Naïve Bayes	24.0	4.7
Logistic regression	10.9	9.1

[a]HIV: human immunodeficiency virus.

[b]THC: tetrahydrocannabinol.

Associating Words With Health Behavior Risk

A dimensionality reduction technique, *t*-distributed stochastic neighbor embedding, was used to observe the relationship between word embeddings [39]. Figure 5 shows the resulting vector space as words are projected onto a 2D plane. Intuitively, the resulting proximity between words can be interpreted as their similarity. As a result, natural groupings between drug and sex-related words form. For instance, the word "Smurff," a term used to describe oral sex, is grouped with the term "head," a known colloquial term for oral sex.

Health behavior risk scores were calculated using the Word2Vec word vector space. Risk scores were created each day for a given user in addition to their average risk score along with risk scores provided by public health experts. While the constructed daily scores vary across time, there is a visible similarity between the generated risk score and expert scores, on average (Figure 6).

Figure 5. A t-distributed stochastic neighbor embedding visualization of the resulting vector space after running Word2Vec: drug-related words, sex-related words, and randomly chosen words.

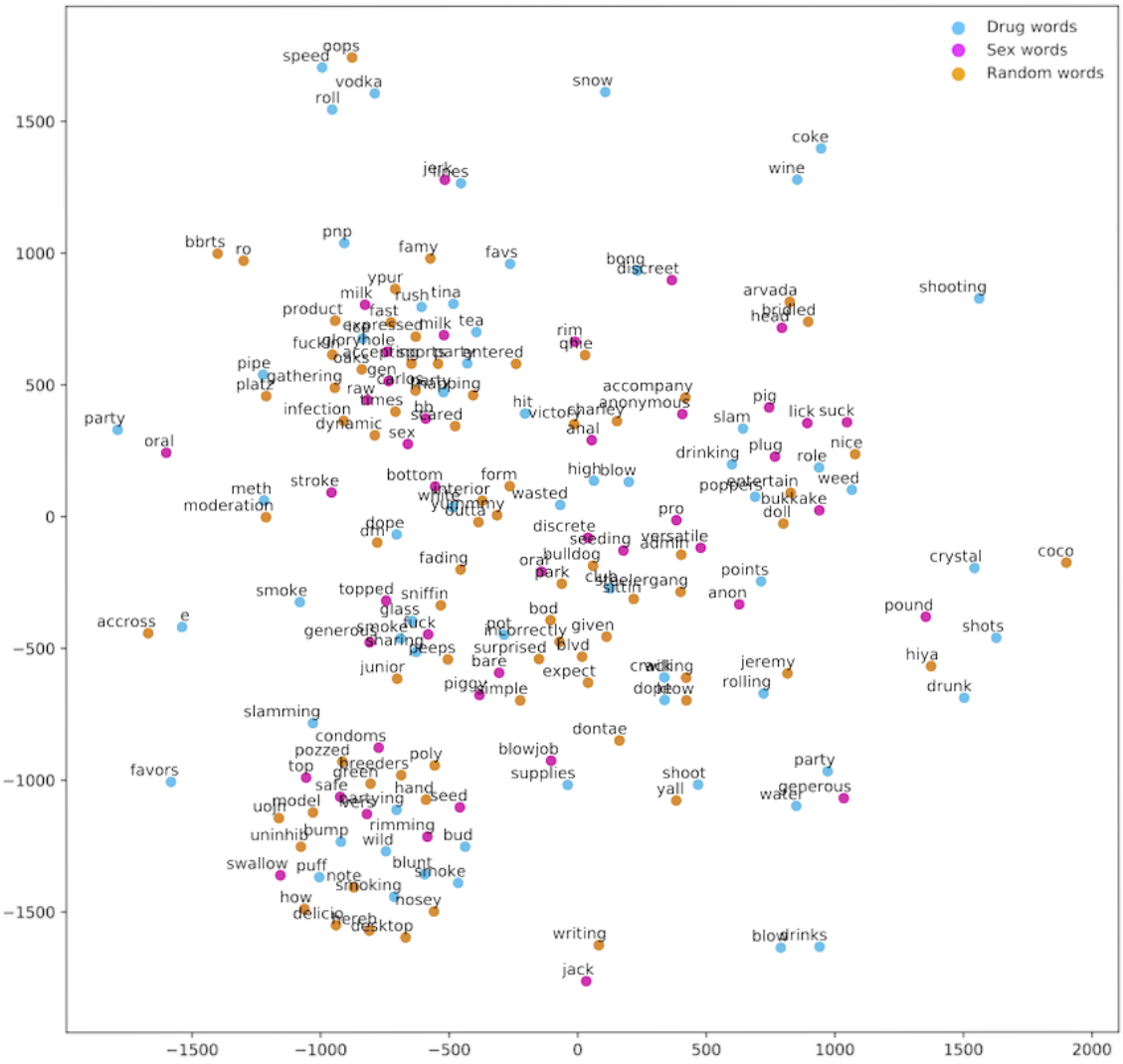
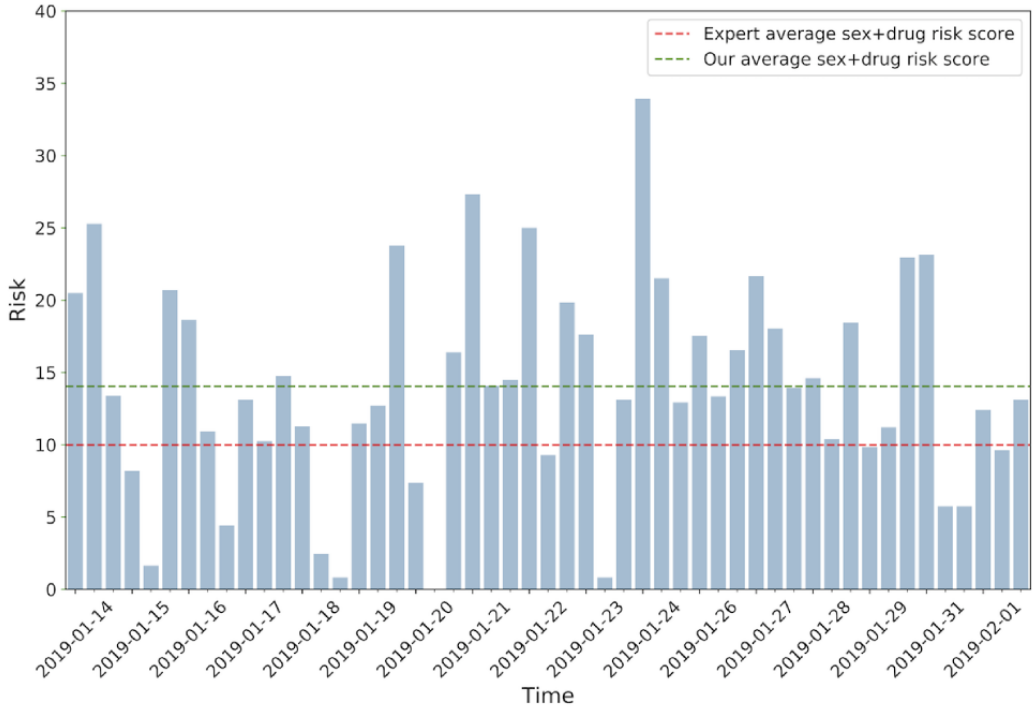


Figure 6. Normalized expert risk evaluation for drug and sex at 1 month and 3 months.



Discussion

Principal Results

To our knowledge, our study is the first empirical evaluation of a social media–based public health intervention framework for men who have sex with men. Our qualitative work highlighted the ways in which men who have sex with men use coded language online to refer to specific substance use behavior and HIV risk. We investigated the association between social media data and offline health risk behavior by operationalizing social media data across several networking sites. We built a system that automated social media data collection, which allowed us to predict offline substance use and sexual health biomarkers and construct daily health behavior risk scores.

In conducting an exploratory data analysis to check for data quality, we validated that the system was indeed able to collect participant messages consistently and reliably using the proposed system architecture. It was reasonable to observe that most messages came from men who have sex with men–specific social networks, as individuals may be more reserved around topics related to health risk behaviors over platforms with a larger audience such as Facebook and Twitter. Furthermore, we observed that after the topic of anal intercourse, the topic of methamphetamines was the most frequent in risk behavior conversation (Figure 5).

We tested the extent to which social media data could provide meaningful insight into health risk behaviors by predicting offline sexual health and substance use biomarkers. We found that across the models and clinical timelines, there was a consistently high F1 score when predicting HIV, amphetamine, and methamphetamine use. These results are validated by the fact that methamphetamines are one of the most commonly used drugs in among men who have sex with men [5]. Moreover, these findings align with existing literature pointing to the association between methamphetamine use and increased risk for HIV [4-6,10,13]. Most importantly, we validate the significance of social media data in its association with substance use and sexual health biomarker outcomes.

Daily health risk scores were created for participants using only social media data. The method was validated by the observation that the constructed risk scores, on average, were comparable to adapted risk scores created by the CDC risk assessment tool. The results were promising at a fundamental level, as we observed how words in social media messages can indeed cluster with known drug and sex-related words when mapped to a vector space (Figure 5). In constructing an average risk score for participants, we used their text messages and respective contexts to extract risk; we assumed that the public health risk dictionary was most accurate and encompassing of health risk behavior terminology. The health behavior risk scores were constructed on a daily level. Only social media data were used to construct the risk score, while public experts used biologically verified results and self-reported data to construct monthly risk scores by adapting a version of the CDC risk assessment tool. Yet, even with this significant difference in methodology, the average risk was still quite similar to the expertly assessed risk. This suggests that the community-based participatory approach used to create the data dictionary was crucial to the risk score creation process. Hence, the success of using such an approach heavily relies on involving end users in the creation of social media tools.

Limitations

The technical setup and maintenance needed for a social media data collection platform is an important consideration for scalability. Extensive effort and resources were required to achieve HIPAA compliance for our system. Additionally,

creating the data collector required custom handling of each data source. For instance, Facebook's policy change called for a completely different approach to data collecting in the middle of the study. To provide a scalable solution, the data collection platform should be flexible enough to adjust to a new collection regime due to circumstances outside of the study's control.

Several outcomes did not pass our exclusion criteria filter at the beginning of the study due to a significant imbalance in their respective distributions, leading to challenges for biomarker prediction. Additionally, employing methods that handle class imbalance may improve our current F1 scores. At a higher level, it is also important to note that the scope of this paper is somewhat limited since we only collected data from a few websites, which may not be used by the entire community of men who have sex with men.

Implications

Self-report is a common form of feedback for health behavior interventions. While useful, it is often flawed because of biases such as recall bias and social desirability bias. Therefore, the health behavior data in this study were combined with biologically verified data to calculate risk. We found that social media may serve as a valid intervention modality as it provides both valuable and relevant feedback. Overall, we determined that combining self-report data with biologically verified outcomes and social media data mining gives a more nuanced and accurate picture of health behavior risk. These findings suggest that it is feasible to use social media data for future public health intervention.

Conclusions

In this study, we created an end-to-end system that leverages social media data for health behavior risk identification, serving as a proof of concept for social media–based behavioral intervention. We demonstrated that it is possible to build an integrative system across multiple platforms that effectively collects meaningful social media data. We determined that social media messages are a valuable source of examining the relationship between health risk behaviors and biologically verified sexual diseases such as HIV and illicit usage of amphetamines and methamphetamines among men who have sex with men. Adapted CDC health risk scores were compared against social media–based behavioral risk scores and found to be, on average, similar to the expertly assessed scores. This validates the feasibility of employing a social media–based behavioral intervention. The contributions made in this paper are stepping stones toward building an automated, cost-effective, fully scalable social media intervention system that serves the public health domain.

Acknowledgments

We thank all the participants of this study for their time and cooperation. We are grateful for our collaborators and researchers. We would like to thank Christian Paneda, Omkar Patil, Tom Weng, and Amrutha Srinivasan for helping to maintain the data collection system. IWH and ESCW are supported by the California HIV/AIDS Research Program (RP15-LA-007) and the University of California, Los Angeles Center for HIV Identification, Prevention, and Treatment Services (P30 MH058107). This study was also partially supported by National Institute on Drug Abuse (R03 DA039752).

Conflicts of Interest

None declared.

References

1. 2018 STD surveillance report. Centers for Disease Control and Prevention. 2019 Nov 12. URL: https://www.cdc.gov/nchhstp/newsroom/2019/2018-STD-surveillance-report.html [accessed 2020-11-12]
2. HIV surveillance reports. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/hiv/library/reports/hiv-surveillance.html [accessed 2020-11-05]
3. Purcell DW, Johnson CH, Lansky A, Prejean J, Stein R, Denning P, et al. Estimating the population size of men who have sex with men in the United States to obtain HIV and syphilis rates. Open AIDS J 2012 Sep 7;6(1):98-107 [FREE Full text] [doi: 10.2174/1874613601206010098] [Medline: 23049658]
4. Buchacz K, McFarland W, Kellogg T, Loeb L, Holmberg SD, Dilley J, et al. Amphetamine use is associated with increased HIV incidence among men who have sex with men in San Francisco. AIDS 2005 Sep 02;19(13):1423-1424. [doi: 10.1097/01.aids.0000180794.27896.fb] [Medline: 16103774]
5. Hoenigl M, Chaillon A, Moore D, Morris S, Smith D, Little S. Clear links between starting methamphetamine and increasing sexual risk behavior. J AIDS 2016;71(5):551-557. [doi: 10.1097/qai.0000000000000888]
6. Berg RC, Weatherburn P, Marcus U, Schmidt AJ. Links between transactional sex and HIV/STI-risk and substance use among a large sample of European men who have sex with men. BMC Infect Dis 2019 Aug 05;19(1):686-689 [FREE Full text] [doi: 10.1186/s12879-019-4326-3] [Medline: 31382923]
7. Grov C, Breslow AS, Newcomb ME, Rosenberger JG, Bauermeister JA. Gay and bisexual men's use of the internet: research from the 1990s through 2013. J Sex Res 2014;51(4):390-409 [FREE Full text] [doi: 10.1080/00224499.2013.871626] [Medline: 24754360]

8. Bien CH, Best JM, Muessig KE, Wei C, Han L, Tucker JD. Gay apps for seeking sex partners in China: implications for MSM sexual health. AIDS Behav 2015 Jun 9;19(6):941-946 [FREE Full text] [doi: 10.1007/s10461-014-0994-6] [Medline: 25572834]
9. Rosser BRS, Wilkerson JM, Smolenski DJ, Oakes JM, Konstan J, Horvath KJ, et al. The future of Internet-based HIV prevention: a report on key findings from the Men's INTernet (MINTS-I, II) Sex Studies. AIDS Behav 2011 Apr 1;15 Suppl 1(S1):S91-100 [FREE Full text] [doi: 10.1007/s10461-011-9910-5] [Medline: 21360127]
10. Bourne A, Reid D, Hickson F, Torres-Rueda S, Weatherburn P. Illicit drug use in sexual settings ('chemsex') and HIV/STI transmission risk behaviour among gay men in South London: findings from a qualitative study. Sex Transm Infect 2015 Dec 09;91(8):564-568. [doi: 10.1136/sextrans-2015-052052] [Medline: 26163510]
11. Shrestha R, Lim SH, Altice FL, Copenhaver M, Wickersham JA, Saifi R, et al. Use of smartphone to seek sexual health information online among Malaysian men who have sex with men (MSM): implications for mHealth intervention to increase HIV testing and reduce HIV risks. J Community Health 2020 Feb 2;45(1):10-19 [FREE Full text] [doi: 10.1007/s10900-019-00713-x] [Medline: 31375976]
12. Grosskopf NA, LeVasseur MT, Glaser DB. Use of the Internet and mobile-based "apps" for sex-seeking among men who have sex with men in New York City. Am J Mens Health 2014 Nov 20;8(6):510-520 [FREE Full text] [doi: 10.1177/1557988314527311] [Medline: 24658284]
13. Taylor MM, Aynalem G, Smith LV, Montoya J, Kerndt P. Methamphetamine use and sexual risk behaviours among men who have sex with men diagnosed with early syphilis in Los Angeles County. Int J STD AIDS 2007 Feb 25;18(2):93-97 [FREE Full text] [doi: 10.1258/095646207779949709] [Medline: 17331279]
14. Liau A, Millett G, Marks G. Meta-analytic examination of online sex-seeking and sexual risk behavior among men who have sex with men. Sex Transm Dis 2006 Sep;33(9):576-584. [doi: 10.1097/01.olq.0000204710.35332.c5] [Medline: 16540884]
15. Zhang D, Bi P, Lv F, Zhang J, Hiller JE. Differences between internet and community samples of MSM: implications for behavioral surveillance among MSM in China. AIDS Care 2008 Oct 29;20(9):1128-1137. [doi: 10.1080/09540120701842829] [Medline: 18825519]
16. Warden C. ReferralMD. URL: https://getreferralmd.com/2017/01/30-facts-statistics-on-social-media-and-healthcare/ [accessed 2020-07-03]
17. Pagoto S, Waring ME, May CN, Ding EY, Kunz WH, Hayes R, et al. Adapting behavioral interventions for social media delivery. J Med Internet Res 2016 Jan 29;18(1):e24 [FREE Full text] [doi: 10.2196/jmir.5086] [Medline: 26825969]
18. Health care costs 101: spending keeps growing. California Health Care Foundation. 2020. URL: https://www.chcf.org/wp-content/uploads/2019/05/HealthCareCostsAlmanac2019.pdf [accessed 2019-05-20]
19. Lemley SM, Klausner JD, Young SD, Stafylis C, Mulatya C, Oden N, et al. Comparing web-based platforms for promoting HIV self-testing and pre-exposure prophylaxis uptake in high-risk men who have sex with men: protocol for a longitudinal cohort study. JMIR Res Protoc 2020 Oct 19;9(10):e20417-e20417 [FREE Full text] [doi: 10.2196/20417] [Medline: 33074164]
20. Rhodes SD, McCoy TP, Tanner AE, Stowers J, Bachmann LH, Nguyen AL, et al. Using social media to increase HIV testing among gay and bisexual men, other men who have sex with men, and transgender persons: outcomes from a randomized community trial. Clin Infect Dis 2016 Jun 01;62(11):1450-1453. [doi: 10.1093/cid/ciw127] [Medline: 26980878]
21. Christley R, Pinchbeck G, Bowers R, Clancy D, French NP, Bennett R, et al. Infection in social networks: using network analysis to identify high-risk individuals. Am J Epidemiol 2005 Nov 15;162(10):1024-1031. [doi: 10.1093/aje/kwi308] [Medline: 16177140]
22. Drumright LN, Frost SDW. Rapid social network assessment for predicting HIV and STI risk among men attending bars and clubs in San Diego, California. Sex Transm Infect 2010 Dec 21;86 Suppl 3(Suppl 3):iii17-iii23. [doi: 10.1136/sti.2010.045914] [Medline: 20966457]
23. Holloway IW, Pulsipher CA, Gibbs J, Barman-Adhikari A, Rice E. Network influences on the sexual risk behaviors of gay, bisexual and other men who have sex with men using geosocial networking applications. AIDS Behav 2015 Jun 9;19 Suppl 2(S2):112-122 [FREE Full text] [doi: 10.1007/s10461-014-0989-3] [Medline: 25572832]
24. Holloway IW. Substance use homophily among geosocial networking application using gay, bisexual, and other men who have sex with men. Arch Sex Behav 2015 Oct;44(7):1799-1811 [FREE Full text] [doi: 10.1007/s10508-015-0581-6] [Medline: 26216146]
25. Wang H, Zhang L, Zhou Y, Wang K, Zhang X, Wu J, et al. The use of geosocial networking smartphone applications and the risk of sexually transmitted infections among men who have sex with men: a systematic review and meta-analysis. BMC Public Health 2018 Oct 16;18(1):1178 [FREE Full text] [doi: 10.1186/s12889-018-6092-3] [Medline: 30326887]
26. Centola D. An experimental study of homophily in the adoption of health behavior. Science 2011 Dec 2;334(6060):1269-1272 [FREE Full text] [doi: 10.1126/science.1207055] [Medline: 22144624]
27. Eichstaedt JC, Smith RJ, Merchant RM, Ungar LH, Crutchley P, Preoţiuc-Pietro D, et al. Facebook language predicts depression in medical records. Proc Natl Acad Sci U S A 2018 Oct 30;115(44):11203-11208 [FREE Full text] [doi: 10.1073/pnas.1802331115] [Medline: 30322910]

28. Kalyanam J, Katsuki T, R G Lanckriet G, Mackey TK. Exploring trends of nonmedical use of prescription drugs and polydrug abuse in the Twittersphere using unsupervised machine learning. Addict Behav 2017 Feb;65:289-295. [doi: 10.1016/j.addbeh.2016.08.019] [Medline: 27568339]
29. Platteau T, Herrijgers C, de Wit J. Digital chemsex support and care: the potential of just-in-time adaptive interventions. Int J Drug Policy 2020 Nov;85:102927 [FREE Full text] [doi: 10.1016/j.drugpo.2020.102927] [Medline: 32932125]
30. Feller D, Zucker J, Yin M, Gordon P, Elhadad N. Using clinical notes and natural language processing for automated HIV risk assessment. J AIDS 2018;77(2):160-166. [doi: 10.1097/qai.0000000000001580]
31. Korhonen A, Séaghdha DO, Silins I, Sun L, Högberg J, Stenius U. Text mining for literature review and knowledge discovery in cancer risk assessment and research. PLoS One 2012 Apr 12;7(4):e33427 [FREE Full text] [doi: 10.1371/journal.pone.0033427] [Medline: 22511921]
32. Yates A, Cohan A, Goharian N. Depression and self-harm risk assessment in online forums. In: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. 2017 Presented at: Empirical Methods in Natural Language Processing; September 7-11; Copenhagen, Denmark. [doi: 10.18653/v1/d17-1322]
33. Maher CA, Lewis LK, Ferrar K, Marshall S, De Bourdeaudhuij I, Vandelanotte C. Are health behavior change interventions that use online social networks effective? a systematic review. J Med Internet Res 2014 Feb 14;16(2):e40 [FREE Full text] [doi: 10.2196/jmir.2952] [Medline: 24550083]
34. Selenium. Github. URL: https://github.com/SeleniumHQ/selenium [accessed 2019-12-01]
35. Richardson L. Beautiful Soup documentation. Crummy. 2007. URL: https://www.crummy.com/software/BeautifulSoup/bs4/doc/ [accessed 2019-12-01]
36. HIV risk reduction tool. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/hivrisk/estimator.html [accessed 2020-06-20]
37. Mikolov T, Sutskever I, Chen K, Corrado G, Dean J. Distributed representations of words and phrases and their compositionality. In: Proceedings of Advances in Neural Information Processing System. 2013 Presented at: Neural Information Processing Systems; December 5-8; Lake Tahoe, Nevada p. 3111-3119.
38. Kiros R, Salakhutdinov R, Zemel R. Unifying visual-semantic embeddings with multimodal neural language models. 2014 Presented at: 31st International Conference on Machine Learning; June 21-26; Beijing, China p. 595-603.
39. Maaten L. Visualizing data using t-SNE. J Mach Learn Res 2008 Nov:2579 [FREE Full text]

Abbreviations

API: application programming interface
CDC: Centers for Disease Control and Prevention
HIPAA: Health Insurance Portability and Accountability Act
HIV: human immunodeficiency virus

Edited by R Kukafka; submitted 04.07.20; peer-reviewed by A Chang, M Pradhan; comments to author 09.10.20; revised version received 25.11.20; accepted 31.01.21; published 26.04.21.

Please cite as:
Ovalle A, Goldstein O, Kachuee M, Wu ESC, Hong C, Holloway IW, Sarrafzadeh M
Leveraging Social Media Activity and Machine Learning for HIV and Substance Abuse Risk Assessment: Development and Validation Study
J Med Internet Res 2021;23(4):e22042
URL: https://www.jmir.org/2021/4/e22042
doi:10.2196/22042
PMID:33900200

Original Paper

Participant Perceptions of Facilitators and Barriers to Adherence in a Digital Mental Health Intervention for a Nonclinical Cohort: Content Analysis

Melanie Elise Renfrew[1], BEd, Grad Dip; Darren Peter Morton[1], MAppSci, PhD; Maria Northcote[1], DipT, BEd, MEd, PhD; Jason Kyle Morton[1], BEd, BSc, PhD; Jason Scott Hinze[1], BEd, MLMEd, PhD; Geraldine Przybylko[1], BCom, MBA, MPH

Lifestyle Medicine and Health Research Centre, Avondale University College, Cooranbong, NSW, Australia

Corresponding Author:
Melanie Elise Renfrew, BEd, Grad Dip
Lifestyle Medicine and Health Research Centre
Avondale University College
582 Freemans Drive
Cooranbong, NSW, 2265
Australia
Phone: 61 405445151
Email: melanie.renfrew@avondale.edu.au

Abstract

Background: Digital mental health promotion interventions (MHPIs) present a scalable opportunity to attenuate the risk of mental health distress among nonclinical cohorts. However, adherence is frequently suboptimal, and little is known about participants' perspectives concerning facilitators and barriers to adherence in community-based settings.

Objective: This study aimed to examine participants' perceptions of facilitators and barriers to adherence in a web- and mobile app–based MHPI for a nonclinical cohort.

Methods: This qualitative study used inductive, reflexive thematic analysis to explore free-text responses in a postintervention evaluation of a 10-week digital MHPI. The intervention was administered using a web and mobile app from September to December 2018. Participants (N=320) were Australian and New Zealand members of a faith-based organization who self-selected into the study, owned a mobile phone with messaging capability, had an email address and internet access, were fluent in English, provided informed consent, and gave permission for their data to be used for research. The postintervention questionnaire elicited participants' perceptions of facilitators and barriers to adherence during the intervention period.

Results: Key factors that facilitated adherence were engaging content, time availability and management, ease of accessibility, easy or enjoyable practical challenges, high perceived value, and personal motivation to complete the intervention. The primary perceived barrier to adherence was the participants' lack of time. Other barriers included completing and recording practical activities, length of video content, technical difficulties, and a combination of personal factors.

Conclusions: Time scarcity was the foremost issue for the nonclinical cohort engaged in this digital MHPI. Program developers should streamline digital interventions to minimize the time investment for participants. This may include condensed content, optimization of intuitive web and app design, simplified recording of activities, and greater participant autonomy in choosing optional features. Nonetheless, participants identified a multiplicity of other interindividual factors that facilitated or inhibited adherence.

(J Med Internet Res 2021;23(4):e25358) doi:10.2196/25358

KEYWORDS

web-based mental health; health promotion; eHealth; adherence; participant perceptions; mobile phone

Introduction

Background

Mental distress and disorders are increasingly prevalent [1], and mental health promotion interventions (MHPIs) present opportunities to enhance the well-being of nonclinical groups, decreasing the risk of mental health distress. The widespread availability of digital technology offers unprecedented accessibility, portability, and cost-effectiveness for implementing such interventions on a wide scale. However, adherence, defined by Kelders et al [2] as "the extent to which individuals should experience the content (of the intervention) to derive maximum benefit from the intervention, as defined or implied by its creators," is often suboptimal [3], and little is known about participants' perceptions of facilitators of and barriers to adherence for nonclinical cohorts in community-based settings.

Considerable research has been undertaken by researchers and program developers to facilitate optimum adherence to digital health initiatives. Methods used to promote participant adherence include the adoption of a robust set of persuasive system design principles [2,4,5], provision of rewards and incentives, positive feedback [6], implementation of gamification techniques [7,8], and various forms of human support [9]. However, adherence to digital interventions remains to be problematic [3], and there is a paucity of research elucidating participant perspectives on the factors that impact adherence [10] in community-based, nonclinical groups. Importantly, factors that may hamper adherence in a clinical population (eg, disease and disorder symptoms) [11,12] may be dissimilar to the influences on adherence in healthy nonclinical cohorts.

Several qualitative studies have addressed participant perspectives on perceived barriers to and facilitators of adherence in MHPIs for nonclinical cohorts in workplace settings. A stress management intervention, using mindfulness training, found that participants' attitudes (eg, motivation, previous interest in mindfulness, and positivity toward change), awareness of treatment effectiveness, curiosity, and alignment with personal preferences for managing stress were facilitators of adherence [13]. Conversely, barriers included insufficient detail or evidence provided as a rationale for the strategy used, length of time required for completing activities, and program intensity [13]. Carolan and de Visser [14] explored employee perspectives on workplace digital mental health interventions and found that lack of time was considered the core barrier to adherence. Notably, anonymity, convenience, and flexibility were deemed to have dual impacts (ie, considered both a facilitator and barrier). Employees described their ideal workplace intervention as one that was short, that was easily accessible, that was available for an indefinite period, updated regularly, and that provided some form of e-coaching support [14]. Research by Blankenhagel et al [15], which focused on the perspectives of various stakeholders in digital stress management programs, found that certain requirements of interventions were needed for nonclinical users who often have substantial work responsibilities. These intervention requirements were prioritizing time efficiency (5- to 15-minute sessions), splitting larger goals into smaller goals, providing flexibility for users to design their own personal notification or reminder schedule, featuring an intuitive and simple web or app design, prioritizing individualization (ie, tailoring), providing high autonomy, fostering some degree of human interaction, facilitating comparability to peers, and offering high mobility [15]. The aforementioned requirements related to a work-based setting for a nonclinical cohort; however, although there may be similarities, variant influences on adherence may be present in a community-based context.

This study sought to broaden the understanding of factors that both facilitate and hinder adherence in a nonclinical cohort participating in a digital MHPI within a community-based situation. This study builds on quantitative data collected during a randomized comparative study that examined the impact of different levels of human support on well-being outcomes [16] and attrition and adherence [17] in a digital MHPI for a nonclinical cohort. Quantitative data revealed improvements in mental well-being irrespective of the level of human support offered, and adherence did not differ significantly between groups. Notably, using the aforementioned definition of adherence by Kelders et al [2], participants were advised to view one video weekly and complete daily and weekly experiential activities over a 10-week period to gain optimal benefit from the intervention. Overall, attrition and adherence were suboptimal [17]. A total of 605 subjects indicated an interest in the intervention by filling out a preliminary web-based enrollment form, with 24.3% (147/605) of these subjects not fully registering by failing to complete the prequestionnaire. Of those who completed the prequestionnaire (n=458), a further 30.1% (138/458) did not complete the postquestionnaire, resulting in a completion rate of 69.9% (320/458). Primary adherence was measured by the number of videos viewed, with 34% (109/320) of the cohort watching less than half of the videos, 18.9% (60/320) viewing 5 to 9 videos, and 47.1% (151/320) viewing all 10 videos. Secondary adherence was measured by the points scored for completing assigned daily and weekly activities. Participants could score a total of 1000 points (ie, 100 points weekly throughout the 10-week intervention) to be considered fully adherent. The mean score for the cohort was 362, indicating low adherence to challenge activities [17].

Objectives

This study aims to increase knowledge regarding participant adherence in digital MHPIs by revealing participant perceptions about the facilitators and barriers to watching the videos and their completion of daily and weekly experiential activities. The findings will assist researchers and developers to refine and improve the design and implementation of interventions to optimize the adherence of participants in future MHPIs in nonclinical settings.

Methods

Overview

This qualitative study stems from the aforementioned study, which reported the quantitative results of the intervention

[16,17]. These qualitative responses were recorded as part of the postquestionnaire administered at the conclusion of the intervention. Participants were allocated a 2-week period, immediately after the intervention period, to complete the postquestionnaire on either the provided app or the web platform. Therefore, the following section of this paper is a summary of the methods used in the aforementioned study, with details only provided for those areas related to the qualitative components of this study.

Setting and Participants

Advertising was used to direct potential participants (Australian and New Zealand members of the Seventh-day Adventist Church) to a website that explained the intervention, outlined the inclusion criteria, and provided an opportunity to self-enroll as a potential participant. To be included in the study, subjects were required to be aged above 18 years, own a mobile phone with SMS text messaging capability, have internet access, be an Australian or New Zealand resident, be fluent in English, provide informed consent, and give permission for their data to be used for research [16]. Participants who completed the MHPI were invited to provide free-text responses to 2 statements within the postquestionnaire.

Intervention

The intervention program, known as the *Live More Project* or *Lift Project*, introduced participants to aspects of neuroscience and a combination of evidence-based strategies to enhance mental well-being from the disciplines of lifestyle medicine and positive psychology. Designed to be implemented as a weekly session for 10 weeks, the content of the program was shared in a 30-minute audiovisual presentation followed by participation in daily and weekly practical activities to consolidate learning. Video sessions were released weekly with a lock/unlock system so that participants watched the presentations in sequential order. An intervention and activity overview is provided in Multimedia Appendix 1.

Subjects chose to access the program on a web-based learning management system (eLMS) or a mobile app. Along with the audiovisual content, the eLMS or mobile app included a place to complete the pre- and postquestionnaire, a section to log experiential activity and earn points, an opportunity to interact with other participants through a public feed, and gamification strategies (eg, points on a leaderboard or badges) to induce adherence. As an optional extra, previously introduced experiential activities could be repeated and logged in forthcoming weeks to earn extra points. This resulted in an increasingly longer list of activities to complete as the intervention progressed. Screenshots of the web- and mobile app design are shown in Multimedia Appendix 2.

Participants were divided into 3 groups that differed according to the mode of human support allocation. The first group received automated email support only, the second group received automated email support along with text messages (2-3 times weekly), and the third group received weekly videoconference support along with email support [16].

Data Collection

After completing the intervention, each participant was asked to complete the postquestionnaire within a 2-week period on either the web or mobile app. As part of the postquestionnaire, participants responded to 2 adherence-related statements in an expanding text box with no imposed word limits. These statements were as follows:

- *Statement 1*: thinking about your personal experience as a participant in the Live More Project over the last 10 weeks, if it was *easy* for you to watch the video presentations and complete the daily and weekly challenges, please describe/list all the factors that contributed to this.
- *Statement 2*: thinking about your personal experience as a participant in the Live More Project over the last 10 weeks, if it was *difficult* for you to watch the video presentations and complete the daily and weekly challenges, please describe/list all the factors that contributed to this.

Data Analysis

An inductive, reflexive thematic analysis (RTA) approach [18-20] was used to extract meaning from the data set. The inductive approach is sometimes described as *bottom-up*, or data-driven, as the themes are generated from within the data [20]. The RTA method suited one researcher (MR) performing the initial coding, followed by 2 additional coders reviewing the progress. MR had multiple roles in the study and functioned as the technical support person; therefore, this approach, in which multiple coders reviewed the data (as recommended by O'Brien et al [21]), reduced the possibility of personal bias.

The coding researcher adopted the 6 phases of thematic analysis by Braun and Clarke [20] as a structured approach and used a spreadsheet (Microsoft Excel) and a qualitative analysis program (NVivo software) to assist in collation, analysis, and coding of the deidentified data. All identifying information was removed before coding, and participant identification numbers were used instead. In phase 1, the researcher read through the data set twice to maximize familiarity with the data. During phase 2, all comments were systematically analyzed by constructing and labeling codes in a descriptive and interpretive manner until all data were coded. In an additional immersion stage, coding was reviewed and code names were edited. In phase 3, the researcher scrutinized the codes for patterns, relationships, and similarities. The codes were clustered together based on broad themes. During phase 4, after repeated scrutiny to ensure that the code names accurately matched the data and that the developing themes represented the participants' descriptions of their most typical experiences, clusters were formed into broad themes and several subthemes to develop a thematic map, as recommended by Braun and Clarke [18]. To ensure that the themes and subthemes accurately reflected the original data set as a whole, and the coded data, a final revision of the themes and codes was undertaken by the original coding researcher in consultation with 2 other researchers. During this process, the coders analyzed the themes and subthemes in relation to the original intention of the study. This procedure enabled the coders to identify any mismatches between the theme and subtheme names and the original data. Furthermore, the coders identified any themes, subthemes, codes, or data within codes that were

deemed of low relevance to the overall themes or this study's intention. This relevance-based scrutiny was followed by analysis of any codes that were characterized by very few participant quotes. This way, we were able to determine the *weightiness* or key ideas across the entire data set. In cases where relevance and the number of participant responses were considered to be low, the subtheme or code was discarded from the overall thematic map.

In the final phase of coding, the original coder and consulting coders collaboratively identified the main focus and extent of each theme. Some names of the main themes and subthemes were also modified to increase the alignment between the documented theme and subtheme labels and the data they each contained. This refinement of the theme and subtheme definitions informed the final write-up phase.

The Standards for Reporting Qualitative Research checklist was used as a guide for reporting (Multimedia Appendix 3) [21], and ethics approval was granted by the Avondale Human Research Ethics Committee (approval 2018.09).

Results

Overview

Most participants were female (262/320, 81.9%), and 50.3% (161/320) of the participants were aged 40 to 60 years. Almost one-third of the cohort (91/320, 28.4%) were aged 25 to 39 years, 15.6% (50/320) were aged 61 to 81 years, and 5.6% (18/320) were aged 18-24 years. Two-thirds of the cohort (212/320, 66.2%) worked either full or part time, 6.2% (20/320) identified as students, and the remaining 27.6% (88/320) identified as being unemployed or doing *home duties*. The majority of the cohort (240/320, 75%) provided a codable response describing facilitators of adherence, and almost all participants (314/320, 98.1%) provided codable responses related to barriers.

As the participants were asked to respond to 2 separate statements (one focusing on facilitators and the other focusing on barriers), responses to these 2 statements were analyzed separately. However, in some cases, participant responses overlapped and issues were reported as both a facilitator and a barrier. Broad themes that emerged were time, program components, personal factors, system design, technology, and human support. Notably, comment frequency regarding barriers to adherence were more than double the comments specifying facilitators of adherence. Table 1 (facilitators of adherence) and Table 2 (barriers to adherence) outline the thematic framework of responses.

During the development of the thematic map, it became evident that there was strong alignment between themes and subthemes in both sets of responses, that is, the thematic map illustrates the two-fold nature of most themes, with the order of reporting of the themes reflecting the most important themes first, as assessed by the coding researcher, according to relevance and volume of comments received. The issue of time emerged as a dominant theme in the data, with the vast majority of participants mentioning *time* across their responses.

Table 1. Perceived facilitators of adherence—themes, subthemes, and codes.

Themes, subthemes, and codes or category	Comments, n
Time	
Availability	
Opportune circumstances	20
Weekend	12
Efficiency	
Priority and planning	14
Multitasking	10
Program components	
Videos	
Highly engaging	70
Length	5
Challenge or experiential factors	
Achievable	18
Enjoyable	12
Part of normal lifestyle	6
Accountability	5
Overall program	
Easy or engaging	8
Quality	5
Personal factors	
Motivation	
High interest in personal health	9
Commitment to complete	5
Perceived value	
Personally beneficial	11
Reinforcement of positive habits	5
System design	
Convenience	
Accessibility	16
App	14
Usability	
Extra features	7
Ease of use	7

Table 2. Perceived barriers to adherence—themes, subthemes, and codes.

Themes, subthemes, and codes or category	Comments, n
Time	
Daily life	
Lack of time in general	129
Work and study combination	58
Family responsibilities	40
Travel	16
Full schedule	8
Work and family combination	6
Life event	
Death	9
Medical	6
Relocation	5
Other major change	5
Period	
Christmas	15
Holiday travel	6
Program components	
Challenge or experiential factors	
Perceived as difficult	33
Time consuming	22
Too many activities	19
Disliked recording or logging activity	18
Forgot to implement or record	10
Confusion	5
Video factors	
Length	17
Unappealing content or style	10
Extra segments	7
Personal factors	
Well-being	
Illness	16
Fatigue	10
Mental health distress or disorder	7
Stress	5
Capacity	
Lack of motivation	18
Lagging behind	17
Not a priority, so forgot	16
System design	
Lock restrictions	10
Features missing	9
Gamification	8

Themes, subthemes, and codes or category	Comments, n
Login and navigation	8
Video playback	5
Recording activity	5
Technology	
Internet	
Internet and data access	28
Personal device	
Faulty	18
Interrupted viewing	6
Human support	
Videoconferencing difficulties[a]	11
Lack of prompts	8
Lack of accountability	6
Preference for face-to-face	5

[a]Videoconferencing support was provided to 1 subgroup during the intervention (n=103).

Facilitators of Adherence

Time

Time was perceived as a facilitator of adherence when participants believed that they had time available and efficiency was optimized. Each weekly lesson was released on a Sunday, and this timing was viewed positively. Expediency afforded by being retired, living alone, or not working also facilitated adherence:

> *I liked that the videos were released on Sunday, even if some Sundays were busy and we couldn't make time to watch them that day. Sunday is a day it's easier to make time to watch the full set of lesson videos.* [ID 2614]

Some participants prioritized, scheduled, and multitasked to ensure that the program components were completed each week. Specific factors that permitted time availability or efficiency included being at home, not traveling, completion of usual duties, incorporation of tasks into daily activities, a high level of organization or planning, and having children in a regular sleep routine:

> *When I made specific time to sit down and watch the program it made it easier to participate and plan how I can incorporate the challenges.* [ID 2278]

Program Components

This theme refers specifically to video content and the completion of practical activities. A total of 3 subthemes related to the program's components emerged from the analysis: the overall program, video presentations, and experiential activity.

Although some participants perceived that the overall program was of high quality, easy to understand, and engaging, the key program components perceived as facilitators of adherence were the video presentations and the experiential challenges. Engaging video content has been frequently reported as a facilitator of adherence. The terms used to describe the videos included *humorous*, *entertaining*, *interesting*, *motivating*, and *easy to understand*. Participants appreciated the lighthearted presentation style, which included just enough information and interesting facts, to avoid information overload:

> *The videos were funny and engaging. Had enough info to be interesting but not overly filled with over-your-head stuff.* [ID 2380]

The daily and weekly experiential challenges that participants completed and recorded were perceived as fostering adherence because they were relevant, easy to accomplish, enjoyable, simple to log or record, and readily incorporated into daily life. Some challenges could be easily doubled up and, for some participants, recording their challenge activity provided a sense of accomplishment:

> *The small daily challenges were able to be absorbed into the day without having to find extra time and it felt like many could become habits.* [ID 2302]

> *I found a lot of the challenges crossed into another which made it easier such as being in the green and blue and going for a walk covered two challenges.* [ID 2550]

Personal Factors

There are numerous personal reasons that may influence adherence, including beliefs, motivation, personal well-being, and a broad array of diverse circumstances. During coding, several key personal subthemes were developed that included motivational factors, beliefs about the value or relevance of the program, level of personal well-being, and individual capacity.

Motivation and perceived value were identified as subthemes that positively influenced adherence at a personal level. Participants reported positive motivation through a high level of readiness, commitment to the intervention, and a personal

desire to improve their health. Some participants were motivated to maintain their adherence to the program based on their belief that the intervention would be personally beneficial to them. They recognized a number of enabling aspects of the program, including the way it reinforced positive lifestyle choices; recognized that even small changes would be worthwhile; and identified the value of immediate effects. Participants also benefited from the multifactorial approach:

> *I looked forward to the video presentations because they were so relevant to my life, very interesting and well presented. I have been experiencing problems with sleep issues for many years and even taking herbal sleeping supplements, but after applying the information I received from the Live More sessions, I've been a good 7 to 8 hours sleep every night. This has been mainly the result of all the walking I've been doing - between 1 to 2 hours a day and more. I'm now taking more notice of nature and enjoying more green and blue. Having recently moved from [name of Country], I had no friends and felt lonely. So I went to the local Library and Community Centre and got a booklet with the different activities for retired people. I joined a group of like-minded people and enjoyed the activities they provided.* [ID 2419]

System Design

This theme relates to the various design and accessibility features of the web or mobile app and primarily includes the methods used to disseminate weekly video content and record experiential activities in the digital setting. Additional features such as gamification (ie, earning points), the ability to view what others were doing, and accessibility to extra resources were also included. Similar to other themes, the system design facilitated adherence for some participants and was perceived as a barrier for an almost equal number.

Two key design concepts were reported by users as positive influences on adherence: accessibility and convenience of the mobile app. Participants appreciated flexible access to view and review the content. This was amplified with easy portability through a mobile app:

> *I could do it when it was suitable for me. Whether 10pm or 6am. The flexibility made it possible.* [ID 2195]

Barriers to Adherence

Time

Participants overwhelmingly perceived lack of time as problematic to adherence, with noticeably more comments referencing time as a barrier to adherence rather than a facilitator. Furthermore, time was mentioned substantially more often than any other factor recognized as a barrier. The 3 subthemes generated in relation to lack of time were daily life, life events, and the intervention period.

The demands of daily life were the foremost barriers and predominantly included work, study, and family commitments and the various combinations of all three:

> *Small child needing attention, lack of help from spouse due to work commitments; other extra-curricular activity; helping my elderly parents with household tasks and other daily activities.* [ID 2538]

Major life events were perceived as hindering adherence and included the death of a loved one, relocation, major illness, or a combination of trials that consumed major amounts of time:

> *The timing of this project was the one of the busiest 10 weeks in my life! Moved house. Solo parented for 4 weeks. My sister's wedding. School holidays. The sickest our family has been in 10 years. Term 4 after-hours school events as part of my work responsibilities. We moved house the week the program started and then went straight into school holidays of which we were away the whole time. Then we were really sick for three weeks whilst I was solo parenting. So I just never got the chance to get started.* [ID 2214]

The intervention was conducted from September to early December 2018. The lead up to Christmas and the added pressures that often accompany the Christmas period in the work and social environment were deemed barriers. In addition, for this cohort (ie, a faith-based population), church activities were perceived as adding further to the time pressures already experienced at this time of year:

> *Making a commitment at this time of year was more difficult than I thought; this time of year is hectic for my job which also probably means my answers at the end could actually be worse than at the start, because of how tired I have been.* [ID 2630]
>
> *If it was at another time of year I probably would have completed it. This time of year is crazy for me at work, I'm a preschool assistant, and I am involved with helping with Road to Bethlehem [Christmas program]. The last few weeks I just found it difficult to think much about what I was doing.* [ID 2417]

Program Components

Some users deemed the overall program too intense and too long. Video segments were perceived as lengthy, and some users found the content unappealing and disliked the presentation style, describing it as *cliché*, *corny*, and *over the top*:

> *The videos were too long. I'd prefer 2-3 shorter ones per topic. That would be easier to watch while in train or fitting between other things I had to do.* [ID 2318]
>
> *While the actual science was interesting, much of the video content felt patronising and the jokes fell flat.* [ID 2142]

Most perceptions about the program components were directly related to difficulties with experiential activities. Time was viewed as a hindrance for both completing and logging the activity, and some participants perceived the activities as being too hard. Although some participants completed the daily and weekly experiential challenges, they disliked recording their

activities, felt overwhelmed by the increasingly growing list of activities to record, or simply forgot to log their activities:

> *Having the challenges compound each week to record them meant that it was too overwhelming and I gave up logging on to record. I think having reminders about the previous weeks challenges is great, but filling in that many challenges every day is too hard!* [ID 2287]

Personal Factors

Numerous personal factors hindering adherence were grouped under 2 subthemes: personal well-being and capacity. Personal illness was the dominant factor affecting well-being. Other factors influencing well-being included the presence of a mental disorder, stress, and fatigue, with several respondents specifically identifying screen fatigue. Comments regarding capacity were equally divided among 3 key areas: a reported lack of motivation, discouragement at lagging behind in viewing content or completing experiential activities, and simply not prioritizing the intervention:

> *When I had downtime I didn't have the motivation to intently watch the videos & do the thinking & work required so it was easy to fall behind.* [ID 2204]

> *Also, with working full time (looking at a screen) all day, I really didn't want to sit in my home office to look at another screen, so I failed to keep on top of the weekly presentations.* [ID 2563]

System Design

Several design features and difficulty navigating the website were perceived as negative influences on adherence. Most comments were made in relation to the lock restrictions, the lack of some desirable features, dissatisfaction with the gamification setup, and the perceived tedious system for logging experiential activity. Although participants were provided with a recommended 10-week schedule for engaging with the intervention content, some participants emphatically opposed the lack of autonomy in how and when they consumed the information:

> *I don’t mind having to unlock videos - just please don’t constrict it to a week at a time!!! And let me do any or all of the challenges at the start or when I unlock things in my own time - don’t make me wait 7 days / a fortnight / 3 weeks etc to engage in all the challenges just because! Let me unlock them at my pace.* [ID 2142]

Desirable features that were noted as missing included an audio-only option, the ability to predownload the video content, and the option to read the material instead of viewing a video. A readable script was actually provided, although it was not explicitly recommended as an alternative to viewing video content. Although gamification was included as a way to enhance adherence, for some participants it was perceived negatively and a deterrent to further adherence:

> *It's very demotivating having a leaderboard for someone who's not near the top.* [ID 2549]

The system setup for recording experiential activity was seen as a barrier because it was perceived as difficult to use, with one participant describing it as “glitchy and not particularly intuitive” with “long and clunky lists” [ID 2142]:

> *Prompts on challenge input boxes were too vague about what the user needs to write. Should the user explain anything or just state that the challenge was completed? It was not clear whether the user's statement in the challenge input box would we shared with others (user had to just try it and see what happens during the first session). I could not log more than one challenge on the interface at a time (I had to change to another page and then come back to the challenges to log another one).* [ID 2391]

Technology

Technology (ie, data availability, internet access, and device suitability) was deemed a barrier, mainly because of poor internet accessibility or personal device problems.

Internet access problems included slow download speeds, internet outages, and no internet access at home. Personal device problems included old or faulty smartphones that could not download the mobile app, small screen size, sound problems, and perceived inability to correctly operate the device:

> *We had some difficulty watching the videos on my tablet, the volume is not loud enough, we really need a PC or to data project to a bigger screen for optimal viewing/listening.* [ID 1767]

Human Support

Human support offered to participants varied depending on the group allocation they received. All participants received automated email reminders to engage with video content and record their practical activities. One group received regular text messages, and the other group was offered weekly videoconference sessions as an extra measure of support. Despite human support being the focus of the original study [16], perceptions about the support offered did not feature strongly. However, some participants reported that adherence was impeded because of human support factors or lack thereof. A lack of prompting and accountability were stated as reasons for difficulty in adherence. In the group that received videoconferencing support, some of the group members reported that videoconferencing support was problematic for a range of reasons, most commonly that the meetings were held at an unsuitable time:

> *Timing of online groups sessions didn’t work for me with my responsibilities during that time period. I needed a later evening session. 9pm! Or a 5:30am session. The timing of the first 5 weeks would have been difficult for me regardless.* [ID 2214]

Discussion

Principal Findings

This study provides novel insights into participants’ perspectives regarding facilitators and barriers to adherence in a digital MHPI

for a community-based nonclinical population. Themes often had dual impacts, deemed as both facilitators and barriers. Facilitators of adherence included engaging in video content, achievable experiential activities, time availability, user-friendly web and mobile app design, and high personal interest or motivation. Conversely, a perceived lack of time was the main barrier to adherence. Completing and logging practical activities, a range of personal influences, problematic design features, technological issues, and a lack of support or prompting were also identified as barriers.

As barriers were the predominant focus of participants' perspectives, it is clear that an MHPI is a difficult endeavor for a nonclinical group and that time is the foremost issue in relation to adherence. Similarly, other qualitative studies in workplace settings [14,15] have concluded that time scarcity is a major inhibitor of adherence and that participants need interventions requiring only small time commitments. It is evident that certain features or problems (eg, difficult website navigation and restrictions on video viewing) consumed inordinate amounts of the participants' time, thereby negatively impacting their adherence. Similarly, such issues have previously been exposed as barriers to the success of web-based learners in educational settings [22,23]. Therefore, time efficiency must be a core consideration by researchers in every aspect of the intervention design. It is vital to design interventions to minimize time and energy investment for participants while concurrently optimizing outcomes.

An important element that underpinned the data was the participants' desire to easily incorporate core components of the program (ie, videos and challenge activities) into their everyday lives. A key priority in the design and development of future interventions should be to collaborate with potential users in designing interventions that enable autonomous, optimal integration of the essential intervention elements into daily life. Practical examples may include the provision of multiple options to access the content (eg, video, print, or audio), choice of methods to log activities, and recommendations on combining some activities to make optimal use of time.

The preponderance of dual themes in this study sheds more light on an important element that has been identified previously [24], that is, individual differences have a substantial influence on adherence to MHPIs. Although previous research by Banerjee et al [13] identified several twofold themes in a qualitative study focusing on engagement with a mindfulness intervention, a study by Zarski et al [24] on adherence to a digital stress management intervention noted that, even after accounting for socioeconomic demographics, personal well-being status, and human support, many individual variances in nonadherence were inexplicable. This study highlights some of these differences.

Understandably, the high prevalence of individual variance confounds efforts to crystallize a precise set of factors influencing adherence to digital interventions. Plausibly, future research should embrace individual differences by examining ways to prioritize participant autonomy in tailoring interventions according to personal preferences. This could be achieved by offering an eclectic, self-directed approach that encourages participant choice in activating or deactivating a range of optional features (eg, gamification or type and frequency of human support) as desired. This constructivist-style approach, commonly used in educational settings, is well supported by research as an efficacious method [25,26].

The original study, from which this qualitative study was derived, focused on the influence of varying levels of human support on outcomes and adherence to an MHPI [16,17]. However, in this qualitative analysis, human support did not have a strong influence on adherence. Considering that the quantitative analysis revealed that adherence was not impacted by different levels of support and that participants achieved improvements in outcomes irrespective of the human support provided, this is not surprising. Although numerous studies have demonstrated that human support may improve outcomes [9,27,28], there is substantial evidence that well-designed, self-guided interventions can be successful without added human support [9,29-32] and their associated costs.

Strengths and Limitations

The large cohort and inclusive age range (18-81 years) of participants were strengths of this study. Furthermore, participants were not restricted to the length of their responses and could write as little or as much as desired. All comments were coded, and participants received no guidance on what topics to comment on, eliminating preconceptions by researchers.

Recruitment bias could be seen as a potential limitation to this study as participants were self-selected into the study, and the target group were Seventh-day Adventist Church members who typically share common health practices (eg, abstinence from alcohol or smoking) and may be more inclined to take a greater interest in health than the general population. In addition, recruits were largely tertiary educated (262/320, 81.9%), White (280/320, 87.5%), and female (262/320, 81.9%). Although a high female proportion is the characteristic of an internet intervention [33,34], it limits transferability to the wider population. In addition, participants were not given the opportunity to indicate whether they used the web-based platform, the mobile app, or a combination of both, which prohibited a comparison of data based on the different delivery methods.

By allowing participants to comment freely on an open-ended statement, the assumption must be made that participants only reported on factors most relevant to their personal experiences. Therefore, the study does not contain all the views of the participants. As researchers, we adopt the position that the predominant issues are exposed and that the volume of comments depicts saturation on crucial issues.

Notably, 2 other factors may have inadvertently influenced participants' perceptions. First, barriers, particularly *time*, may have been augmented by the period that the intervention was conducted (mid-September to early December 2018). Avoiding administering an intervention that ends in December is a learning for future implementation. Second, the website and mobile app design, in permitting recording of an increasing array of experiential activities as the weeks progressed (rather than just

the activities for the current week), may have augmented dissatisfaction with the logging or recording component, introducing an element that went beyond the intentions of the original intervention design.

Conclusions

Although there are a diversity of factors influencing adherence, highly engaging content encourages adherence and time efficiency is a principal consideration for a nonclinical, community-based cohort. Therefore, intervention designers should place a concerted focus on streamlining interventions to reduce the time required by participants and simplify all program elements. Other considerations to increase adherence include allowing participants to create their own experience by choosing what optional features to include or exclude, enhancement of the website and mobile app to facilitate a more intuitive experience, shorter video presentations that are easily accessible and provide high autonomy in the way they are consumed, simplified and limited recording of experiential activities, and options for personal choice in accountability. In future research, it will be important to consider the views of users in relation to adherence and respond with an adequate level of agility that assures an ever-improving experience for participants to ensure optimal adherence and sustainability.

Acknowledgments

This work was supported by the South Pacific Division of the Seventh-day Adventist Church that provided free access to all participants who participated in the study. The authors would like to thank Jared Madden for coordinating the technical support throughout the study.

Authors' Contributions

MR and DM conceptualized this study. MR completed the data analysis in consultation with DM and MN. All authors contributed to the revision and editing of the manuscript.

Conflicts of Interest

DM manages a *profit-for-purpose* trust that uses a version of the intervention without receiving personal remuneration. GP is employed with the South Pacific Division of the Seventh-day Adventist Church (administrator of the intervention). No other authors have any financial interests or conflicts of interest to declare.

Multimedia Appendix 1
Program overview.
[PDF File (Adobe PDF File), 110 KB - jmir_v23i4e25358_app1.pdf]

Multimedia Appendix 2
Web or app screenshots.
[PDF File (Adobe PDF File), 1141 KB - jmir_v23i4e25358_app2.pdf]

Multimedia Appendix 3
Checklist—standards for reporting qualitative research.
[PDF File (Adobe PDF File), 566 KB - jmir_v23i4e25358_app3.pdf]

References

1. Jorm AF, Patten SB, Brugha TS, Mojtabai R. Has increased provision of treatment reduced the prevalence of common mental disorders? Review of the evidence from four countries. World Psychiatry 2017 Feb;16(1):90-99 [FREE Full text] [doi: 10.1002/wps.20388] [Medline: 28127925]
2. Kelders SM, Kok RN, Ossebaard HC, Van GJEWC. Persuasive system design does matter: a systematic review of adherence to web-based interventions. J Med Internet Res 2012;14(6):e152 [FREE Full text] [doi: 10.2196/jmir.2104] [Medline: 23151820]
3. Linardon J, Fuller-Tyszkiewicz M. Attrition and adherence in smartphone-delivered interventions for mental health problems: A systematic and meta-analytic review. J Consult Clin Psychol 2020 Jan;88(1):1-13. [doi: 10.1037/ccp0000459] [Medline: 31697093]
4. Asbjørnsen RA, Smedsrød ML, Solberg Nes L, Wentzel J, Varsi C, Hjelmesæth J, et al. Persuasive system design principles and behavior change techniques to stimulate motivation and adherence in electronic health interventions to support weight loss maintenance: Scoping review. J Med Internet Res 2019 Jun 21;21(6):e14265 [FREE Full text] [doi: 10.2196/14265] [Medline: 31228174]
5. Oinas-Kukkonen H, Harjumaa M. Persuasive systems design: Key issues, process model, and system features. CAIS 2009;24:485-501. [doi: 10.17705/1CAIS.02428]

6. Mohr DC, Cuijpers P, Lehman K. Supportive accountability: A model for providing human support to enhance adherence to eHealth interventions. J Med Internet Res 2011;13(1):e30 [FREE Full text] [doi: 10.2196/jmir.1602] [Medline: 21393123]
7. Brown M, O'Neill N, van Woerden H, Eslambolchilar P, Jones M, John A. Gamification and adherence to web-based mental health interventions: A systematic review. JMIR Ment Health 2016;3(3):e39 [FREE Full text] [doi: 10.2196/mental.5710] [Medline: 27558893]
8. Schmidt-Kraepelin M, Warsinsky S, Thiebes S, Sunyaev A. The role of gamification in health behavior change: A review of theory-driven studies. In: Proceedings of the 53rd Hawaii International Conference on System Sciences. 2020 Presented at: 53rd Hawaii International Conference on System Sciences; 2020; Hawaii p. 1256-1265 URL: https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=w
9. Shim M, Mahaffey B, Bleidistel M, Gonzalez A. A scoping review of human-support factors in the context of Internet-based psychological interventions (IPIs) for depression and anxiety disorders. Clin Psychol Rev 2017 Nov;57:129-140. [doi: 10.1016/j.cpr.2017.09.003] [Medline: 28934623]
10. Esponda GM, Hartman S, Qureshi O, Sadler E, Cohen A, Kakuma R. Barriers and facilitators of mental health programmes in primary care in low-income and middle-income countries. Lancet Psychiatry 2020 Jan;7(1):78-92. [doi: 10.1016/S2215-0366(19)30125-7] [Medline: 31474568]
11. Beatty L, Binnion C. A systematic review of predictors of, and reasons for, adherence to online psychological interventions. Int J Behav Med 2016 Dec;23(6):776-794. [doi: 10.1007/s12529-016-9556-9] [Medline: 26957109]
12. Glowacki K, Duncan MJ, Gainforth H, Faulkner G. Barriers and facilitators to physical activity and exercise among adults with depression: A scoping review. Mental Health and Physical Activity 2017 Oct;13:108-119. [doi: 10.1016/j.mhpa.2017.10.001]
13. Banerjee M, Cavanagh K, Strauss C. A qualitative study with healthcare staff exploring the facilitators and barriers to engaging in a self-help mindfulness-based intervention. Mindfulness 2017;8(6):1653-1664 [FREE Full text] [doi: 10.1007/s12671-017-0740-z] [Medline: 29201248]
14. Carolan S, de Visser RO. Employees' perspectives on the facilitators and barriers to engaging with digital mental health interventions in the workplace: Qualitative study. JMIR Ment Health 2018 Jan 19;5(1):e8 [FREE Full text] [doi: 10.2196/mental.9146] [Medline: 29351900]
15. Blankenhagel K, Theilig M, Koch H, Witte A, Zarnekow R. Challenges for preventive digital stress management systems-identifying requirements by conducting qualitative interviews. 2019 Presented at: Proceedings of the 52nd Hawaii International Conference on System Sciences; 2019; Hawaii. [doi: 10.24251/hicss.2019.461]
16. Renfrew ME, Morton DP, Morton JK, Hinze JS, Beamish PJ, Przybylko G, et al. A web- and mobile app-based mental health promotion intervention comparing email, short message service, and videoconferencing support for a healthy cohort: Randomized comparative study. J Med Internet Res 2020 Jan 06;22(1):e15592 [FREE Full text] [doi: 10.2196/15592] [Medline: 31904578]
17. Renfrew ME, Morton DP, Morton JK, Hinze JS, Przybylko G, Craig BA. The influence of three modes of human support on attrition and adherence to a web- and mobile app-based mental health promotion intervention in a nonclinical cohort: Randomized comparative study. J Med Internet Res 2020 Sep 29;22(9):e19945 [FREE Full text] [doi: 10.2196/19945] [Medline: 32990633]
18. Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Research in Psychology 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
19. Braun V, Clarke V. Reflecting on reflexive thematic analysis. Qualitative Research in Sport, Exercise and Health 2019 Jun 13;11(4):589-597. [doi: 10.1080/2159676x.2019.1628806]
20. Braun V, Clarke V. Thematic Analysis. In: Cooper H, editor. APA handbook of research methods in psychology, Vol 2: Research designs: Quantitative, qualitative, neuropsychological,biological. Washington DC: American Psychological Association; 2012:57-71.
21. O'Brien BC, Harris IB, Beckman TJ, Reed DA, Cook DA. Standards for reporting qualitative research: A synthesis of recommendations. Acad Med 2014 Sep;89(9):1245-1251 [FREE Full text] [doi: 10.1097/ACM.0000000000000388] [Medline: 24979285]
22. Markova T, Glazkova I, Zaborova E. Quality issues of online distance learning. Procedia - Social and Behavioral Sciences 2017 Feb;237:685-691. [doi: 10.1016/j.sbspro.2017.02.043]
23. Muilenburg LY, Berge ZL. Student barriers to online learning: A factor analytic study. Distance Education 2005 Jan;26(1):29-48. [doi: 10.1080/01587910500081269]
24. Zarski A, Lehr D, Berking M, Riper H, Cuijpers P, Ebert DD. Adherence to internet-based mobile-supported stress management: A pooled analysis of individual participant data from three randomized controlled trials. J Med Internet Res 2016;18(6):e146 [FREE Full text] [doi: 10.2196/jmir.4493] [Medline: 27357528]
25. Swan K, Garrison D, Richardson J. A constructivist approach to online learning: The community of inquiry framework. In: Payne CR, editor. Information technology and constructivism in higher education: Progressive learning frameworks. Hershey PA: IGI global; 2009:43-57.
26. Schrader D. Constructivism and learning in the age of social media: Changing minds and learning communities. New Directions for Teaching and Learning 2015 Dec 13;2015(144):23-35.

27. Baumeister H, Reichler L, Munzinger M, Lin J. The impact of guidance on Internet-based mental health interventions — A systematic review. Internet Interventions 2014 Oct;1(4):205-215. [doi: 10.1016/j.invent.2014.08.003]
28. Heber E, Ebert DD, Lehr D, Cuijpers P, Berking M, Nobis S, et al. The benefit of web- and computer-based interventions for stress: A systematic review and meta-analysis. J Med Internet Res 2017 Feb 17;19(2):e32 [FREE Full text] [doi: 10.2196/jmir.5774] [Medline: 28213341]
29. Ebert DD, Heber E, Berking M, Riper H, Cuijpers P, Funk B, et al. Self-guided internet-based and mobile-based stress management for employees: Results of a randomised controlled trial. Occup Environ Med 2016 May;73(5):315-323. [doi: 10.1136/oemed-2015-103269] [Medline: 26884049]
30. Dear BF, Staples LG, Terides MD, Fogliati VJ, Sheehan J, Johnston L, et al. Transdiagnostic versus disorder-specific and clinician-guided versus self-guided internet-delivered treatment for Social Anxiety Disorder and comorbid disorders: A randomized controlled trial. J Anxiety Disord 2016 Aug;42:30-44 [FREE Full text] [doi: 10.1016/j.janxdis.2016.05.004] [Medline: 27261562]
31. Dear BF, Fogliati VJ, Fogliati R, Johnson B, Boyle O, Karin E, et al. Treating anxiety and depression in young adults: A randomised controlled trial comparing clinician-guided versus self-guided Internet-delivered cognitive behavioural therapy. Aust N Z J Psychiatry 2018 Jul;52(7):668-679. [doi: 10.1177/0004867417738055] [Medline: 29064283]
32. Hadjistavropoulos HD, Schneider LH, Mehta S, Karin E, Dear BF, Titov N. Preference trial of internet-delivered cognitive behaviour therapy comparing standard weekly versus optional weekly therapist support. J Anxiety Disord 2019 Apr;63:51-60 [FREE Full text] [doi: 10.1016/j.janxdis.2019.02.002] [Medline: 30844613]
33. Mitchell J, Stanimirovic R, Klein B, Vella-Brodrick D. A randomised controlled trial of a self-guided internet intervention promoting well-being. Computers in Human Behavior 2009 May;25(3):749-760. [doi: 10.1016/j.chb.2009.02.003]
34. Waters LA, Galichet B, Owen N, Eakin E. Who participates in physical activity intervention trials? J Phys Act Health 2011 Jan;8(1):85-103. [Medline: 21297189]

Abbreviations

eLMS: web-based learning management system
MHPI: mental health promotion intervention
RTA: reflexive thematic analysis

Edited by R Kukafka; submitted 28.10.20; peer-reviewed by I Choi, S Poduval, L Morrison; comments to author 10.12.20; revised version received 17.12.20; accepted 18.03.21; published 14.04.21.

Please cite as:
Renfrew ME, Morton DP, Northcote M, Morton JK, Hinze JS, Przybylko G
Participant Perceptions of Facilitators and Barriers to Adherence in a Digital Mental Health Intervention for a Nonclinical Cohort: Content Analysis
J Med Internet Res 2021;23(4):e25358
URL: https://www.jmir.org/2021/4/e25358
doi:10.2196/25358
PMID:33851925

Original Paper

Effect of a Virtual Reality–Enhanced Exercise and Education Intervention on Patient Engagement and Learning in Cardiac Rehabilitation: Randomized Controlled Trial

Victoria Gulick[1], BA, MS; Daniel Graves[2], PhD; Shannon Ames[2], MBA, DPT; Pavitra Parimala Krishnamani[3], MD, MS

[1]Information Services & Technologies, Jefferson Health, Philadelphia, PA, United States

[2]Rehabilitation Medicine, Thomas Jefferson University, Philadelphia, PA, United States

[3]Department of Emergency Medicine, Baylor College of Medicine, Houston, TX, United States

Corresponding Author:
Victoria Gulick, BA, MS
Information Services & Technologies
Jefferson Health
833 Chestnut St
Suite 1000
Philadelphia, PA, 19107
United States
Phone: 1 215 503 3224
Email: victoria.gulick@jefferson.edu

Abstract

Background: Cardiac rehabilitation (CR) is clinically proven to reduce morbidity and mortality; however, many eligible patients do not enroll in treatment. Furthermore, many enrolled patients do not complete their full course of treatment. This is greatly influenced by socioeconomic factors but is also because of patients' lack of understanding of the importance of their care and a lack of motivation to maintain attendance.

Objective: This study aims to explore the potential benefits of virtual reality (VR) walking trails within CR treatment, specifically with regard to patient knowledge retention, satisfaction with treatment, and the overall attendance of treatment sessions.

Methods: New CR patients were enrolled and randomized on a rolling basis to either the control group or intervention group. Intervention patients completed their time on the treadmill with VR walking trails, which included audio-recorded education, whereas control patients completed the standard of care therapy. Both groups were assisted by nursing staff for all treatment sessions. Primary outcomes were determined by assessing 6-minute walk test improvement. In addition, secondary outcomes of patients' cardiac knowledge and satisfaction were assessed via a computer-based questionnaire; patient adherence to the recommended number of sessions was also monitored. Cardiac knowledge assessment included a prerehabilitation education quiz, and the same quiz was repeated at patients' final visit and again at the 2-month follow-up. The satisfaction questionnaire was completed at the final visit.

Results: Between January 2018 and May 2019, 72 patients were enrolled—41 in the intervention group and 31 in the control group. On the basis of the results of the prerehabilitation and postrehabilitation 6-minute walk test, no significant differences were observed between the intervention and control groups (P=.64). No statistical differences were observed between groups in terms of education (P=.86) or satisfaction (P=.32) at any time point. The control group had statistically more favorable rates of attendance, as determined by the risk group comparison (P=.02) and the comparison of the rates for completing the minimum number of sessions (P=.046), but no correlation was observed between the study group and reasons for ending treatment.

Conclusions: Although no improvements were seen in the VR intervention group over the control group, it is worth noting that limitations in the study design may have influenced these outcomes, not the medium itself. Furthermore, the qualitative information suggests that patients may have indeed enjoyed their experience with VR, even though quantitative satisfaction data did not capture this. Further considerations for how and when VR should be applied to CR are suggested in this paper.

Trial Registration: ClinicalTrials.gov NCT03945201; https://clinicaltrials.gov/ct2/show/NCT03945201

(J Med Internet Res 2021;23(4):e23882) doi:10.2196/23882

KEYWORDS

virtual reality; VR; cardiac rehabilitation; patient experience; patient education; outpatient therapy; exercise

Introduction

Background

Cardiac rehabilitation (CR) is an underutilized but vital part of recovery following a cardiac event. However, many patients do not attend or complete the recommended number of treatment sessions. According to the Million Hearts Initiative, only approximately one-third of all eligible patients participate in CR treatment [1]. Previous studies have shown that CR decreases "morbidity and mortality, improves clinical outcomes, enhances psychological recovery, and decreases the risk for secondary cardiac events" [1] and that a 12-week program can reduce cardiovascular risk factors for over a year post treatment [2]. A myriad of social, economic, and cultural factors contributes to the low percentage of patients who participate in CR [3]. However, even among patients who can attend and afford sessions, many do not complete their entire course of treatment. Research has shown that a lack of motivation [4] and a lack of understanding [5] contribute to this behavior.

A fundamental part of outpatient CR treatment is access to education. At many CR centers, such as the Jefferson Health Methodist CR facility, nursing staff conduct periodic lectures while patients exercise. However, because of the number of patients or timing of sessions, most educational materials are distributed in paper format, either on bulletin boards or as handouts.

To address the lack of interest in treatment and difficulties in providing access to patient education, we decided to pursue the use of digital technologies as an alternative, potentially more engaging way of conveying information to patients. This was founded on the work of the Digital Innovation and Consumer Experience (DICE) Group. The DICE Group used technology to advance the future of health care and education at Jefferson Health. Originally formed within the DICE Group, the XR (extended reality) Lab works to create a collaborative environment in which health care meets XR technology. XR is an all-encompassing term for augmented, virtual, and mixed reality technology. This team studies the potential benefits of XR technology for education and patient care and has pursued projects using XR, including patient experiences during bone marrow biopsies, advanced cardiac life support training for resident physicians, anatomy education for medical students, in-clinic vestibular therapy exercises, and nursing education for patient falls prevention.

Virtual reality (VR) has previously been employed in areas of health care, ranging from preoperative patient anxiety [6] to poststroke patient education [7], but it has rarely been used in conjunction with CR. Most notably, several studies have examined VR in combination with stage 3 CR when patients continue their physical activity at home. One survey revealed that 64% of patients felt motivated to continue exercising with the VR application after primary program completion [8]. At the time of this project's development, there was a lack of conclusive evidence about VR's effect on stage 2 outpatient rehabilitation.

Objectives

This study intends to address the lack of interest and comprehension exhibited by many CR patients, which may contribute to incomplete treatment. Education was directly incorporated into patient time on the treadmill, thereby increasing the number of patient exposures to cardiac health information. In addition, digital technologies, including VR, were explored as an alternative way to increase patient engagement with care. Overall, this study aims to evaluate whether employing a VR program that incorporated patient education could increase patients' motivation, understanding, and adherence to CR treatment.

Methods

Study Design

The study procedures were approved by the Jefferson Institutional Review Board (IRB) in December 2017. Beginning in January 2018, participants were selected from patients enrolled in CR at the Jefferson Health Methodist CR program on a rolling basis. Sample size calculations indicated that 68 patients were required to achieve a target power of 90%. Patients were randomized via the Google random number generator, limited to a range of 1-2—1 signifying the control group and 2 signifying the intervention group. To prevent patient bias, the consenting process did not include VR terminology or any reference to VR walking trails.

Participants

Participants were recruited between January 2018 and May 2019. All patients were emailed or mailed a copy of the consent form to review with the rest of their welcome materials before visiting the facility. At patients' first visit, the facility staff stratified patients according to their risk: low, moderate, or high, as per facility standard guidelines. Initially, only moderate-risk patients were enrolled, but this was subsequently expanded to all risk-level stratifications (following approval from the IRB in July 2018) to increase enrollment numbers. Patients were deemed eligible to participate after agreement of 2 study staff members (Textbox 1). If the staff determined that a patient was eligible to participate, they were consented in a private room on the same day.

Textbox 1. Eligibility criteria for participation in trial.

Eligibility criteria

- Aged 18 years or older
- Ability to use a treadmill independent of aid (eg, walker and cane)
- Medically safe to use a treadmill for 15 minutes
- Ability to understand English
- Ability to give their own consent

At this time, patients were scheduled according to their assigned study group—control patients at 9 AM and 11 AM sessions and intervention patients at 8 AM, 10 AM, and 1:30 PM sessions. Patients who could not attend during one of their randomized time slots were excluded from the study and scheduled at their preferred time. Normally, the facility accommodates 10 patients per session. To account for space limitations, VR participants were limited to 5 per session to allow for adequate time with 1 of the 2 VR systems. The other slots in sessions that already contained 5 VR participants were filled by nonparticipants.

Admission to the study continued until 72 patients were enrolled: 31 in the control group and 41 in the intervention group. Patients in both the control and intervention groups completed a 6-minute walk test (6MWT) at their introductory visit to establish a baseline.

Intervention

Study Procedures

Immediately after consenting to participate, patients were given a 5-question education quiz (Textbox 2) to establish their baseline cardiac education level.

Textbox 2. Education questions for pretest, posttest, and 2-month follow-up test.

Education quiz

- How often should you exercise?
- What are the most important things to consider when grocery shopping for heart-healthy foods?
- How often should you take your blood pressure medication?
- What are the symptoms of heart failure?
- What type of medicine is used to control your cholesterol?

For their first 3 visits, participants spent approximately 5 minutes on the treadmill, increasing at intervals of 1 minute if deemed appropriate by the staff. The remainder of each session took place on other types of exercise equipment, as per standard of care. Beginning at their fourth visit, patients were allowed to use the treadmill for up to 15 minutes at each visit. Patients walked at a fixed walking speed as self-paced treadmills are not used at the facility. In addition, research has shown that participants using fixed walking speeds had increased stride length compared with participants on self-paced treadmills [9].

All patients were scheduled to complete between 18 and 36 sessions, as decided by their care team and insurance provider. Typically, low-risk patients were encouraged to complete between 18 and 24 sessions, moderate-risk patients were encouraged to complete between 24 and 30 sessions, and high-risk patients were encouraged to complete between 30 and 36 sessions. CR staff recommended that patients complete the maximum number of possible sessions for their risk level, but the final number was decided by the patient.

Patients were contacted again 2 months after their final CR treatment. It was decided to follow up with patients at 2 months postrehabilitation, as, on average, patients were expected to complete approximately 2 months of treatment. This phone call was designed to follow up on patients' health following treatment and to assess their retention of education by repeating the 5-question education test conducted at their initial and final visits. Previous studies have shown that the use of VR increases the retention of education in students learning complex anatomical concepts [10]. Patients who could not be reached by phone were contacted via email. Those who stopped attending CR sessions before their determined treatment length were also contacted to follow-up and debrief from the study. At least three attempts were made to contact each participant before considering them lost to follow-up.

Control

Patients allocated to the control group received standard of care CR, with the option for additional time on the treadmill, up to 15 minutes, beginning at session 4. Standard of care CR includes completing time on 4 types of exercise equipment, including treadmills, stationary bikes, ellipticals, and hand rowing machines.

Another standard part of CR is education on heart health. Patients are exposed to educational materials to help them understand how to better care for their heart during recovery and to give them tools to continue improving their health in the future. Control patients received standard of care education through handouts, bulletin board displays, and periodic group lectures.

Finally, participants were debriefed at their last scheduled session and informed about the VR intervention. At this time, they were able to try the VR walking trails if desired.

VR Intervention

Participants in the VR intervention group completed the standard of care CR for the majority of their exercise, except for time spent on the treadmill. While using the treadmill, they used the Bionautica Trails system [11], a VR walking trails platform designed and produced by Plas.md [12]. Nursing staff logged each patient with their unique study ID and allowed them to select 1 of 6 walking trails, including 5 nature trails of different themes and 1 outer space–themed trail (Multimedia Appendix 1).

At each log-in, patients had the option to continue from their saved place on a trail from the previous session or select a new one. As patients walked, visual tokens popped up on screen, and when the patient *walked through* them, a randomly chosen piece of cardiac education information was triggered to start (Multimedia Appendix 2). These were programmed to filter through all 109 audio files before repeating. These audio files were taken directly from the handouts and CR textbooks used by the nursing staff. A 2015 study showed that patients who received both auditory and visual cardiac education were more likely to participate in their treatment [13].

Materials

The VR platform, Bionautica Trails, was set up as a high-definition flat-screen television, oriented vertically on a stand in front of 1 of 2 treadmills designated for the study (Figure 1). This system was selected over head-mounted VR options for the safety of the patient. Each of the 2 systems was run on an individual computer hard drive and connected to a Bluetooth keyboard with a built-in trackpad. For the purposes of the study, the Bionautica Trails platform was not updated and remained frozen in the version initially used to maintain consistency between participant experience regardless of when they started treatment.

The educational audio files each intervention group patient heard during their treatment were recorded in a deidentified cloud-based server, via Microsoft Azure, to track individual exposure to education and frequency. Patients used on-ear, wireless headphones, connected to the walking trails via Bluetooth, to better immerse themselves and allow other patients to continue exercising without distraction. These were noise reducing but not noise canceling to ensure that patients could communicate properly with nursing staff. Headphones were disinfected using AF3 antibacterial wipes provided by the facility, as per standard infectious disease guidelines.

Figure 1. Vertically oriented television screen in front of a treadmill showing an example of Bionautica Trails.

Instruments

Survey Tool

At their last scheduled session, after finishing their exercise, patients completed a web-based survey housed on the Qualtrics platform [14], a Health Insurance Portability and Accountability Act–compliant survey tool, through an anonymous link. This survey was designed by the authors to address the patient population at the Jefferson Health Methodist CR facility. Patients completed the survey using their study ID number rather than any identifying information, and only 1 member of the study staff had access to responses for analysis to protect patient privacy. The key variables measured in this self-report questionnaire were knowledge retention, patient satisfaction, and engagement (Multimedia Appendix 3).

Knowledge Retention

Knowledge retention for this study was defined as the maintenance of cardiac health information over time. Patients were asked a series of 5 questions (Textbox 2) at 3 time points: their first visit, last visit, and 2-months post treatment. These questions were identified by nursing staff as being vital to patients' recovery and common areas of confusion for patients. Knowledge retention was scored out of 5 for all 3 time points, with each question worth 1 point.

Patient Satisfaction

Patient satisfaction is contentment with the care received and the overall experience of a health care interaction. Patients completed 6 questions rated on a 4-point Likert-type scale (extremely dissatisfied=1; dissatisfied=2; satisfied=3; extremely satisfied=4). Each question covered a different subtopic of satisfaction and was scored individually.

Engagement

Patient engagement is defined as an intrinsic interest in and participation in care. Patients completed 3 questions on their engagement, rating each feature on a scale of 1-10 wherein 1 was “not at all engaged” and 10 was “extremely engaged.”

Statistical Analysis

This study was powered by an increase in the distance traveled during the 6MWT. The 6MWT is considered the functional walk test of choice for assessing clinical improvement in cardiorespiratory patients [15]. To estimate the expected improvement, equivalent data from patients with chronic obstructive pulmonary disease were used, which suggested an average improvement of 446 m (SD 82 m) for moderate-risk patients [16]. It was considered reasonable that the experimental group would achieve an additional 15% increase in the distance walked over that of the control group or an additional 67 m in 6 minutes because of their increased motivation and education from the intervention. Power analysis showed that to detect a between-group difference in the 6MWT at a power level of 0.90 and an α of .05, 34 participants were required in each group.

The 6MWT analysis was conducted using a *t* test allowing unequal variances; the Levene test for the equality of variances was used to determine equal variances for 6MWT improvement. For knowledge retention, a 2-way multivariate analysis of variance (MANOVA) test was used to compare improvement between groups at each of the following time intervals: pretest to posttest, posttest to follow-up, and pretest to follow-up. The MANOVA root included calculations for the Pillai trace statistic, the Wilks λ statistic, the Hotelling trace criterion, and the Roy largest test.

Results

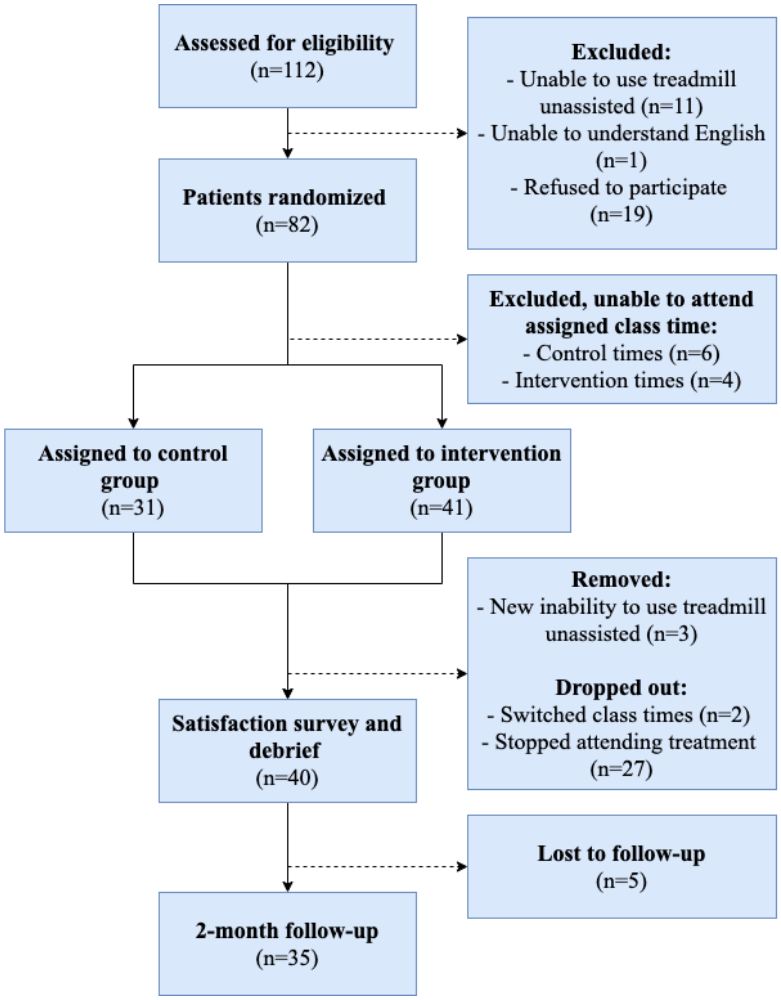
Patient Demographics

Between January 2018 and May 2019, 72 patients (male: n=52; female: n=20; age: range 32-81 years) were enrolled in the study to either the control (n=31) or intervention (n=41) group (Table 1). Of these, 3 participants were excluded from the study because of an inability to continue exercise on the treadmill, and 35 participants (control: n=19; intervention: n=16) completed all stages of the study (Figure 2).

Table 1. Patient demographic data (N=72).

Characteristic	Patients
Age (years), mean (SD)	61 (9.9)
Age (years), n (%)	
<40	2 (3)
40-49	7 (10)
50-59	19 (26)
60-69	32 (44)
70-79	11 (15)
≥80	1 (1)
Sex, n (%)	
Male	52 (72)
Female	20 (28)
Race, n (%)	
Asian	2 (3)
Black (non-Hispanic)	21 (29)
Hispanic	3 (4)
White (non-Hispanic)	45 (63)
Other	1 (1)

Figure 2. Flow diagram of research design and patient completion status.



Outcomes

6MWT

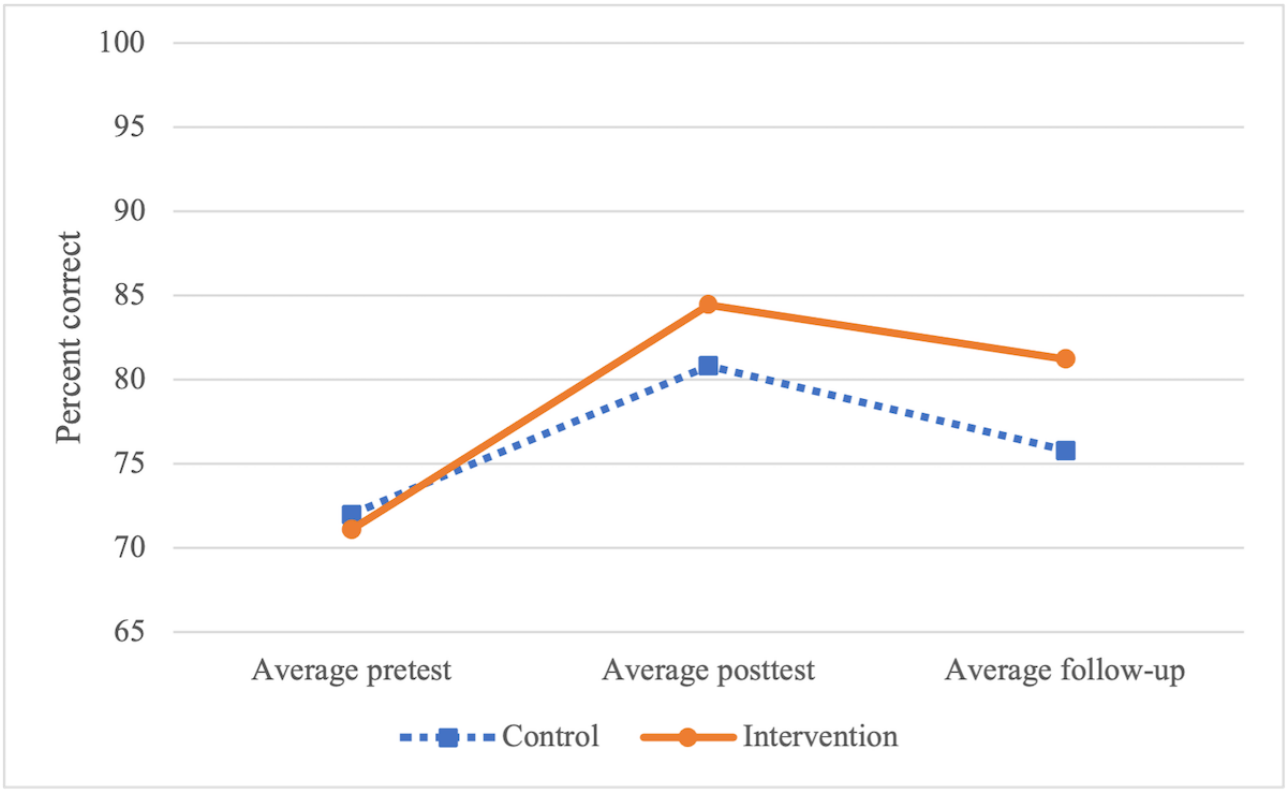
All patients (N=72) completed the 6MWT during their first treatment session. All patients who completed the program (n=34) completed a postrehabilitation 6-minute walk test (post-6MWT) as well. In addition, patients who chose to terminate their program early by notifying staff completed the post-6MWT (n=6) at their last scheduled session. All patients who did not finish treatment and did not notify staff before terminating their participation did not complete the post-6MWT (n=32). For these patients, their final session's distance walked was used to estimate their post-6MWT results for primary outcome improvement evaluation.

A blinded interim analysis was conducted in January 2019, which showed no statistical difference between improved distance walked relative to the 6MWT (P=.60; 95% CI −196.42 to 329.12). As no increased risk was seen in either group, it was decided to continue with the original intended sample size for complete analysis of the primary and secondary outcomes. At the final review, patients in both groups showed improvement between pre- and post-6MWT distances. The control group improved by an average of 298 feet and intervention group by 340 feet, without significance between groups (P=.64; 95% CI −224.14 to 139.52).

Knowledge Retention

Patient knowledge retention was evaluated at 3 time points: at the first visit, at the last visit, and at 2 months post treatment. Patients were asked to answer 5 multiple-choice questions at all time points (Textbox 2). No significant difference was seen between groups for baseline (P=.88; 95% CI −0.53 to 0.62), final visit (P=.56; 95% CI −0.79 to 0.43), or follow-up (P=.50; 95% CI −1.09 to 0.54) quizzes (Figure 3). Overall, patients in the intervention group improved by an average of 10.1% between pretest and follow-up compared with 3.8% for the control group (P=.86; 95% CI 0.71-0.59). Patients in the intervention group received an average of 115 pieces of audio education from a bank of 109 audio files, at an average of every 100 seconds of walk time.

Figure 3. Average education test scores for pretest, posttest, and 2-month follow-up.



Patient Satisfaction and Engagement

At their final visit, the patients completed a survey on their satisfaction. Patients were asked to rate their satisfaction with the education they received and their CR experience, based on several subtopics (Table 2). Participants from both groups responded positively to their treatment, and no significant difference was observed between the groups. Patients were also asked to rate their enjoyment of time spent on the treadmill (Table 3) and their engagement with this portion of their treatment (Multimedia Appendix 4). Additional questions were asked about patient engagement overall and specific to the education they received to help understand the breakdown of patient engagement. For education specific engagement, 56% (22/39) of patients selected a rating of 10 out of 10. For overall engagement, 70% (28/40) of patients selected a rating of 10 out of 10. Again, no significant differences were observed between the groups.

Table 2. Satisfaction survey questions (control: n=21 and intervention: n=19).

Question	Control, mean score (SD)	Intervention, mean score (SD)	*P* value
How satisfied are you with your CR[a] experience?			
How satisfied are you with your CR experience overall?	3.7 (0.7)	3.9 (0.3)	.32
How satisfied are you with your time spent on the treadmill?	3.5 (0.7)	3.6 (0.5)	.45
How satisfied are you with your interactions with staff?	3.9 (0.6)	4.0 (0.0)	.35
How satisfied are you with your CR education?			
How satisfied are you with the delivery of your CR education?	3.8 (0.4)	3.5 (0.5)	.06
How satisfied are you with the personalization of your CR education?	3.7 (0.5)	3.5 (0.5)	.23
How satisfied are you with the clarity of information?	3.8 (0.4)	3.6 (0.5)	.23

[a]CR: cardiac rehabilitation.

Table 3. Patients' response to "During most of your treatment sessions, you were allowed to use the treadmill for up to 15 minutes. Did you enjoy the time spent on the treadmill?"

Answer choice	Control, n (%)	Intervention, n (%)
"Yes, I would have enjoyed even more than the time I was allowed to spend on the treadmill."	6 (29)	5 (26)
"Yes, I enjoyed the time I spent on the treadmill. I did not need more than 15 minutes. It was the right amount of time for me."	13 (62)	14 (74)
"No, I did not enjoy it, so I did not use the full 15 minutes."	0 (0)	0 (0)
"No, I did not enjoy the treadmill. I do not enjoy that form of exercise."	1 (5)	0 (0)
"I am undecided."	1 (5)	0 (0)

Patients in the intervention group also completed the VR-specific questions (Table 4). Most participants responded negatively ("disagree") to negative statements and positively ("agree") to positive statements. Notably, the statement "I wanted to spend longer on the treadmill" resulted in a mix of positive, negative, and undecided responses.

Table 4. Virtual reality–specific satisfaction (n=19).

Question—"Because of the virtual reality walking trails..."	Disagree, n (%)	Undecided, n (%)	Agree, n (%)
"I looked forward to treatment sessions more."	1 (5)	5 (26)	13 (68)
"I enjoyed my treatment sessions more."	1 (5)	0 (0)	18 (95)
"I felt more engaged in my treatment."	1 (5)	1 (5)	17 (89)
"I wanted to spend longer on the treadmill."	7 (37)	6 (32)	6 (32)
"I gained a better understanding of my cardiac health."	2 (11)	0 (0)	17 (89)
"I did not see any impact on my treatment."	13 (68)	1 (5)	5 (26)
"I was unwantedly distracted from my treatment."	16 (84)	1 (5)	2 (11)
"I did not enjoy how I received my education."	17 (89)	0 (0)	2 (11)
"I felt isolated from the rest of my class."	15 (79)	0 (0)	4 (21)
"I dreaded my time on the treadmill."	15 (79)	1 (5)	3 (16)

Attendance

In addition, total patient attendance was recorded by group (Table 5). Each risk stratification had a different recommended number of sessions, but the Jefferson Health Methodist CR facility requires all patients to complete at least 18 sessions to be considered finished with treatment. As such, attendance was evaluated through both the recommended number of sessions and the completion of at least 18 sessions. Of the patients who were removed from or dropped out of the study (n=32), 4 control and 3 intervention patients continued with or returned to CR treatment after study dropout and were thus omitted from the attendance totals.

The comparison of attendance between each risk group showed that the control group had significantly higher completion rates (P=.02; 95% CI 0.04-0.53), as did the comparison of the rates for completing 18 sessions (P=.046; 95% CI 0.00-0.47). However, patients who stopped attending treatment sessions were asked to explain their choice, and no causal relationship was seen between poor attendance and VR walking trails. The reasons for ending treatment sessions early were mainly associated with other health issues (n=7), such as chronic pain, or a need to return to work (n=8); of those who needed to return to work, 2 patients switched to a new class time and 1 patient later returned to treatment but had already been debriefed about the study. Of the 29 patients who stopped attending treatment sessions, 11 left without explanation and could not be reached for follow-up.

Table 5. Patient attendance by completion of minimum required sessions and by recommended number of sessions.

Attendance	Control, n (%)	Intervention, n (%)
Completion of the minimum required treatment sessions		
Fewer than the required 18 treatment sessions completed	5 (19)	16 (42)
18 or more treatment sessions completed	22 (81)	22 (58)
Completion of the recommended number of treatment sessions		
Fewer than the recommended number of treatment sessions	8 (30)	22 (58)
Met or exceeded the recommended number of treatment sessions	19 (70)	16 (42)

Discussion

Principal Findings

This randomized controlled trial (RCT) focused on improving patients' experiences during CR treatment through an alternative VR solution. To the best of our knowledge, this is the first RCT focusing on VR for knowledge retention and motivation conducted in stage 2 outpatient rehabilitation. Previous studies related to VR's usage in stage 3 CR positively affected patient motivation compared with controls. The results demonstrated that there were no statistically significant outcomes, although this study provides valuable insights into the application of a virtual tool in outpatient CR.

Primary Outcome

It was hypothesized that VR-enhanced outpatient CR would be superior to conventional outpatient CR in improving 6MWT distances. This was chosen as the primary outcome as a means of standardization, as it is a common measure of improvement for patients undergoing CR. However, the results indicated that the intervention and control groups had comparable improvements in their 6MWT distances. These results may be related to limitations of the 6MWT, including its use of patient-initiated walking speed rather than the manual speed of the treadmills, and the fact that patients completed exercises on an average of 4 different types of equipment per session. These factors may have limited the 6MWT's ability to capture the value that VR can bring to a CR environment.

Furthermore, 44% (32/72) of patients did not complete a post-6MWT, severely limiting the sample size. We decided to use these patients' final treadmill sessions and extrapolate their distance walked in 6 minutes. However, as these were from manually set speeds, this extrapolation did not capture patient effort during the 6MWT, which may have been impacted by using Bionautica Trails.

Secondary Outcomes

Patient Engagement

Patient engagement was chosen as a secondary outcome because of evidence that VR has motivational benefits for patients in other rehabilitation settings, including during motor rehabilitation [17] and stroke rehabilitation [18].

Subjective measures of engagement were assessed through the patient survey, but an attempt was made to evaluate an objective measure of engagement based on the time spent on the treadmill. Patients were told they had up to 15 minutes of time on the treadmill, a duration chosen by the nursing care team to ensure patient safety and also fit within regularly scheduled sessions. By allowing patients to choose how long to spend on the treadmill, the hope was that they would spend time on the treadmill proportional to their interest level, based on the understanding that when participants are intrinsically motivated, they adhere better to exercise regimens [19].

This inference proved to be incorrect, as 62.5% (45/72) of patients, regardless of group, used the full 15 minutes at 60% of sessions or more. Those who did not use the full time typically cited coexisting health issues, such as arthritic joints, as reasoning for completing less than the full 15 minutes. Most patients likely chose to complete the full 15 minutes because many patients, particularly those of this age group, hear health care providers' instructions as recommendations rather than as choices [20]. To improve how engagement can be better assessed in a similar outpatient environment, future studies should explore how the phrasing of patient instructions affects their choices.

Patient Satisfaction

It was unlikely that the intervention would improve overall satisfaction, as at baseline, the Jefferson Health Methodist CR center already had incredibly high satisfaction scores. This has been largely attributed to the incredible care of the nursing staff. Patients reported that "the positive attitude and support of the staff was THE most critical factor" and the "staff took time to check-in regularly on physical and emotional health - Program clearly help[ed] me regain and improve my overall health not just cardiac."

Although no differences were seen between groups in terms of satisfaction, it is worth noting that many patients responded positively to Bionautica Trails. One participant reported, "It took me out of the room for a little bit; I really imagined hiking somewhere with my brothers," and others highlighted how "the walking trails made the walking experience enjoyable." The few patients who did not respond positively cited that they felt isolated from the social environment of the center or felt that the trails became repetitive. Although VR may not satisfy all patients, it should be mentioned that Bionautica Trails is still in use at the Jefferson Health Methodist CR center. Patients in both the treatment and maintenance programs use it regularly, and the nursing staff appreciate having the virtual trails as a tool for maintaining the focus of patients who need more attention.

Knowledge Retention

Previous research suggests that VR has the potential to improve patient comprehension of information [21] and long-term knowledge retention in students [10,22], possibly through spatial learning [22] or by increasing motivation to learn [23]. The knowledge retention test was limited to only 5 questions to minimize the amount of time added to required sessions, but it appears that this may have been too few questions to show a distinction between groups.

In addition, the questions themselves may have been too simplistic; at their initial visit, 86% (62/72) of patients knew that statins are used to treat cholesterol and 94% (68/72) knew how often blood pressure medication should be taken. It is difficult to show a significant change when there is little room for improvement. A more challenging set of questions, including a wider variety of complexities, should be used in future studies measuring the educational value of VR in this setting.

Attendance

Patient completion of the recommended number of rehabilitation sessions indicated statistical significance in favor of the control group. One possible explanation for this is that the enjoyment of VR could not overcome the considerable social and economic

factors that patients face, which contribute to low attendance [3]. Patients who stopped attending treatment sessions were asked to explain their choice, and no causal relationship was seen between poor attendance and virtual walking trails. One of the most common reasons for dropping out was the need to return to work (n=8). Patients would often say that they had to miss work to attend treatment or that their job would only cover so much sick leave. In addition, although only 1 patient reported dropping out of treatment because of their insurance, several patients completed only 2 days per week of treatment—rather than the recommended 3—to keep copay costs at a minimum.

Given the complexity of balancing work, finances, and treatment schedules, VR did not work as a tool to increase patient attendance at CR sessions. However, it highlights the importance of introducing digital tools in conjunction with practical strategies that address patients' needs.

Patient Safety

In total, 3 patients were excluded from the study because of a new inability to use the treadmill. Two of these patients experienced a change in their ability to use the treadmill related to their preexisting health conditions. The third patient discontinued participation in the trial because of self-described motion sickness, also referred to in the literature as "cybersickness" [24]. During the course of the study, 41 patients collectively completed 139.6 hours of virtual walking trails—almost 6 full days of content—and the third patient was the only patient to report any adverse side effects. Although cybersickness should still be considered when conducting research, the exceptionally low incidence rate in this study supports the use of VR regardless of age or previous experience using VR technology.

Blinding

Although not directly correlated to outcomes, it is worth discussing the difficulties faced when designing RCTs using VR. The nature of VR means that neither patients nor onsite study staff could be blinded to the study arm. Instead, the terms "virtual reality" and "virtual walking trails" were never mentioned during the consent process to blind patients to the study goal and limit response bias. Therefore, the patients were debriefed on their participation at their final treatment visit. At that time, those in the control group were offered the chance to try the walking trails. Most participants did not wish to try them when approached at the end of their last session, likely because they had already completed their workout and did not wish to spend additional time on the treadmill. However, post-6MWTs were occasionally conducted 1-2 sessions before the patient's expected end date, typically when multiple patients were expected to finish on the same day or if there was concern that the patient might not return for the final treatment session. Those who completed their post-6MWT early often chose to use Bionautica Trails for their final time(s) on the treadmill. It is unclear why this trend occurred, but based on qualitative responses from patients, it supports positive patient engagement with the trails. One such patient, when asked if they would have enjoyed the VR walking trails during regular treatment sessions, commented:

> *I feel like I missed out. It made the time pass so fast. It was great. It's different when you're looking at a picture, it makes it go quick [sic]*

Future Research

The use of VR in rehabilitation is a growing field, and there are numerous potential avenues for further research. Adaptations to this project could include expanding Bionautica Trails to other types of exercise equipment, such as a recumbent bike, to assess how continuity between exercise equipment and increased exposure would affect results. It would also be valuable to conduct a similar trial with a more robust education component, particularly expanding the number of questions and including variable question complexity.

Beyond this specific trial, research should continue to explore the gamification of treatment. In the version of Bionautica Trails used for this study, each token that patients *walked through* contributed to an overall score. These scores remained private, but the creation of a social network scoreboard could potentially increase the motivation to continue walking [25]. Furthermore, the newest version of Bionautica Trails includes a story mode that may encourage repeated use and promote behavioral changes through narrative communication [26]. Although it may still not increase attendance rates of this particular patient population, given the socioeconomic factors discussed previously, this addition may address patients' complaints about the repetitive nature of the trails. In addition, the use of a story mechanic lends itself particularly well to at-home or stage 3 treatment for maintaining user engagement without the external motivators inherent to the gym setting. Further research is needed to determine the most appropriate setting and population for using VR in the CR treatment process.

Conclusions

Given the importance of CR treatment, it is vital to continue studying methods of improving patient access to care and patient experience while in care. Despite the need for CR, many patients do not complete treatment. Although this particular tool did not show statistically significant improvements in outcomes, it provided anecdotal positive reactions from patients and their health care providers. The limitations of the study design likely contributed to the final results. Further studies are needed to determine whether virtual walking trails are worth implementing in stage 2 CR, particularly accounting for the numerous socioeconomic factors that influence patient access to care. VR tools such as Bionautica Trails should continue to be explored to improve the overall patient experience.

Acknowledgments

This study was internally funded by the DICE Group at Jefferson Health. Hardware and software for Bionautica Trails were provided by Plas.md for use in this study. Publication was made possible in part by support from the Thomas Jefferson University

Open Access Fund. The authors would like to thank the Jefferson Health Methodist CR facility staff, especially Mary Lou O'Neill, Barb Pettit, and Dottie Grimes, for their support and expertise.

Authors' Contributions

All authors contributed to this study and were involved in the conception and study design. Primary data collection was completed by VG. Statistical analysis was completed by Maclain Capron, the statistical analyst for DG's office. The manuscript was drafted by VG. All authors approved the final version of the manuscript for submission.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Additional images of Bionautica Trails.
[PDF File (Adobe PDF File), 18282 KB - jmir_v23i4e23882_app1.pdf]

Multimedia Appendix 2
Video demonstrating Bionautica Trails in action.
[MP4 File (MP4 Video), 22619 KB - jmir_v23i4e23882_app2.mp4]

Multimedia Appendix 3
Complete patient survey.
[PDF File (Adobe PDF File), 123 KB - jmir_v23i4e23882_app3.pdf]

Multimedia Appendix 4
Patient response to the question "On a scale of 1 to 10, how engaged were you with your cardiac rehabilitation experience?".
[PNG File , 206 KB - jmir_v23i4e23882_app4.png]

Multimedia Appendix 5
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 462 KB - jmir_v23i4e23882_app5.pdf]

References

1. Ades PA, Keteyian SJ, Wright JS, Hamm LF, Lui K, Newlin K, et al. Increasing cardiac rehabilitation participation from 20% to 70%: a road map from the million hearts cardiac rehabilitation collaborative. Mayo Clin Proc 2017 Feb;92(2):234-242 [FREE Full text] [doi: 10.1016/j.mayocp.2016.10.014] [Medline: 27855953]
2. Blum MR, Schmid J, Eser P, Saner H. Long-term results of a 12-week comprehensive ambulatory cardiac rehabilitation program. J Cardiopulm Rehabil Prev 2013;33(2):84-90. [doi: 10.1097/HCR.0b013e3182779b88] [Medline: 23385556]
3. Karmali KN, Davies P, Taylor F, Beswick A, Martin N, Ebrahim S. Promoting patient uptake and adherence in cardiac rehabilitation. Cochrane Database Syst Rev 2014(6):CD007131. [doi: 10.1002/14651858.CD007131.pub3] [Medline: 24963623]
4. Kerins M, McKee G, Bennett K. Contributing factors to patient non-attendance at and non-completion of Phase III cardiac rehabilitation. Eur J Cardiovasc Nurs 2011 Mar;10(1):31-36. [doi: 10.1016/j.ejcnurse.2010.03.006] [Medline: 20457543]
5. Mikkelsen T, Korsgaard Thomsen K, Tchijevitch O. Non-attendance and drop-out in cardiac rehabilitation among patients with ischaemic heart disease. Dan Med J 2014 Oct;61(10):A4919. [Medline: 25283618]
6. Ganry L, Hersant B, Sidahmed-Mezi M, Dhonneur G, Meningaud J. Using virtual reality to control preoperative anxiety in ambulatory surgery patients: A pilot study in maxillofacial and plastic surgery. J Stomatol Oral Maxillofac Surg 2018 Sep;119(4):257-261. [doi: 10.1016/j.jormas.2017.12.010] [Medline: 29317347]
7. Thompson-Butel AG, Shiner CT, McGhee J, Bailey BJ, Bou-Haidar P, McCorriston M, et al. The role of personalized virtual reality in education for patients post stroke-a qualitative case series. J Stroke Cerebrovasc Dis 2019 Feb;28(2):450-457. [doi: 10.1016/j.jstrokecerebrovasdis.2018.10.018] [Medline: 30415917]
8. Vieira Á, Melo C, Machado J, Gabriel J. Virtual reality exercise on a home-based phase III cardiac rehabilitation program, effect on executive function, quality of life and depression, anxiety and stress: a randomized controlled trial. Disabil Rehabil Assist Technol 2018 Feb;13(2):112-123. [doi: 10.1080/17483107.2017.1297858] [Medline: 28285574]
9. Sloot LH, van der Krogt MM, Harlaar J. Effects of adding a virtual reality environment to different modes of treadmill walking. Gait Posture 2014 Mar;39(3):939-945. [doi: 10.1016/j.gaitpost.2013.12.005] [Medline: 24412269]

10. Ekstrand C, Jamal A, Nguyen R, Kudryk A, Mann J, Mendez I. Immersive and interactive virtual reality to improve learning and retention of neuroanatomy in medical students: a randomized controlled study. CMAJ Open 2018 Feb 23;6(1):E103-E109 [FREE Full text] [doi: 10.9778/cmajo.20170110] [Medline: 29510979]
11. Plas.md Inc. Bionautica trails. URL: https://www.bionautica.com [accessed 2020-08-28]
12. Plas.md Inc. Plas.md Inc. URL: https://www.plas.md/ [accessed 2020-08-28]
13. Dankner R, Drory Y, Geulayov G, Ziv A, Novikov I, Zlotnick AY, et al. A controlled intervention to increase participation in cardiac rehabilitation. Eur J Prev Cardiol 2015 Sep;22(9):1121-1128. [doi: 10.1177/2047487314548815] [Medline: 25183694]
14. Online survey software. Qualtrics. URL: https://www.qualtrics.com/core-xm/survey-software/ [accessed 2020-08-28]
15. Solway S, Brooks D, Lacasse Y, Thomas S. A qualitative systematic overview of the measurement properties of functional walk tests used in the cardiorespiratory domain. Chest 2001 Jan;119(1):256-270. [doi: 10.1378/chest.119.1.256] [Medline: 11157613]
16. Valerio G, Bracciale P, Valerio F. Metabolic requirements during six minutes walking tests in patients affected by chronic obstructive pulmonary disease in different stages. Open J Resp Disease 2012;2(4):83-90 [FREE Full text] [doi: 10.4236/ojrd.2012.24012]
17. Zimmerli L, Jacky M, Lünenburger L, Riener R, Bolliger M. Increasing patient engagement during virtual reality-based motor rehabilitation. Arch Phys Med Rehabil 2013 Sep;94(9):1737-1746. [doi: 10.1016/j.apmr.2013.01.029] [Medline: 23500181]
18. Burke JW, McNeill MD, Charles DK, Morrow PJ, Crosbie JH, McDonough SM. Optimising engagement for stroke rehabilitation using serious games. Vis Comput 2009 Aug 27;25(12):1085-1099. [doi: 10.1007/s00371-009-0387-4]
19. Ryan R, Frederick C, Lepes D, Rubio N, Sheldon K. Intrinsic motivation and exercise adherence. Int J Sport Psychol 1997;28:335-354 [FREE Full text]
20. Agarwal AK, Murinson BB. New dimensions in patient-physician interaction: values, autonomy, and medical information in the patient-centered clinical encounter. Rambam Maimonides Med J 2012 Jul;3(3):e0017 [FREE Full text] [doi: 10.5041/RMMJ.10085] [Medline: 23908841]
21. Vertemati M, Cassin S, Rizzetto F, Vanzulli A, Elli M, Sampogna G, et al. A virtual reality environment to visualize three-dimensional patient-specific models by a mobile head-mounted display. Surg Innov 2019 Jun;26(3):359-370. [doi: 10.1177/1553350618822860] [Medline: 30632462]
22. Krokos E, Plaisant C, Varshney A. Virtual memory palaces: immersion aids recall. Virtual Real 2018 May 16;23(1):1-15. [doi: 10.1007/s10055-018-0346-3]
23. Stepan K, Zeiger J, Hanchuk S, Del Signore A, Shrivastava R, Govindaraj S, et al. Immersive virtual reality as a teaching tool for neuroanatomy. Int Forum Allergy Rhinol 2017 Oct;7(10):1006-1013. [doi: 10.1002/alr.21986] [Medline: 28719062]
24. Weech S, Kenny S, Barnett-Cowan M. Presence and cybersickness in virtual reality are negatively related: a review. Front Psychol 2019;10:158 [FREE Full text] [doi: 10.3389/fpsyg.2019.00158] [Medline: 30778320]
25. Tate DF, Lyons EJ, Valle CG. High-tech tools for exercise motivation: use and role of technologies such as the internet, mobile applications, social media, and video games. Diabetes Spectr 2015 Jan;28(1):45-54. [doi: 10.2337/diaspect.28.1.45] [Medline: 25717278]
26. Hinyard LJ, Kreuter MW. Using Narrative Communication as a Tool for Health Behavior Change: A Conceptual, Theoretical, and Empirical Overview. Health Educ Behav 2007 Oct;34(5):777-792. [doi: 10.1177/1090198106291963] [Medline: 172000094]

Abbreviations

6MWT: 6-minute walk test
CR: cardiac rehabilitation
DICE: Digital Innovation and Consumer Experience
IRB: Institutional Review Board
MANOVA: multivariate analysis of variance
Post-6MWT: postrehabilitation 6-minute walk test
RCT: randomized controlled trial
VR: virtual reality
XR: extended reality

Edited by R Kukafka; submitted 28.08.20; peer-reviewed by C Miranda, A Azzam; comments to author 15.10.20; revised version received 10.12.20; accepted 16.03.21; published 15.04.21.

Please cite as:
Gulick V, Graves D, Ames S, Krishnamani PP
Effect of a Virtual Reality–Enhanced Exercise and Education Intervention on Patient Engagement and Learning in Cardiac Rehabilitation: Randomized Controlled Trial
J Med Internet Res 2021;23(4):e23882
URL: https://www.jmir.org/2021/4/e23882
doi:10.2196/23882
PMID:33856355

Original Paper

"Doc McStuffins: Doctor for a Day" Virtual Reality (DocVR) for Pediatric Preoperative Anxiety and Satisfaction: Pediatric Medical Technology Feasibility Study

Jeffrey I Gold[1,2], PhD; Erin T Annick[1], BA; Arianna S Lane[1], BA; Katherine Ho[1]; Ryan T Marty[1]; Juan C Espinoza[3,4], MD, FAAP

[1]Department of Anesthesiology Critical Care Medicine, The Saban Research Institute, Children's Hospital Los Angeles, Los Angeles, CA, United States

[2]Departments of Anesthesiology, Pediatrics, and Psychiatry & Behavioral Sciences, Keck School of Medicine, University of Southern California, Los Angeles, CA, United States

[3]Division of General Pediatrics, Department of Pediatrics, The Saban Research Institute at Children's Hospital Los Angeles, The West Coast Consortium for Technology & Innovation Pediatrics, Los Angeles, CA, United States

[4]Department of Pediatrics, Keck School of Medicine, University of Southern California, Los Angeles, CA, United States

Corresponding Author:
Jeffrey I Gold, PhD
Department of Anesthesiology Critical Care Medicine
The Saban Research Institute
Children's Hospital Los Angeles
4650 Sunset Blvd.
Los Angeles, CA, 90027
United States
Phone: 1 13233616341
Email: JGold@chla.usc.edu

Abstract

Background: Preoperative anxiety is a common occurrence among children and is associated with a host of maladaptive postoperative behaviors. Consequently, increased attention has been placed on interventions to reduce preoperative anxiety and its associated outcomes. Child Life preparation prior to surgery includes evidence-based practices such as age-appropriate distraction and therapeutic play. Virtual reality (VR) is a promising addition to the Child Life toolbox to address anxiety prior to surgery. The current study evaluates the implementation and feasibility of a VR experience, "Doc McStuffins: Doctor for a Day Virtual Reality Experience" (*Doc*VR), developed by Disney Junior in collaboration with Children's Hospital Los Angeles, to target pediatric preoperative anxiety.

Objective: The primary aim of this study was to examine the feasibility and efficacy of *Doc*VR for preoperative anxiety. A secondary aim was to improve patient, caregiver, and health care provider satisfaction with the preoperative experience.

Methods: In this study, 51 patients (age 6-14 years) scheduled for surgery in the ambulatory surgery center and the main operating room at Children's Hospital Los Angeles were approached to participate in Disney's *Doc*VR experience. The patients played the *Doc*VR experience for an average of 18 minutes (3-55 minutes). Irrespective of surgical procedure, patients and their families were eligible, as long as they had no known marked cognitive or visual impairments that would interfere with completing the survey and engaging in the *Doc*VR experience.

Results: Patients who tried the *Doc*VR experience (n=51) responded overwhelmingly positively to both the VR technology and to the game itself. Patients experienced a statistically significant decrease in anxiety following DocVR game play (Z=–3.26, *P*=.001). On the Facial Affective Scale, the percentage of patients who chose the face with the most positive facial expression to represent their affect increased from 23% (12/51) pre-VR to 49% (25/47) post-VR. Furthermore, 97% (38/39) of patients reported feeling more comfortable at the hospital, and 74% (28/38) reported feeling less scared at the hospital after playing the game. The game was enjoyed by 94% (46/49) of patients, and 88% (30/34) of patients reported feeling both "Interested" and "Involved" in the game.

Conclusions: *Doc*VR is a feasible and beneficial VR experience to relieve pediatric preoperative anxiety and improve satisfaction in the preoperative area. The VR experience resulted in a decrease in overall anxiety and an increase in overall positive affect during the preoperative time. Patients also responded positively to the game, confirming their interest in the content and affirming

the quality of the *Doc*VR experience. The positive response to the game indicates that *Doc*VR has the potential to make the overall preoperative experience less anxiety-producing and more comfortable, which leads to improved patient satisfaction. Naturally, improved patient outcomes lead to improved caregiver and health care provider satisfaction.

(J Med Internet Res 2021;23(4):e25504) doi:10.2196/25504

KEYWORDS

virtual reality (VR); pediatric; anxiety; preoperative; satisfaction; Doc McStuffins

Introduction

Nearly 5 million children undergo surgery in the United States annually, and 50-75% of these children experience preoperative anxiety [1]. Preoperative anxiety not only causes distress and suffering in children prior to surgery but is also associated with a slower, more painful recovery and negative postoperative behavior changes, such as separation anxiety, sleep disturbances, eating difficulties, and aggression towards authority [2,3]. Additionally, children who experience high preoperative anxiety are more likely to develop emergence delirium, a state of dissociated consciousness characterized by uncooperativeness, inconsolable crying, irritation, and incoherency that occurs upon waking from anesthesia [4,5].

Techniques for reducing preoperative anxiety in pediatric patients fall into 4 broad categories: preoperative sedatives, parental presence, preparation programs [6], and pain or anxiety management interventions. Although preoperative sedatives are regularly administered prior to surgery, anti-anxiety drugs can synergistically interact with anesthetics to produce undesirable side effects that prolong recovery [7]. Additionally, the use of sedatives may delay hospital discharge and consequently increase operational costs [6,8,9]. To date, research on the effect of parental presence for preoperative anxiety has yielded mixed results; although some studies have documented its efficacy, other studies have found no significant differences between anxiety levels of children whose parents were present versus absent at various stages of the preoperative process [10]. Thus, clinicians and researchers have long been interested in using other nonpharmacological interventions, such as procedural preparation programs and anxiety and stress management interventions, to combat preoperative anxiety. These combination interventions, which include Child Life programs and virtual reality (VR), among others, can work in tandem to provide potentially greater reductions in pediatric preoperative anxiety and distress, thus ultimately diminishing known features associated with medical trauma.

A variety of health care providers, including nurses, physicians, and psychologists, deploy VR for an array of health care–related interventions. In the preoperative space, Certified Child Life Specialists (CCLSs) are often the frontline professionals providing preparation and anxiety management strategies for patients in pediatric facilities. CCLSs are professional health care providers who help patients and their families cope with the challenges of the medical world, including, but not limited to, initial diagnoses, illness, hospitalization, and medical procedures [11]. CCLSs have spent decades continuously adding new innovations to their toolboxes to help children and families manage pain and anxiety associated with medical procedures [12-14]. For example, CCLSs capitalize upon the gate control theory of pain to help children cope with the unpleasant sensations caused by medical environments [15]. By providing engaging distractions like age-appropriate toys, electronic devices, and deep breathing techniques, CCLSs can divert the child's attention away from their upcoming procedure and close the child's "gate" for pain sensation. CCLSs also recognize the role of mental preparation and education in reducing preprocedural pain and anxiety. Numerous studies have shown that preparation programs can significantly reduce children's negative responses to medical procedures [16-20]. Mentally preparing patients and their families for upcoming procedures can involve giving tours of the operating room (OR), allowing children to familiarize themselves with medical equipment, and facilitating medical play sessions with dolls [16]. Whether providing distraction or education, CCLSs help align the perspectives of patients, caregivers, and health care providers to mitigate the oftentimes debilitating stresses of the medical environment.

VR has emerged as a promising nonpharmacological intervention to relieve pediatric preoperative anxiety and improve satisfaction. Incorporation of visual, auditory, and tactile stimulation within a 3D environment allows a child to "escape" to another world. The current study utilized a head-mounted display VR technology, which is the most common mode of technology for VR intervention-based studies. VR has been shown to reduce both pain and anxiety in pediatric populations for a variety of procedures, including phlebotomy [21], peripheral intravenous catheterization access [22], invasive dental procedures [23], and wound care and dressing changes [24]. A majority of pediatric VR studies for anxiety management focus on periprocedural timeframes, as cited earlier in this manuscript, and only a very limited number of studies have examined the use of VR for preoperative anxiety [25,26]. Additionally, there has been a call for a greater focus on anxiety management in VR research [27]. The current study's focus on both the preoperative timeframe and anxiety management makes it particularly relevant as VR research continues to charge forward.

Disney Junior, a television network owned by The Walt Disney Company, is a familiar staple in children's entertainment. Their television shows and online interactive games keep children engaged while reiterating life lessons of bravery, kindness, and friendship [28-30]. One of these shows is "Doc McStuffins: Toy Hospital," which follows a young girl who acts out her dream of becoming a pediatrician on her toys. Cast members from multiple Disney teams created "Doc McStuffins: Doctor for a Day" (*Doc*VR), a VR experience to engage, educate, entertain, and immerse patients into a "Doctor for the Day" role,

virtually conducting routine medical procedures, while waiting for their planned outpatient surgery. A multidisciplinary team from Children's Hospital Los Angeles (CHLA), including psychology, Child Life, and general pediatrics, implemented the Disney Junior *Doc*VR experience. Considering the medical context of the Doc McStuffins show, *Doc*VR is well-matched for pediatric medical centers, as it exposes and normalizes medical experiences, all in the tried-and-true format of an already beloved show. Incorporating VR, and particularly *Doc*VR, into the CCLS's preoperative processes has promising applications to more effectively manage preoperative anxiety.

This undertaking marks the first collaboration between Disney Junior and a children's hospital to build and pilot, respectively, a VR experience. This study assessed the feasibility of the *Doc*VR experience for preoperative anxiety and overall presurgical satisfaction. We hypothesized that *Doc*VR would reduce patient anxiety and overall distress, while improving patient, caregiver, and health care provider satisfaction in the ambulatory surgery center (ASC) or the main OR.

Methods

This study included data from 51 patients, ranging in age from 6 to 14 years old, collected from March 15, 2019 to April 12, 2019. Patients were accompanied by at least one caregiver in the room during the VR experience. Given the time-sensitive nature of the OR, patients were available to play *Doc*VR and to answer study questions for variable lengths of times. All activities for the current study were approved by the institutional review board.

Recruitment

Pediatric patients waiting for surgical procedures in either the CHLA ASC or the CHLA main OR were approached to participate in the feasibility pilot. Though most patients were approached in the waiting rooms, some were approached in the preoperative holding areas. The research team worked with CCLSs in the ASC and OR to screen for and approach eligible patients.

All patients between the ages of 6 and 18 years with normal vision and typical cognitive development were eligible to participate, irrespective of surgery type. Cognitive and visual impairments rendered patients ineligible as these would interfere with completing the survey or playing the *DocVR* game. Due to the varied consensus on the minimum age for VR headsets, 6 years old was the minimum age for participation.

Development of DocVR

Preparation

In 2016, Disney Junior and CHLA had a series of meetings to discuss creating a VR experience with Doc McStuffins for the purposes of entertainment and to comfort children in the hospital setting. After multiple brainstorming meetings, the Disney Junior creative staff (termed "cast members") iterated on the VR environment and organized a demonstration day at the hospital for CHLA stakeholders. Having received feedback regarding the user experience of the prototype and what children typically enjoy, the cast members finished the virtual environment, and the team planned for a launch date, including decorating the waiting rooms for the ASC and the main OR area. The Disney team decorated the waiting areas with specific Doc McStuffins decals, transforming the virtual experience into a "real" physical space.

Virtual Reality

Inside the VR experience, users choose to enter either a toy hospital (main experience; Figure 1) or a theater (supplementary experience). If the toy hospital is chosen, users help Doc McStuffins treat toy patients by completing a series of game-like VR tasks. If the theater is chosen instead, users can watch a selection of Doc McStuffins episodes and clips.

Figure 1. Doc McStuffins in front of the Toy Hospital along with Lambie Lamb and Stuffy.

Upon choosing the toy hospital experience, users are greeted by Dottie ("Doc") McStuffins, who instructs them to click on a set of doors to enter the virtual hospital (Figure 1). After entering the hospital, users arrive at the OR front desk, where Doc explains that users will assist as her "medical student for the day." Users are also introduced to Nurse Hallie Hippo and fellow medical student Lambie Lamb. Users then click on a book of patients to pick 1 of 5 toy characters to treat.

The 5 characters are a robot, Robot Ray; a blue dinosaur stuffed animal, Stuffy; a purple plastic shark, Mr. Chomp; a green toy monster, Globo; and a superhero action figure, Awesome Guy. Each character has a health issue that the user helps Doc treat (Figure 2). Robot Ray is sick with "Drainy Battery-itis," treatment of which requires users to open Robot Ray's back panel with a screwdriver, replace his batteries, and reclose the battery compartment door. Stuffy is afflicted with "Ripped Plush-Anemia," healing of which requires the sewing up of a rip in Stuffy's fur. Mr. Chomp suffers from "Stuck-Junk-itis," for which users can remove junk objects from Mr. Chomp's mouth. Globo has a diagnosis of "No-Glow-Atosis," which is treated by pointing a Sunlight Power-Upper at his many moving hands for recharging. Awesome Guy is aching from "Crackety-Crackatosis," recovery of which requires players to scrub his cracks clean and seal them with paste. Each character takes about 5 minutes to treat, amounting to a VR experience lasting up to 25-30 minutes.

Figure 2. A few examples of what a user sees when playing DocVR Experience.

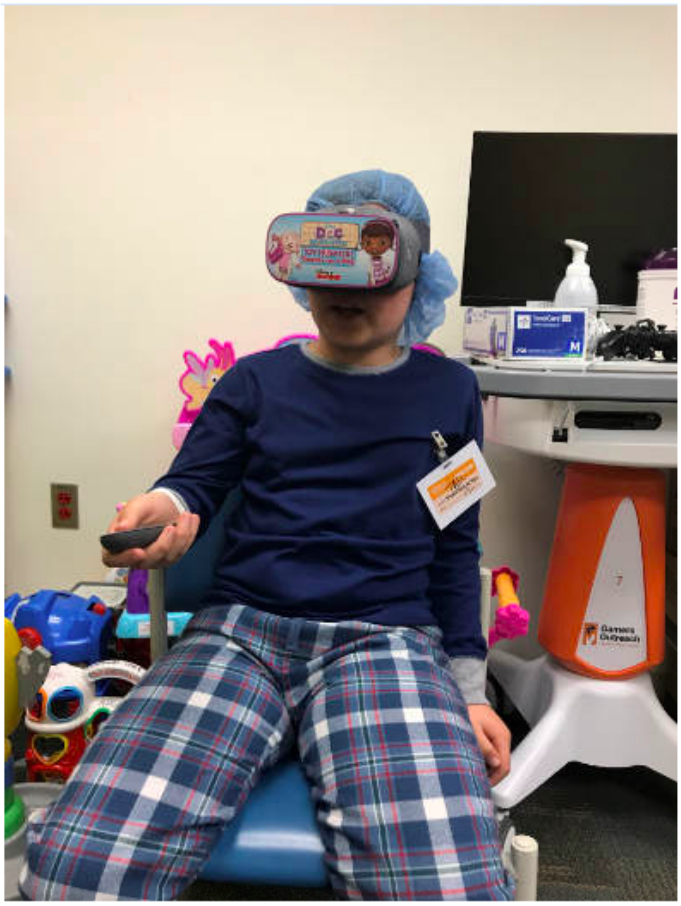
Besides playing engaging games to treat the toys' afflictions, users are also exposed to the medical day-to-day of check-ups and diagnoses. For example, they use blood pressure cuffs to record the toys' energy levels, stethoscopes to listen to heartbeats, an x-ray machine to capture the toys' internal features, and magnifying glasses to examine cuts, cracks, and rips. Furthermore, Doc puts some of the toys to sleep before treating them, paralleling the experience of general anesthesia in a real OR.

Study Procedures

In collaboration with CCLSs, patients were identified and approached to pilot *Doc*VR. After confirming the patient's interest, a research assistant administered a pre-*Doc*VR web-developed survey on Qualtrics using an iPad. Questions in the survey assessed the patient's current levels of anxiety (Visual Analogue Scale [VAS] for Child Anticipatory Anxiety/Procedural Anxiety), current overall mood (Facial Affective Scale [FAS]), thoughts and feelings about the upcoming surgery, and familiarity with both the Doc McStuffins character and VR itself.

After the patient completed the pre-*Doc*VR experience survey, the research team launched the *Doc*VR application on a Google Pixel 2 phone and inserted it into the Google Daydream View VR, before fitting the headset on the patient (Figure 3). Using the handheld controllers, the patient was then instructed to click through 2 floating bandage icons to begin the *Doc*VR experience.

Figure 3. A Children's Hospital Los Angeles patient playing the DocVR Experience.



Following gameplay, the research assistant administered the patient's post-*Doc*VR experience survey. The post-*Doc*VR survey, similar to the pre-*Doc*VR survey, evaluated the patient's current/post-VR anxiety, satisfaction, and affect/emotions; issues pertaining to immersion in the *Doc*VR experience; issues related to overall satisfaction; and an opportunity for general comments and feedback about the *Doc*VR experience.

At the conclusion of the VR experience, the patient was given a Doc McStuffins–themed badge lanyard inscribed with Dr. (child's name), sticker sheets, and toys to celebrate completing the *Doc*VR experience.

To maintain good hygiene and eliminate infection disease concerns, each patient wore a disposable felt face mask under the VR headset and a hair net. Additionally, research staff sanitized the controller and headset between each patient with Sani-Cloth Germicidal Disposable Wipes.

Data Collection and Measures

VAS for Child Anticipatory Anxiety/Procedural Anxiety

The VAS anticipatory anxiety measure is a vertical VAS, anchored with 0 at the bottom indicating the least amount and 10 at the top indicating the greatest amount, in response to the instruction to rate "how nervous, afraid, or worried" they were about the upcoming medical procedure or surgery. This continuous measure also has color cues, graded from yellow at the bottom to dark red at the top, as well as a neutral face at the bottom and a face showing a negative expression at the top. Prior research has used the VAS to rate anticipatory anxiety and pain in children [19,31,32].

Facial Affective Scale (FAS)

The FAS is a cartoon face scale with 9 faces ranging from smiling widely (Face 1: least distressed) to crying (Face 9: most distressed) [33]. The scale measures both pain intensity and emotional affect, giving insight to the child's overall level of discomfort, both physical and emotional. The scale can be used objectively by caregivers and health care providers, or the child can point to the face he feels represents how he feels at a given point in time.

Additional Questions: VR Experience Survey

A Disney Junior/CHLA-developed survey was created to assess various aspects of the patient's experience with *Doc*VR. The questions assessed overall enjoyment of the *Doc*VR game and thoughts and feelings that occurred while playing the game. The survey also included questions that asked patients about game-specific elements, such as which characters they helped, which characters were their favorite, what they liked most and least about the game, and whether or not they found the game difficult. Patients were asked various questions about how familiar they were with both VR and Doc McStuffins before playing the game and how often they thought about their surgeries before and during the VR experience. Additionally, the post-VR survey included a 14-item questionnaire that assessed the degree of immersion in the VR experience. All self-reported answers on the immersion questionnaire were

scored on a 3-point Likert Scale reflecting "A Lot," "A Little," or "No/Not at All." Finally, a research assistant also noted the patient's gender, age, and number of minutes spent playing the *Doc*VR experience.

Statistical Analysis

Descriptive statistics were used to summarize quantitative data from patient surveys. A nonparametric Wilcoxon signed-rank test was conducted with SPSS version 26 to calculate the change in the pre- and postanxiety scores (VAS for anticipatory anxiety), since this was a continuous variable with nonnormally distributed data.

Results

Pre-DocVR

Though this study included 51 patients, due to the OR schedule and based on which questions patients wanted to answer, the response rates for each question varied and are noted as such.

Of the 51 patients, 76% (39/51) had heard of Disney Junior, 90% (46/51) had heard of Doc McStuffins, and 32 had watched Doc McStuffins on television. Regarding VR play, 71% (36/51) of patients had never played VR, and 29% (15/51) of patients had played VR in the past. Those patients had used VR an average of 2.4 times, but 4 patients had used it 4 or more times. Overall, the patients were highly familiar with Disney Junior and Doc McStuffins, but few had actually used VR in the past.

Time Spent Thinking About Their Surgeries

Before playing the VR experience (n=50), based on the 5-level categorical ranking of time spent that morning thinking about "today's" surgery, 86% (43/50) of patients reported thinking about their surgery "Sometimes," "Often," or "Almost Always," with just 14% (7/50) reporting "Never" or "Almost Never" (Figure 4). While playing *Doc*VR (n=38), 52% (20/38) of patients reported thinking about the surgery "Sometimes," "Often," or "Almost Always" while playing the game, while 48% (18/38) of patients reported that they "Never" or "Almost Never" thought about their surgery during gameplay, demonstrating a 34% decrease in "time spent thinking about [their] surgery" between the pre-*Doc*VR (n=50) and post-*Doc*VR (n=38) groups of patients (Figure 4). When patients were asked what they were thinking about during *Doc*VR, they reported a variety of statements ranging from "fun" to "nothing" (Table 1).

Figure 4. Self-reported time spent thinking about their surgery before and during the DocVR experience.

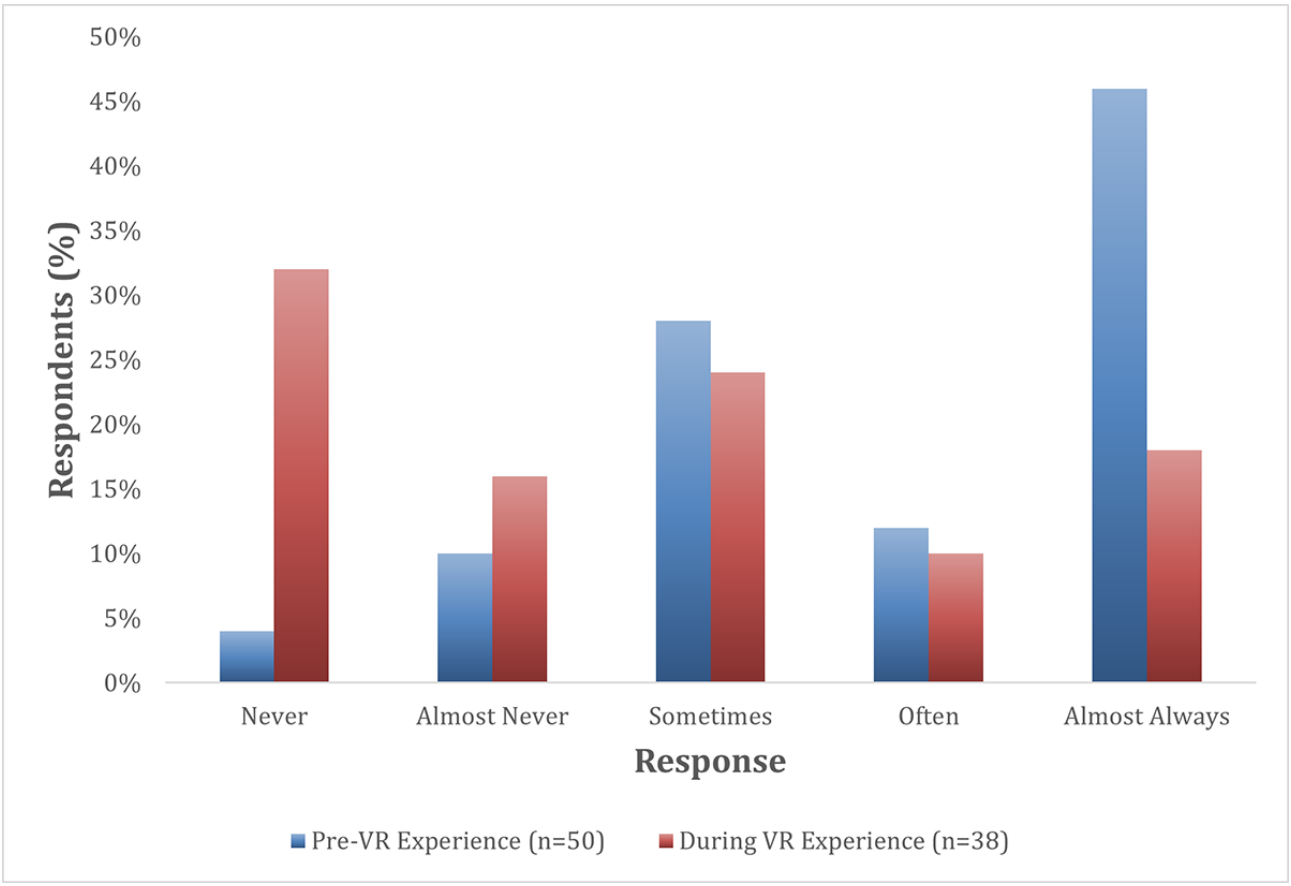


Table 1. Patients' thoughts during gameplay with DocVR, in response to the question "What did you think about during 'Doc McStuffins: Doctor for a Day?'"

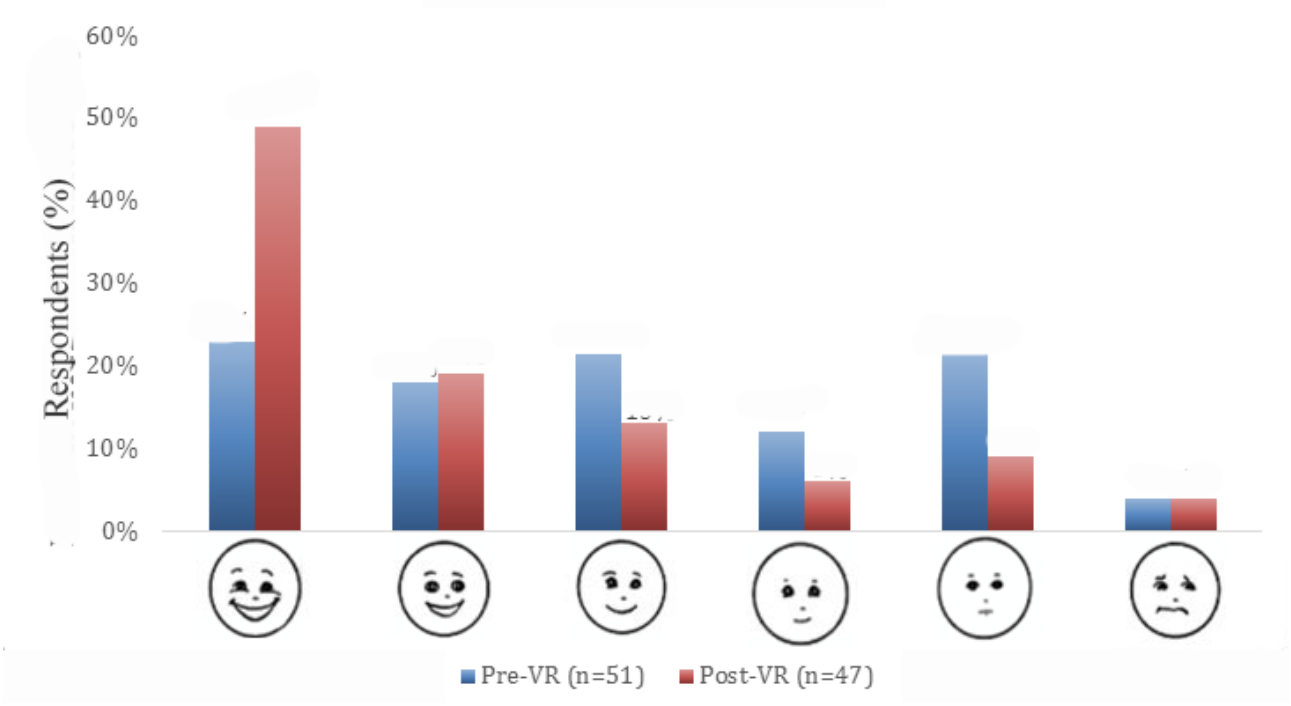
Response	Number of patients (n=24)
"It was fun"	10
"I felt good"	4
"Playing the game"	2
"Nothing"	1
"Excited"	1
"It was cool"	1
"It was hard"	1
"Surgery"	1
"I felt like a real doctor"	1
"That Doc was helping a lot"	1
"Nervous to help patients"	1

Pre- and Post-VR Affect and Anxiety

The median level of anxiety (n=51) reported by patients prior to *Doc*VR gameplay was 4.0 (IQR 0.60-7.43), while the median level of anxiety following gameplay (n=46) was 0.91 (IQR 0.30-4.60). Patients displayed a statistically significant decrease in anxiety after *DocVR* gameplay (Z=–3.26, *P*=.001).

Patients reported increases in positive affect from 23% (12/51) to 49% (23/47) following *Doc*VR gameplay (Figure 5). Similarly, a breakdown of the 3 most positive faces chosen versus the 3 most negative faces chosen shifted from 63% (32/51) and 38% (19/51) to 81% (38/47) and 19% (9/47), respectively, after playing the VR experience (Figure 5). Of the 9 faces in the FAS, no participants reported Faces 7-9 (faces that represent the 3 most distressed facial affects) pre- and post-*Doc*VR.

Figure 5. Self-reported affect pre- and post-DocVR gameplay experience. Faces 7-9 (faces that represent the 3 most distressed affects) were not selected by any participants pre- or post-DocVR and are thus not represented in the graph.



After *Doc*VR gameplay (n=39), 97% (38/39) of patients reported that playing the game made them feel more comfortable at the hospital: 51% (20/39) reported feeling "A Lot" more comfortable, and 46% (18/39) reported feeling "A Little" more comfortable. In addition, 74% (28/38) of patients reported that playing the game made them feel less scared at the hospital, with 29% (15/38) reporting "A Lot" less scared and 45% (17/38) reporting "A Little" less scared. When patients were asked about their feelings during *Doc*VR, they reported a variety of statements ranging from "a little better" to "good" (Table 2).

Table 2. Patients' feelings during gameplay with DocVR, in response to the question "How did you feel while playing 'Doc McStuffins: Doctor for a Day?'"

Response	Number of patients (n=23)
"Good"	6
"Happy"	5
"Fun and excited"	3
"Like a real doctor"	2
"Nervous"	2
"Cool"	2
"Like I was helping"	1
"Shy"	1
"A little better, because they had to go into a little surgery too"	1

Self-Reported Satisfaction With DocVR

On a 3-point anchor scale with responses of "No," "A Little," and "A Lot," 94% (46/49) of patients enjoyed playing the game, with 76% (37/49) reporting that they enjoyed the game "A Lot," and 18% (9/49) reporting "A Little." Of the 29% (15/51) of all patients who had played VR before, 100% (15/15) enjoyed *Doc*VR. Additionally, 88% (30/34) of patients reported feeling both "Interested" and "Involved" in the game, while 56% (19/34) reported their levels of interest and involvement as "A Lot," and 32% (11/34) reported it as "A Little." Of the patients, 97% (34/35) reported that playing the game "grabbed [their] attention"; 51% (18/35) reported it did so "A Lot," and 46% (16/35) reported it did so "A Little."

Of the patients, 88% (28/32) reported that *Doc*VR was also interesting compared to other computer games they had played; 41% (13/32) reported the game as being "A Lot" more interesting, and 47% (15/32) reported it as being "A Little" more interesting.

Of the patients, 97% (33/34) reported that they felt like they were "really" there in the VR world; 68% (23/34) reported feeling "A Lot" like they were there, and 29% (10/34) reported feeling "A Little" like they were there.

Of the patients, 53% (18/34) reported feeling sad or disappointed when the game ended; 18% (6/34) reported the level of their feelings of sadness or disappointment as being "A Lot," and 35% (12/34) reported the level of these feelings as "A Little."

No children reported known side effects to VR like dizziness, nausea, or light-headedness.

When patients were asked about they liked most about *Doc*VR, they reported a variety of statements ranging from "airplanes" to "helping people/patients/characters" (Table 3).

Table 3. Self-reported comments describing what the patients liked the most about the DocVR experience, in response to the question "What did you like MOST about 'Doc McStuffins: Doctor for a Day?'"

Response	Number of patients (n=28)
"Helping patients/people/characters"	12
"Helping Mr. Chomp"	3
"Everything"	2
"Fixing Stuffy"	2
"Playing"	2
"Helping Robot Ray with the batteries"	2
"Liked the patient going to sleep so it didn't feel any pain"	1
"It feels like I was really there"	1
"Airplanes"	1
"The surgery"	1
"The whole 3D environment"	1

"Doctor for a Day" Implementation Feasibility

Of the patients, 88% (29/33) reported that they would like to play the game again, with 61% (20/33) reporting "A Lot" and 27% (9/33) reporting "A Little." Of the patients, 97% (32/33) reported that they would recommend the game to another patient in the hospital or friend, with 70% (23/33) reporting "A Lot" and 27% (9/33) reporting "A Little." Responses to other feasibility questions are shown in Table 4.

Table 4. Patients' responses to a variety of virtual reality feasibility questions.

Question	Response, n (%)		
	A Lot	A Little	No
Did you get used to the game quickly? (n=38)	27 (71)	5 (13)	6 (16)
Were the controls easy to use? (n=38)	22 (58)	7 (18)	9 (24)
Did the things you saw look real? (n=33)	18 (55)	10 (30)	5 (15)
Was the headset comfortable? (n=34)	15 (44)	15 (44)	4 (12)
Were you worried about putting on the headset? (n=34)	4 (12)	3 (9)	27 (79)
Did it feel like you were in control? (n=34)	18 (53)	10 (29)	6 (18)
Did the way things moved look real? (n=33)	17 (52)	11 (33)	5 (15)

Discussion

The primary aim of this study was to assess the usability and feasibility of *Doc*VR, a novel, interactive, fully immersive 3D experience developed by Disney Junior and implemented by CHLA.

Key Findings

Findings suggest that the *Doc*VR is feasible, effective, and enjoyable and shows promise as a tool to alleviate pediatric preoperative anxiety and improve overall patient, caregiver, and health care provider experience during the preoperative period.

*Doc*VR significantly decreased patients' anxiety, leading them to think about things other than their apprehension about their procedures. Additionally, the virtual hospital setting of the VR experience made patients more comfortable in and less scared of the real medical space around them. Patients also responded positively to the DocVR content, reporting that they enjoyed the experience and found the game interesting. The game was very immersive, with a vast majority of patients reporting that they felt like they were "really there." The hardware itself was deemed both comfortable and user-friendly.

The authors found only 3 other studies that used VR as an exposure tool to alleviate pediatric preoperative anxiety. However, the VR experience for all 3 of these studies was an immersive guided tour of the operating theater [25,26,34]. While 1 of these 3 VR tours was conducted by another familiar childhood cartoon [26], the interactive nature of the *Doc*VR game allows patients to take an active role in medical play and provides an important addition to VR's cognitive load that ultimately is the key to alleviating cognitive states like pain and anxiety [35,36].

Patients' reports confirmed the excellent quality of the game. Many of the patients agreed that the virtual environment looked "real," and a large majority reported feeling like they were really in the virtual world. Patients were both interested and engaged, and a majority of patients said the game was also interesting compared to other video games they had played.

Effectiveness of DocVR

Patients experienced a strong statistically significant decrease in anxiety following *Doc*VR gameplay. After playing the *Doc*VR experience, patients also reported more positive overall affect, less fear, and less time spent thinking about their upcoming procedure. The VR experience also made patients more comfortable in the hospital, with 1 patient, when asked his favorite part of the game, saying, "I liked the patient going to sleep, so [the patient] didn't feel any pain." This response demonstrates the power of familiarizing pediatric patients with their upcoming procedures, putting them at ease for when the time for the procedure comes. *Doc*VR also made patients feel "Good" and "Happy."

While the hardware required for VR, including headsets and controllers, is rapidly deemed out of date due to the rapid expansion of VR innovation and development, the software nature of *Doc*VR makes it an effective product that can stay relevant as the VR world surges forward.

VR Age Restrictions

There is still some debate around the appropriate age for the use of VR in pediatrics. Most VR headset manufacturers (Sony, Google, Samsung, HTC, Oculus) recommend that their products should not be used by individuals younger than 12-14 years of age. The Google Daydream, the headset used in this study, "should not be used by children under the age of 13," according to the manufacturer's website [37]. The concerns for pediatric use can generally be thought of as being related to (1) legal and liability concerns and (2) safety concerns. The 1998 Children's Online Privacy Protection Act (COPPA) governs the collection and use of data generated by children under the age of 13 years by websites, mobile applications, and smart devices [38]. COPPA created rules for privacy policies, data collection, parental consent, and parent access to and control over a child's data. Importantly, COPPA does not prohibit the collection of data but, rather, creates a regulatory framework for how the data should be collected [39]. Since many VR headsets include creating user accounts, the COPPA rules apply to them. In our study, we did not have study patients create user accounts. From a regulatory standpoint, although a handful of VR applications have been approved by the US Food and Drug Administration, the regulatory environment is still evolving. In 2017, the Food and Drug Administration released its Digital Health Innovation Action Plan [40] and, in 2020, held a public workshop to convene industry, investigators, providers, and regulators to discuss best evaluation practices for medical extended reality [41]. When VR is used for distraction or entertainment, such as in this study, it is not considered a medical device [42].

Safety concerns for pediatric VR have focused on the physical fit of the device (headsets may be too big for some young children and, if not worn correctly, can cause physical discomfort), safety around device use (such as accidentally bumping into objects in the real world), and health impacts of VR. Some users experience dizziness, headaches, and motion sickness during or after VR, but over 15 years of pediatric studies have demonstrated the overall safety and minimal side effects of VR in children as young as 6 years [21,22,43-46]. In our study, which included 42 patients under the age of 12 years, no patients reported any symptoms after using the VR headset. While there have been concerns about the possible impact of VR on vision, the American Academy of Ophthalmology states that, although there have been no long-term studies, there is little reason to be worried about VR's impact on eye development or function [47]. A 2019 study of 50 children aged 4-10 years showed no significant impact on visuomotor function after VR use [48]. Finally, from a parenting and development standpoint, the American Academy of Pediatrics has stated that the 2016 media use guidelines [49] apply to a variety of media, including VR [50]. Parents should make sure content is developmentally appropriate for their children; that they adhere to overall screen time recommendations; to balance media use with media-free time for physical activity, education, play, sleep, and bonding; and whenever possible, to coview or coexperience media together. Various methods of delivering VR have been implemented to validate these data and to ensure that VR is a valid, ethical, and safe option for all pediatric patients [51].

Limitations

One limitation of this study is the small sample size that, though not uncommon in feasibility research, may limit the generalizability of the conclusions.

This pilot did not utilize a control group; all patients were given the option to play VR. Therefore, researchers were unable to compare the effects of playing *Doc*VR with the anxiety of children who did not play the game. Additionally, the observed positive impact of the *Doc*VR experience on the patients could have been confounded by the positive impact of research assistants' general active engagement with patients.

Another limitation to the pilot is that not all patients completed the entire *Doc*VR experience. There were 2 scenarios that would have prematurely terminated the VR experience: (1) the child deciding to take the headset off or (2) the child being called to the preoperative area by a health care provider.

A premature end to the VR experience resulted in incomplete playing time and therefore, an incomplete assessment of *Doc*VR for managing pain and anxiety (dose effects). The average time spent playing *Doc*VR may have been higher if the pilot had been conducted in a more controlled environment. However, because the pilot was conducted in a dynamic setting, parts of the protocol were adapted to accommodate clinical care.

In the scenario where children were quickly called back to the preoperative area by the health care provider, there was insufficient time for the post-VR survey to be administered. The results reported in this paper are thus impacted by incomplete data collection.

Other potential confounds include that 90% (46/51) of patients were familiar with Doc McStuffins, and 29% (15/51) of patients had played VR before. This prior exposure may have primed patients to benefit from a pre-existing relationship or experience with Doc, enjoying the *Doc*VR experience, or other content and technology confounds. Nonetheless, this patient population could be representative of a young, TV-watching population, indicating that familiarity may, in fact, be a beneficial component of relieving pediatric preoperative anxiety.

Conclusions

This is the first feasibility study on the use of *Doc*VR to ease preoperative anxiety in pediatric patients and one of a few studies ever to use VR to address preoperative anxiety in pediatric patients prior to surgery. These results demonstrate the potential utility of VR and, particularly, *Doc*VR in the preoperative space. The present study capitalized on the fusion of familiar and lovable characters with immersive VR gameplay to transform the patient preoperative experience from unpleasant and potentially frightening to fun. Given that health care institutions continue to use patient and family satisfaction as a metric of success [52], digital therapeutic solutions that patients find enjoyable and distracting prove to be worthwhile investments. Furthermore, developers in pediatric health care would benefit from collaborating with children's media companies in order to capitalize upon pre-established characters and the emotional experience or relationship between children and best-loved characters. Continued research on emerging technologies and VR experiences is essential to ensure that science and an evidence base drive clinical interventions for pain and stress management in a pediatric environment.

Acknowledgments

The authors are grateful for all the patients and families who agreed to participate in the study and would also like to acknowledge the leadership at both Disney Junior and CHLA who made this collaboration possible. The authors would also like to thank the generous donors who continue to support VR research at the CHLA Biobehavioral Pain Laboratory, including Beatrice and Paul Bennett and Cancer Free Generation and the Board of Tower Cancer Research Foundation. A special thank you must go to the CHLA Child Life specialists and the CHLA Biobehavioral Pain Laboratory volunteers.

Conflicts of Interest

None declared.

References

1. Perry JN, Hooper VD, Masiongale J. Reduction of preoperative anxiety in pediatric surgery patients using age-appropriate teaching interventions. J Perianesth Nurs 2012 Apr;27(2):69-81. [doi: 10.1016/j.jopan.2012.01.003] [Medline: 22443919]
2. Kain ZN, Mayes LC, Caldwell-Andrews AA, Karas DE, McClain BC. Preoperative anxiety, postoperative pain, and behavioral recovery in young children undergoing surgery. Pediatrics 2006 Aug;118(2):651-658. [doi: 10.1542/peds.2005-2920] [Medline: 16882820]
3. Kain ZN, Mayes LC, O'Connor TZ, Cicchetti DV. Preoperative anxiety in children. Predictors and outcomes. Arch Pediatr Adolesc Med 1996 Dec;150(12):1238-1245. [doi: 10.1001/archpedi.1996.02170370016002] [Medline: 8953995]
4. Schmoldt A, Benthe HF, Haberland G. Digitoxin metabolism by rat liver microsomes. Biochem Pharmacol 1975 Sep 01;24(17):1639-1641. [Medline: 10]
5. Kain ZN, Caldwell-Andrews AA, Maranets I, McClain B, Gaal D, Mayes LC, et al. Preoperative anxiety and emergence delirium and postoperative maladaptive behaviors. Anesth Analg 2004 Dec;99(6):1648-1654. [doi: 10.1213/01.ANE.0000136471.36680.97] [Medline: 15562048]
6. Kain ZN, Caldwell-Andrews AA, Mayes LC, Weinberg ME, Wang S, MacLaren JE, et al. Family-centered preparation for surgery improves perioperative outcomes in children: a randomized controlled trial. Anesthesiology 2007 Jan;106(1):65-74 [FREE Full text] [doi: 10.1097/00000542-200701000-00013] [Medline: 17197846]
7. Bevan JC, Veall GR, Macnab AJ, Ries CR, Marsland C. Midazolam premedication delays recovery after propofol without modifying involuntary movements. Anesth Analg 1997 Jul;85(1):50-54. [doi: 10.1097/00000539-199707000-00009] [Medline: 9212121]
8. Viitanen H, Annila P, Viitanen M, Tarkkila P. Premedication with midazolam delays recovery after ambulatory sevoflurane anesthesia in children. Anesth Analg 1999 Jul;89(1):75-79. [doi: 10.1097/00000539-199907000-00014] [Medline: 10389782]
9. McCann ME, Kain ZN. The management of preoperative anxiety in children: an update. Anesth Analg 2001 Jul;93(1):98-105. [doi: 10.1097/00000539-200107000-00022] [Medline: 11429348]
10. Wright KD, Stewart SH, Finley GA, Buffett-Jerrott SE. Prevention and intervention strategies to alleviate preoperative anxiety in children: a critical review. Behav Modif 2007 Jan;31(1):52-79. [doi: 10.1177/0145445506295055] [Medline: 17179531]
11. Boles J, Fraser C, Bennett K, Jones M, Dunbar J, Woodburn A, et al. The Value of Certified Child Life Specialists: Direct and Downstream Optimization of Pediatric Patient and Family Outcomes. Association of Child Life Professionals. 2020 Jan. URL: https://www.childlife.org/docs/default-source/the-child-life-profession/value-of-cclss-full-report.pdf [accessed 2021-03-26]
12. Chan JM. Preparation for procedures and surgery through play. Paediatrician 1980;9(3-4):210-219. [Medline: 7208072]
13. American Academy of Pediatrics Child Life Council and Committee on Hospital Care, Wilson JM. Child life services. Pediatrics 2006 Oct;118(4):1757-1763. [doi: 10.1542/peds.2006-1941] [Medline: 17015572]
14. Kaminski M, Pellino T, Wish J. Play and Pets: The Physical and Emotional Impact of Child-Life and Pet Therapy on Hospitalized Children. Children's Health Care 2002 Dec;31(4):321-335 [FREE Full text] [doi: 10.1207/s15326888chc3104_5]
15. Beickert K, Mora K. Transforming the Pediatric Experience: The Story of Child Life. Pediatr Ann 2017 Sep 01;46(9):e345-e351. [doi: 10.3928/19382359-20170810-01] [Medline: 28892551]
16. Brewer S, Gleditsch SL, Syblik D, Tietjens ME, Vacik HW. Pediatric anxiety: child life intervention in day surgery. J Pediatr Nurs 2006 Feb;21(1):13-22. [doi: 10.1016/j.pedn.2005.06.004] [Medline: 16428010]
17. Schwartz BH, Albino JE, Tedesco LA. Effects of psychological preparation on children hospitalized for dental operations. J Pediatr 1983 Apr;102(4):634-638. [doi: 10.1016/s0022-3476(83)80211-x] [Medline: 6834205]
18. Ellerton ML, Merriam C. Preparing children and families psychologically for day surgery: an evaluation. J Adv Nurs 1994 Jun;19(6):1057-1062. [doi: 10.1111/j.1365-2648.1994.tb01188.x] [Medline: 7930085]
19. Kain ZN, Mayes LC, Caramico LA. Preoperative preparation in children: a cross-sectional study. J Clin Anesth 1996 Sep;8(6):508-514. [doi: 10.1016/0952-8180(96)00115-8] [Medline: 8872693]
20. Lynch M. Preparing children for day surgery. Child Health Care 1994;23(2):75-85. [doi: 10.1207/s15326888chc2302_1] [Medline: 10171871]
21. Gold JI, Mahrer NE. Is Virtual Reality Ready for Prime Time in the Medical Space? A Randomized Control Trial of Pediatric Virtual Reality for Acute Procedural Pain Management. J Pediatr Psychol 2018 Apr 01;43(3):266-275. [doi: 10.1093/jpepsy/jsx129] [Medline: 29053848]
22. Gold JI, Kim SH, Kant AJ, Joseph MH, Rizzo AS. Effectiveness of virtual reality for pediatric pain distraction during i.v. placement. Cyberpsychol Behav 2006 Apr;9(2):207-212. [doi: 10.1089/cpb.2006.9.207] [Medline: 16640481]
23. Shetty V, Suresh L, Hegde A. Effect of Virtual Reality Distraction on Pain and Anxiety During Dental Treatment in 5 to 8 Year Old Children. J Clin Pediatr Dent 2019;43(2):97-102. [doi: 10.17796/1053-4625-43.2.5] [Medline: 30730798]
24. Hoffman HG, Rodriguez RA, Gonzalez M, Bernardy M, Peña R, Beck W, et al. Immersive Virtual Reality as an Adjunctive Non-opioid Analgesic for Pre-dominantly Latin American Children With Large Severe Burn Wounds During Burn Wound Cleaning in the Intensive Care Unit: A Pilot Study. Front Hum Neurosci 2019;13:262 [FREE Full text] [doi: 10.3389/fnhum.2019.00262] [Medline: 31440148]

25. Bekelis K, Calnan D, Simmons N, MacKenzie TA, Kakoulides G. Effect of an Immersive Preoperative Virtual Reality Experience on Patient Reported Outcomes: A Randomized Controlled Trial. Ann Surg 2017 Jun;265(6):1068-1073. [doi: 10.1097/SLA.0000000000002094] [Medline: 27906757]
26. Ryu J, Park S, Park J, Kim J, Yoo H, Kim T, et al. Randomized clinical trial of immersive virtual reality tour of the operating theatre in children before anaesthesia. Br J Surg 2017 Nov 04;104(12):1628-1633. [doi: 10.1002/bjs.10684] [Medline: 28975600]
27. Eijlers R, Utens EMWJ, Staals LM, de Nijs PFA, Berghmans JM, Wijnen RMH, et al. Systematic Review and Meta-analysis of Virtual Reality in Pediatrics. Anesthesia & Analgesia 2019;129(5):1344-1353. [doi: 10.1213/ane.0000000000004165]
28. de Leeuw RNH, van der Laan CA. Helping behavior in Disney animated movies and children's helping behavior in the Netherlands. Journal of Children and Media 2017 Dec 05;12(2):159-174 [FREE Full text] [doi: 10.1080/17482798.2017.1409245]
29. Padilla-Walker LM, Coyne S, Fraser A, Stockdale L. Is Disney the Nicest Place on Earth? A Content Analysis of Prosocial Behavior in Animated Disney Films. J Commun 2013 Mar 11;63(2):393-412 [FREE Full text] [doi: 10.1111/jcom.12022]
30. Ward AR. The Lion King's Mythic Narrative: Disney as Moral Educator. Journal of Popular Film and Television 1996 Jan;23(4):171-178 [FREE Full text] [doi: 10.1080/01956051.1996.9943703]
31. Bringuier S, Dadure C, Raux O, Dubois A, Picot M, Capdevila X. The perioperative validity of the visual analog anxiety scale in children: a discriminant and useful instrument in routine clinical practice to optimize postoperative pain management. Anesth Analg 2009 Sep;109(3):737-744. [doi: 10.1213/ane.0b013e3181af00e4] [Medline: 19690240]
32. Tsao JCI, Myers CD, Craske MG, Bursch B, Kim SC, Zeltzer LK. Role of anticipatory anxiety and anxiety sensitivity in children's and adolescents' laboratory pain responses. J Pediatr Psychol 2004;29(5):379-388 [FREE Full text] [doi: 10.1093/jpepsy/jsh041] [Medline: 15187176]
33. McGrath PA, Seifert CE, Speechley KN, Booth JC, Stitt L, Gibson MC. A new analogue scale for assessing children's pain: an initial validation study. Pain 1996 Mar;64(3):435-443. [doi: 10.1016/0304-3959(95)00171-9] [Medline: 8783307]
34. Eijlers R, Legerstee JS, Dierckx B, Staals LM, Berghmans J, van der Schroeff MP, et al. Development of a Virtual Reality Exposure Tool as Psychological Preparation for Elective Pediatric Day Care Surgery: Methodological Approach for a Randomized Controlled Trial. JMIR Res Protoc 2017 Sep 11;6(9):e174 [FREE Full text] [doi: 10.2196/resprot.7617] [Medline: 28893727]
35. Yuan JC, Rodriguez S, Caruso TJ, Tsui JH. Provider-controlled virtual reality experience may adjust for cognitive load during vascular access in pediatric patients. Can J Anaesth 2017 Dec 31;64(12):1275-1276. [doi: 10.1007/s12630-017-0962-5] [Medline: 28861855]
36. Petrovic P, Petersson K, Ghatan P, Stone-Elander S, Ingvar M. Pain-related cerebral activation is altered by a distracting cognitive task. Pain 2000 Mar;85(1-2):19-30. [doi: 10.1016/s0304-3959(99)00232-8] [Medline: 10692599]
37. Daydream Standalone Safety Information. Google Daydream Help. URL: https://support.google.com/daydream/answer/9009545?hl=en [accessed 2021-03-26]
38. Children's Online Privacy Protection Rule ("COPPA"). Federal Trade Commission. 2020 Mar 06. URL: https://www.ftc.gov/enforcement/rules/rulemaking-regulatory-reform-proceedings/childrens-online-privacy-protection-rule [accessed 2021-03-26]
39. Complying with COPPA: Frequently Asked Questions. Federal Trade Commission. 2020 Jul 22. URL: https://www.ftc.gov/tips-advice/business-center/guidance/complying-coppa-frequently-asked-questions-0 [accessed 2021-03-26]
40. Digital health innovation action plan. U.S. Food and Drug Administration. URL: https://www.fda.gov/media/106331/download [accessed 2021-03-26]
41. Public Workshop - Medical Extended Reality: Toward Best Evaluation Practices for Virtual and Augmented Reality in Medicine. US Food and Drug Administration. 2020 Mar 05. URL: https://www.fda.gov/medical-devices/workshops-conferences-medical-devices/public-workshop-medical-extended-reality-toward-best-evaluation-practices-virtual-and-augmented?utm_campaign=2020-01-16 [accessed 2021-03-26]
42. Software as a Medical Device (SaMD): Key Definitions. International Medical Device Regulators Forum. 2013 Dec 09. URL: http://www.imdrf.org/docs/imdrf/final/technical/imdrf-tech-131209-samd-key-definitions-140901.pdf [accessed 2021-03-26] [WebCite Cache ID 6T6V2i7Nq]
43. Parsons TD, Riva G, Parsons S, Mantovani F, Newbutt N, Lin L, et al. Virtual Reality in Pediatric Psychology. Pediatrics 2017 Nov;140(Suppl 2):S86-S91 [FREE Full text] [doi: 10.1542/peds.2016-1758I] [Medline: 29093039]
44. Caruso TJ, O'Connell C, Qian JJ, Kung T, Wang E, Kinnebrew S, et al. Retrospective Review of the Safety and Efficacy of Virtual Reality in a Pediatric Hospital. Pediatr Qual Saf 2020;5(2):e293 [FREE Full text] [doi: 10.1097/pq9.0000000000000293] [Medline: 32426648]
45. Morris LD, Louw QA, Grimmer-Somers K. The effectiveness of virtual reality on reducing pain and anxiety in burn injury patients: a systematic review. Clin J Pain 2009;25(9):815-826. [doi: 10.1097/AJP.0b013e3181aaa909] [Medline: 19851164]
46. Rodriguez S, Caruso T, Tsui B. Bedside Entertainment and Relaxation Theater: size and novelty does matter when using video distraction for perioperative pediatric anxiety. Paediatr Anaesth 2017 Jun 05;27(6):668-669 [FREE Full text] [doi: 10.1111/pan.13133] [Medline: 28474813]

47. Mukamal R, Stephen L. Are Virtual Reality Headsets Safe for Eyes? American Academy of Ophthalmology. 2017 Feb 28. URL: https://www.aao.org/eye-health/tips-prevention/are-virtual-reality-headsets-safe-eyes [accessed 2021-03-26]
48. Tychsen L, Foeller P. Effects of Immersive Virtual Reality Headset Viewing on Young Children: Visuomotor Function, Postural Stability, and Motion Sickness. Am J Ophthalmol 2020 Jan;209:151-159. [doi: 10.1016/j.ajo.2019.07.020] [Medline: 31377280]
49. Council on Communications and Media. Media Use in School-Aged Children and Adolescents. Pediatrics 2016 Nov;138(5):1 [FREE Full text] [doi: 10.1542/peds.2016-2592] [Medline: 27940794]
50. AAP Announces New Recommendations for Children's Media Use. American Academy of Pediatrics. 2016 Oct 21. URL: https://www.healthychildren.org/English/news/Pages/AAP-Announces-New-Recommendations-for-Childrens-Media-Use.aspx [accessed 2021-03-26]
51. Newbutt N, Bradley R, Conley I. Using Virtual Reality Head-Mounted Displays in Schools with Autistic Children: Views, Experiences, and Future Directions. Cyberpsychol Behav Soc Netw 2020 Jan;23(1):23-33. [doi: 10.1089/cyber.2019.0206] [Medline: 31502866]
52. Shirley ED, Sanders JO. Patient satisfaction: Implications and predictors of success. J Bone Joint Surg Am 2013 May 15;95(10):e69. [doi: 10.2106/JBJS.L.01048] [Medline: 23677370]

Abbreviations

ASC: ambulatory surgery center
CCLS: Certified Child Life Specialist
CHLA: Children's Hospital Los Angeles
COPPA: 1998 Children's Online Privacy Protection Act
DocVR: "Doc McStuffins: Doctor for a Day" Virtual Reality Experience
FAS: Facial Affective Scale
OR: operating room
VAS: visual analogue scale
VR: virtual reality

Edited by R Kukafka; submitted 04.11.20; peer-reviewed by N Newbutt, S Rodriguez; comments to author 01.12.20; revised version received 06.01.21; accepted 16.03.21; published 19.04.21.

Please cite as:
Gold JI, Annick ET, Lane AS, Ho K, Marty RT, Espinoza JC
"Doc McStuffins: Doctor for a Day" Virtual Reality (DocVR) for Pediatric Preoperative Anxiety and Satisfaction: Pediatric Medical Technology Feasibility Study
J Med Internet Res 2021;23(4):e25504
URL: https://www.jmir.org/2021/4/e25504
doi:10.2196/25504
PMID:33730687

Original Paper

Socioeconomic Disparities in eHealth Literacy and Preventive Behaviors During the COVID-19 Pandemic in Hong Kong: Cross-sectional Study

Ziqiu Guo[1], MMed; Sheng Zhi Zhao[1], MPH; Ningyuan Guo[1], BSc; Yongda Wu[1], PhD; Xue Weng[1], PhD; Janet Yuen-Ha Wong[1], PhD; Tai Hing Lam[2], MD; Man Ping Wang[1], PhD

[1]School of Nursing, University of Hong Kong, Hong Kong, China (Hong Kong)

[2]School of Public Health, University of Hong Kong, Hong Kong, China (Hong Kong)

Corresponding Author:
Man Ping Wang, PhD
School of Nursing
University of Hong Kong
21 Sassoon Road
Hong Kong
China (Hong Kong)
Phone: 852 3917 6636
Fax: 852 2872 6079
Email: mpwang@hku.hk

Abstract

Background: eHealth literacy can potentially facilitate web-based information seeking and taking informed measures.

Objective: This study aimed to evaluate socioeconomic disparities in eHealth literacy and seeking of web-based information on COVID-19, and their associations with COVID-19 preventive behaviors.

Methods: The COVID-19 Health Information Survey (CoVHIns), using telephonic (n=500) and web-based surveys (n=1001), was conducted among adults in Hong Kong in April 2020. The Chinese eHealth literacy scale (eHEALS; score range 8-40) was used to measure eHealth literacy. COVID-19 preventive behaviors included wearing surgical masks, wearing fabric masks, washing hands, social distancing, and adding water or bleach to the household drainage system. Adjusted beta coefficients and the slope indices of inequality for the eHEALS score by socioeconomic status, adjusted odds ratios (aORs) for seeking of web-based information on COVID-19 by socioeconomic status, and aORs for the high adherence to preventive behaviors by the eHEALS score and seeking of web-based information on COVID-19 were calculated.

Results: The mean eHEALS score was 26.10 (SD 7.70). Age was inversely associated with the eHEALS score, but education and personal income were positively associated with the eHEALS score and seeking of web-based information on COVID-19 (for all, *P* for trend<.05). Participants who sought web-based information on COVID-19 showed high adherence to the practice of wearing surgical masks (aOR 1.56, 95% CI 1.15-2.13), washing hands (aOR 1.33, 95% CI 1.05-1.71), social distancing (aOR 1.48, 95% CI 1.14-1.93), and adding water or bleach to the household drainage system (aOR 1.67, 95% CI 1.28-2.18). Those with the highest eHEALS score displayed high adherence to the practice of wearing surgical masks (aOR 3.84, 95% CI 1.63-9.05), washing hands (aOR 4.14, 95% CI 2.46-6.96), social distancing (aOR 2.25, 95% CI 1.39-3.65), and adding water or bleach to the household drainage system (aOR 1.94, 95% CI 1.19-3.16), compared to those with the lowest eHEALS score.

Conclusions: Chinese adults with a higher socioeconomic status had higher eHealth literacy and sought more web-based information on COVID-19; both these factors were associated with a high adherence to the guidelines for preventive behaviors during the COVID-19 pandemic.

(J Med Internet Res 2021;23(4):e24577) doi:10.2196/24577

KEYWORDS

COVID-19; eHealth literacy; preventive behaviors; socioeconomic disparities; web-based information seeking

Introduction

Curbing of the spread of COVID-19 depends on the timely adoption of appropriate preventive behaviors by the public. Web-based health information is important in affecting preventive behaviors, particularly when physical distancing and stay-at-home orders during the pandemic have reduced face-to-face health communication [1]. A recent study reported that seeking of web-based information on COVID-19 from social networking apps and internet-based news media are associated with preventive behaviors [2]. A breadth of information and misinformation has been disseminated on the internet and rapidly propagated and evolved on social media platforms [3]. Exposure to misinformation on the internet or conspiracy theories regarding COVID-19 are associated with decreased adherence to prevention guidelines and worse physical and mental health outcomes [4,5]. The ability to seek, understand, and appraise health information on the internet and ultimately take well-informed action to handle health problems can be assessed on the basis of eHealth literacy [6]. A higher eHealth literacy is associated with more active information seeking and scrutiny [7,8]. Lack of access or capacity to understand health information on the internet, in contrast, is associated with negligence toward health warnings and difficulty in making health decisions [9].

Appropriate processing and utilization of health information is complex during the COVID-19 pandemic, given the novel outbreak patterns and evolving information regarding the disease [10]. It is important to identify the characteristics of individuals at the risk of lower eHealth literacy for effective health promotion, including the provision of limited literacy resources [11]. Previous studies have suggested that eHealth literacy is affected by sociodemographic, environmental, and contextual factors [12]. Disparities in eHealth literacy by education and income have been previously reported [13], but incongruent correlations between socioeconomic status and eHealth literacy have been found across populations with different characteristics [8,14,15]. The COVID-19 pandemic has disproportionately affected the lower socioeconomic status (SES) group and has limited access to health care, overcrowded living conditions with a higher risk of disease transmission, and inconvenienced individuals who are in occupations that do not allow working from home [16], which have further accentuated existing socioeconomic inequalities. Since eHealth literacy skill is not static and evolves with changes in new social contexts [6], little is known about the disparities in eHealth literacy in the unique context of widening socioeconomic inequalities and the overwhelming influx of COVID-19–related information (and misinformation) being disseminated.

Hong Kong, the most developed and westernized city of China, has a larger income gap (Gini index 0.539 in 2016) compared to other developed countries [17], but internet use is prevalent across individuals of different SES because of the advanced cyber infrastructure and the low cost of internet access [18]. Nearly all individuals have sought web-based information during the COVID-19 pandemic [19]. Our previous study in 2009-2012 reported that disparities in SES groups affected web-based health information seeking behavior [20]. Considering that the context of COVID-19 may stimulate universal web-based information seeking behavior by triggering effective responses such as fear and anxiety [21], it remains unknown whether SES disparities in web-based health information seeking existed during the COVID-19 pandemic. Our study's research questions were as follows: (1) Are there socioeconomic disparities in seeking web-based information on COVID-19 during the pandemic? (2) Are there socioeconomic disparities in eHealth literacy among these web-based information seekers? (3) Is seeking of web-based information and eHealth literacy associated with preventive behaviors during the COVID-19 pandemic? In a random cohort of adults in Hong Kong, we examined socioeconomic disparities in seeking web-based information on COVID-19 and eHealth literacy, and their associations with personal preventive behaviors during the COVID-19 pandemic.

Methods

Design and Participants

This study was part of the COVID-19 Health Information Survey (CoVHIns)—which is a cross-sectional survey among adults in Hong Kong who are aged ≥18 years—which investigated COVID-19–related information use, preventive behaviors, and well-being. The survey was conducted on April 9-23, 2020, after the peak of the second wave of the outbreak, and when social distancing measures were implemented. Data were collected using telephonic and web-based surveys. All interviews were conducted by trained interviewers of Social Policy Research Limited through the Web-based Computer Assisted Telephone Interview system.

The details of CoVHIns have been reported previously [22,23]. Briefly, a 2-stage sampling method was adopted for the telephonic survey. First, telephone numbers were retrieved from the residential telephone directories and randomly listed for interview. Invalid numbers, lack of responses (after being called for a maximum of 5 times), and ineligible households (including individuals aged <18 years or those who are unable to communicate in Cantonese or Mandarin) were excluded. Second, once a household was successfully contacted, the eligible family member whose birth date was closest to the interview date was invited to complete the interview. Each interview lasted approximately 20 minutes. A total of 816 landline telephone numbers were successfully sampled, and 500 participants consented to and completed the interview (response rate=61.3%).

In addition, web-based surveys randomly sampled participants from a representative panel of >100,000 mobile phone users, which was generated by sending text messages to a random list of mobile phone numbers provided by the Numbering Plan for Telecommunication Services (prefixes 5, 6, 9). Stratified random sampling by sex and age was adopted. Text messages with an invitation were sent to the randomly selected members in the panel. Among 1623 eligible individuals contacted, 1001 participants consented to and completed the questionnaire on the internet (response rate=61.7%). Ethics approval was granted by the Institutional Review Board of the University of Hong Kong/Hospital Authority Hong Kong West Cluster (approval# UW-20-238).

Measurements

Seeking of web-based information on COVID-19 was self-reported (sought or no-sought). eHealth literacy was assessed among individuals who had sought web-based information on COVID-19, considering that eHealth literacy is based on the experience of access to web-based information [24]. We used the Chinese version of the eHealth literacy scale (eHEALS) to measure eHealth literacy levels by asking participants' about their past experience in seeking web-based information on COVID-19 (Multimedia Appendix 1). The eHEALS contains 8 items scored with a 5-point Likert scale (ranging from "strongly disagree" to "strongly agree"). The total score ranged from 8 to 40, with a higher score indicating higher eHealth literacy [25]. A Cronbach α of .95 was used in our study. Consistent with the Chinese version of eHEALS [25], we observed a unidimensional structure of the Chinese eHEALS with adequate model fitness (comparative fit index 0.974 [>0.95, acceptable], root mean square error of approximation 0.097 [close to 0.06, acceptable], and Tucker-Lewis index 0.964 [>0.95 acceptable]) [26]. We divided the eHEALS score into 4 categories (Q1-Q4) based on quartile values (median 28, IQR 22-32) in accordance with previous studies, using the median as the cutoff [7,27]. Specifically, Q1 was the interval of the overall eHEALS score ranging from the lowest value (ie, 8) to a score of 22; Q2, from 22 to 28; Q3, from 28 to 32; and Q4, from 32 to the highest value (ie, 40).

Based on World Health Organization guidelines for COVID-19 prevention [28], we assessed personal preventive behaviors in the past 7 days, including the following: "wearing surgical masks when going out," "wearing fabric masks when going out," "washing hands with alcohol-based sanitizer," "adding water/bleach to the household drainage system," and "keeping a social distance from people in public areas (eg, 1.5 meters)," with responses of "never," "occasionally," "sometimes," and "often" (Multimedia Appendix 2). Adherence to personal preventive behavior was dichotomized (low or high adherence) on the basis of previous studies on the association between eHealth literacy and health behaviors [14,24]. Responses of "never," "occasionally," and "sometimes" were together considered as low adherence and "often" was considered as high adherence.

Education levels and income were considered as indicators of SES. Education levels were measured as categorical variables ("primary or below," "secondary," and "tertiary or above") on the basis of the highest education level attained. We measured monthly personal income in accordance with 6 predefined categories (from ≤HK $10,000-$50,001 [US $1=HK $7.8]). Since few participants had an income of HK $40,001-$50,000 and ≥HK $50,001, the data were recoded in 4 categories including ≤HK $10,000, HK $10,001-$20,000, HK $20,001-$30,000, and >HK $30,000 to obtain robust outcomes on regression analyses.

Other demographic data included sex, age, and marital status (never been married, married or cohabitating, and divorced or separated or widowed). Employment status was categorized as economically active (full-time work or part-time work) and economically inactive (student, homemaker, unemployed, and retiree) [29]. Any chronic diseases were self-reported (any or none).

Statistical Analysis

All data were weighted by sex, age, and education levels in accordance with the 2016 population by census to improve the representativeness of the sample.

First, disparities in seeking web-based information on COVID-19 (dichotomized variable) by SES were assessed through multivariable logistic regression, which yielded adjusted odds ratios (aOR) for seeking web-based information on COVID-19. Second, socioeconomic disparities in eHealth literacy, being a continuous variable, were assessed through linear regression, which yielded unstandardized regression coefficients to reflect the change in the eHEALS score for a unit change in the independent variable. Third, we used the slope index of inequality (SII) to estimate the absolute difference in the eHEALS score between the most advantaged and most disadvantaged groups. The SII has been recommended by the World Health Organization, and a high SII indicates severe inequality [30]. Income categories were first ranked from the lowest to highest, and the cumulative proportion of participants were assigned to each category on the basis of the midpoint of range as the code for each category. The eHEALS score was then regressed against the cumulative proportion of each income category [30]. A similar analysis was performed for education-related SII. As each personal preventive behavior was dichotomized as low and high adherence, the associations (determined with aOR and 95% CI values) of seeking web-based information on COVID-19 and the eHEALS score with each personal preventive behavior were estimated through multivariable logistic regression adjusted for demographic variables, SES, and chronic disease. All analyses were performed using Stata (version 15.1, Stata Corp).

Results

Table 1 shows the weighted sample (N=1501; females: n=829, 52.6%), with 495 (27.7%) participants aged ≥60 years. Approximately two-third (66.1%) participants were married or cohabitating, and 981 (62.9%) were economically active. Most participants had attained secondary or tertiary and above education. In total, 519 (37.5%) participants' monthly personal income was HK ≤$10,000, and 1040 (67.8%) participants self-reported seeking web-based information on COVID-19. The mean eHEALS score was 26.10 (SD 7.70).

Table 1. Demographic characteristics, socioeconomic status, chronic disease outcomes, and seeking of web-based information on COVID-19 among the study participants (N=1501).

Characteristics	Participants		
	Number	Unweighted %	Weighted %[a]
Sex			
Male	672	44.8	47.5
Female	829	55.2	52.6
Age (years)			
18-39	497	33.1	33.8
40-59	509	33.9	38.5
≥60	495	33.0	27.7
Marital status			
Never been married	353	23.5	24.7
Married/cohabitating	1053	70.2	66.1
Divorced/separated/widowed	95	6.3	9.2
Education			
Primary or below	247	16.5	23.2
Secondary	864	57.6	45.4
Tertiary or above	390	26.0	31.4
Income (HK $)[b]			
≤10,000	519	34.6	37.5
10,001-20,000	519	34.6	30.7
20,001-30,000	268	17.9	17.5
>30,000	195	13.0	14.3
Employment status			
Economically active	981	65.4	62.9
Economically inactive	520	34.6	37.1
Chronic diseases[c]			
Any	187	12.5	15.0
None	1314	87.5	85.0
Seeking web-based information on COVID-19			
Yes	1040	69.3	67.8
No	461	30.7	32.2

[a]Weighted by sex, age, and education levels in accordance with the 2016 population by census.

[b]US $1=HK $7.8.

[c]Participants self-reported being diagnosed with any chronic disease (eg, hypertension, diabetes, or cancer).

Table 2 shows the inverse correlation between age and seeking of web-based information on COVID-19 (*P* for trend<.001). Education (secondary education: aOR 1.55, 95% CI 1.10-2.18; tertiary or above education: aOR 2.98, 95% CI 1.84-4.81; *P* for trend<.001), income (*P* for trend=.025), and the absence of chronic diseases (aOR 1.56, 95% 1.11-2.21) were associated with seeking of web-based information on COVID-19.

Table 2. Associations among demographic variables, socioeconomic status, and chronic disease with seeking of web-based information on COVID-19 (N=1501).

Characteristics	Sought (n=1040), n (%)[a]	No-sought (n=461), n (%)[a]	Association	
			Unadjusted OR[b] (95% CI)	Adjusted OR[c] (95% CI)
Sex				
Male	481 (49.6)	191 (42.9)	1	1
Female	559 (50.4)	270 (57.1)	0.82 (0.66-1.03)	0.83 (0.65-1.06)
Age (years)				
18-39	411 (41.4)	86 (17.8)	1	1
40-59	391 (41.4)	118 (32.4)	0.69 (0.51-0.95)[d]	0.86 (0.60-1.23)
≥60	238 (17.1)	257 (49.9)	0.19 (0.14-0.26)[e]	0.40 (0.27-0.61)[e]
Marital status				
Never been married	290 (30.1)	63 (13.5)	1	1
Married/cohabitating	702 (64.4)	351 (69.7)	0.43 (0.32-0.59)[e]	0.90 (0.62-1.30)
Divorced/separated/widowed	48 (5.5)	47 (16.8)	0.22 (0.14-0.36)[e]	0.65 (0.37-1.15)
Education				
Primary or below	105 (13.2)	142 (44.3)	1	1
Secondary	597 (47.1)	267 (41.8)	3.02 (2.26-4.04)[e]	1.55 (1.10-2.18)[d]
Tertiary or above	338 (39.8)	52 (13.9)	8.79 (5.98-12.93)[e]	2.98 (1.84-4.81)[e]
Income (HK $)[f]				
≤10,000	304 (28.7)	215 (55.9)	1	1
10,001-20,000	360 (32.3)	159 (27.4)	1.60 (1.24-2.07)[e]	0.97 (0.69-1.36)
20,001-30,000	208 (20.6)	60 (11.0)	2.45 (1.75-3.43)[e]	1.06 (0.69-1.63)
>30,000	168 (18.4)	27 (5.7)	4.40 (2.83-6.85)[e]	1.79 (1.04-3.06)[d]
Employment status				
Economically inactive	293 (27.5)	227 (57.4)	1	1
Economically active	747 (72.5)	234 (42.6)	2.47 (1.97-3.10)[e]	1.18 (0.84-1.65)
Chronic diseases[g]				
Any	91 (9.5)	96 (26.7)	1	1
None	949 (90.5)	365 (73.3)	2.74 (2.01-3.74)[e]	1.56 (1.11-2.21)[d]

[a]The proportion weighted by sex, age, and education levels in accordance with the 2016 population by census.

[b]OR: odds ratio

[c]Mutually adjusted for the variables in the table.

[d]*P*<.05.

[e]*P*<.001.

[f]US $1=HK $7.8.

[g]Self-reported by participants if having been diagnosed with a chronic disease (eg, hypertension, diabetes, or cancer).

Table 3 shows the inverse correlation between age and the eHEALS score (*P* for trend<.001). Education (secondary education: adjusted β 3.58, 95% CI 1.98-5.18; tertiary or above education: adjusted β 6.22, 95% CI 4.39-8.06; *P* for trend<.001) and income (*P* for trend<.001) were associated with the eHEALS score. The estimated difference in the eHEALS score between participants of the highest and the lowest SES was higher by education than by income (SII 13.27 vs 7.30). Sex, marital status, employment status, and chronic diseases were not associated with the eHEALS score after adjusting for age and SES.

Table 3. Associations among demographic variables, socioeconomic status, and chronic diseases with the eHealth literacy score[a] among participants seeking web-based information on COVID-19 (N=1040).

Characteristics	Mean (SD)	Unadjusted β (95% CI)	Adjusted β (95% CI)[b]	SII[c]
Sex				N/A[d]
Male	26.00 (7.51)	0	0	
Female	26.19 (7.86)	0.19 (–0.75 to 1.13)	–0.01 (–0.83 to 0.80)	
Age (years)				N/A
18-39	28.84 (6.07)	0	0	
40-59	27.18 (6.12)	–1.67 (–2.61 to –0.72)[e]	–0.77 (–1.82 to 0.28)	
≥60	19.60 (8.79)	–9.24 (–10.33 to –8.16)[f]	–5.48 (–6.91 to –4.05)[f]	
Marital status				N/A
Never been married	28.99 (5.93)	0	0	
Married/cohabitating	25.07 (7.93)	–3.92 (–4.95 to –2.90)[f]	–1.03 (–2.09 to 0.02)	
Divorced/separated/widowed	23.65 (9.07)	–5.35 (–7.64 to –3.06)[f]	–1.83 (–3.93 to 0.27)	
Education				13.27[f]
Primary or below	17.56 (8.45)	0	0	
Secondary	25.40 (6.80)	7.84 (6.42 to 9.26)[f]	3.58 (1.98 to 5.18)[f]	
Tertiary or above	29.98 (6.34)	12.42 (10.92 to 13.92)[f]	6.22 (4.39 to 8.06)[f]	
Income (HK $)[g]				7.30[f]
≤10,000	23.86 (8.27)	0	0	
10,001-20,000	25.38 (7.63)	1.52 (0.38 to 2.65)[e]	–0.40 (–1.69 to 0.88)	
20,001-30,000	27.74 (6.74)	3.88 (2.57 to 5.19)[f]	0.62 (–0.86 to 2.10)	
>30,000	29.67 (6.07)	5.81 (4.41 to 7.22)[f]	2.25 (0.63 to 3.88)[e]	
Employment status				N/A
Economically inactive	23.40 (8.75)	0	0	
Economically active	27.16 (6.97)	3.76 (2.74 to 4.78)[f]	0.39 (–0.89 to 1.66)	
Chronic diseases				N/A
None	26.42 (7.50)	0	0	
Any	22.71 (8.89)	–3.71 (–5.35 to –2.07)[f]	–0.85 (–2.31 to 0.60)	

[a]eHealth literacy scores ranged between 8 and 40, with higher scores indicating higher eHealth literacy.

[b]Mutually adjusted for the variables in the table.

[c]SII: slope index of inequality; SII refers to the absolute difference in the eHEALS score between the most advantaged and most disadvantaged groups, a higher score indicating a higher disparity in the eHEALS score.

[d]N/A: not applicable.

[e]*P*<.01.

[f]*P*<.001.

[g]US $1=HK $7.8.

Table 4 shows that participants who had sought web-based information on COVID-19 displayed higher adherence to the practice of wearing surgical masks (aOR 1.56, 95% CI 1.15-2.13), washing hands with alcohol-based sanitizers (aOR 1.33, 95% CI 1.05-1.71), adding water or bleach to the household drainage system (aOR 1.67, 95% CI 1.28-2.18), and social distancing (aOR 1.48, 95% CI 1.14-1.93) than those who did not seek web-based information on COVID-19. Seeking of web-based information on COVID-19 was not associated with the adherence to the practice of wearing a fabric mask. Among those who sought web-based information on COVID-19, the eHEALS score was associated with the adherence to the practice of wearing surgical masks (Q2: aOR 1.44, 95% CI 0.91-2.30; Q3: aOR 2.05, 95% CI 1.26-3.35; Q4: aOR 3.84, 95% CI

1.63-9.05; *P* for trend<.001; overall score: aOR 1.04, 95% CI 1.01-1.07). Regarding the adherence to washing hands with alcohol-based sanitizers, the aOR was 1.77 (95% CI 1.25-2.53) for Q2, 2.16 (95% CI 1.52-3.09) for Q3, 4.14 (95% CI 2.46-6.96) for Q4 (*P* for trend<.001), and 1.06 (95% CI 1.04-1.08) for the overall eHEALS score. Similarly, the eHEALS score was associated with the adherence to adding water or bleach to the household drainage system (Q2: aOR 1.47, 95% CI 1.02-2.15; Q3: aOR 1.89, 95% CI 1.30-2.75; Q4: aOR 1.94, 95% CI 1.19-3.16; *P* for trend=.001; overall score: aOR 1.04, 95% CI 1.02-1.06) and social distancing (Q2: aOR 1.68, 95% CI 1.16-2.44; Q3: aOR 1.58, 95% CI 1.09-2.30; Q4: aOR 2.25, 95% CI 1.39-3.65; *P* for trend=.002; overall score: aOR 1.03, 95% CI 1.01-1.05). We observed no association between the eHEALS score and the practice of wearing fabric masks.

Table 4. Association between the adherence to preventive behaviors by seeking web-based information on COVID-19 and the eHealth literacy score.

Parameter	Wearing a surgical mask[a]			Wearing a fabric mask[a]			Washing hands with alcohol-based sanitizers[a]			Adding water or bleach to the household drainage system[a]			Social distancing (eg, by 1.5 meters)[a]		
	n (%)	aOR[b] (95% CI)	*P* for trend value	n (%)	aOR (95% CI)	*P* for trend value	n (%)	aOR (95% CI)	*P* for trend value	n (%)	aOR (95% CI)	*P* for trend value	n (%)	aOR (95% CI)	*P* for trend value
Sought web-based information on COVID-19 (n=1501)			N/A[c]			N/A			N/A			N/A			N/A
No	359 (77.9)	1		85 (18.4)	1		191 (41.4)	1		122 (26.5)	1		122 (26.5)	1	
Yes	899 (86.4)	1.56 (1.15-2.13)[d]		166 (16.0)	0.84 (0.61-1.15)		572 (55.0)	1.33 (1.05-1.71)[e]		385 (37.0)	1.67 (1.28-2.18)[f]		377 (36.3)	1.48 (1.14-1.93)[d]	
eHealth literacy score categories for seekers of web-based information on COVID-19 (n=1040)			<.001			.39			<.001			.001			.002
Q1[g]	224 (79.4)	1		38 (13.5)	1		109 (38.7)	1		83 (29.4)	1		76 (27.0)	1	
Q2	244 (85.9)	1.44 (0.91-2.30)		44 (15.5)	1.17 (0.72-1.90)		153 (53.9)	1.77 (1.25-2.53)[d]		102 (35.9)	1.47 (1.02-2.15)[e]		108 (38.0)	1.68 (1.16-2.44)[d]	
Q3	309 (89.6)	2.05 (1.26-3.35)[d]		64 (18.6)	1.39 (0.86-2.22)		210 (60.9)	2.16 (1.52-3.09)[f]		142 (41.2)	1.89 (1.30-2.75)[d]		130 (37.7)	1.58 (1.09-2.30)[e]	
Q4	122 (94.6)	3.84 (1.63-9.05)[d]		20 (15.5)	1.12 (0.59-2.12)		100 (77.5)	4.14 (2.46-6.96)[f]		58 (45.0)	1.94 (1.19-3.16)[d]		63 (48.8)	2.25 (1.39-3.65)[d]	
eHealth literacy score (continuous variable) for seekers of web-based information on COVID-19 (n=1040)			N/A			N/A			N/A			N/A			N/A
Overall score		1.04 (1.01-1.07)[d]			1.01 (0.99-1.04)			1.06 (1.04-1.08)[f]			1.04 (1.02-1.06)[f]			1.03 (1.01-1.05)[d]	

[a]All preventive behaviors: high adherence (“often”) vs low adherence (“never,” “occasionally,” and “sometimes”).

[b]aOR: adjusted odds ratio; the aOR was adjusted for sex, age, marital status, employment, education, income, and chronic diseases.

[c]N/A: not applicable.

[d]*P*<.01.

[e]*P*<.05.

[f]*P*<.001.

[g]The eHealth literacy score was divided into 4 categories (Q1-Q4) on the basis of the quartile values (median 28, IQR 22-32); a higher score indicated higher eHealth literacy.

Discussion

Principal Findings

This study is the first to report socioeconomic disparities in seeking web-based information on COVID-19 and eHealth literacy during the COVID-19 pandemic and their association with a high adherence to COVID-19–related preventive behaviors, including wearing surgical masks, washing hands, adding water or bleach to the household drainage system, and social distancing.

Seeking of web-based information on COVID-19 was observed among younger participants in our study, concurrent with previous studies on web-based health information seeking behaviors [31]. A recent study also indicated that younger family members sought web-based information for the elderly during the pandemic [32]. Such an age disparity in information seeking can be attributed to the higher penetration rate of internet devices such as personal computers and smartphones among younger rather than older individuals [18]. Small font sizes, crowded visual presentations, and distracting flashes on most web-based information sources could be barriers to web-based information seeking among the elderly [33]. More frequent health information seeking from traditional media such as the radio and newspapers were observed among the elderly in our previous population-based study [20]. Our finding that higher SES including education levels and income is associated with seeking web-based information on COVID-19 is consistent with that of previous studies on seeking of web-based health information conducted before the COVID-19 pandemic [13,31]. Compared to our previous study, which measured SES disparities in seeking of web-based health information in 2009-2012, the ORs of web-based information seeking were found to decrease (eg, tertiary or above education: 2.98 in 2020 vs 8.00 in 2009-2012) [20]. Such a reduction in the effect size could result from the increased popularity of internet-accessible devices in the general population in Hong Kong [34]. Alternatively, the decreased ORs could be attributed to increased information seeking behaviors in crisis events, which were suggested as a means of reducing situation uncertainty and controlling risk [35].

Furthermore, we found age and SES disparities in eHealth literacy, thus revealing disparities in locating, understanding, and the utility of web-based information among those who sought web-based information on COVID-19. The associations between sociodemographic characteristics (eg, age and SES) and eHealth literacy observed in our study were concurrent with previous findings on health literacy [36,37]. Our study focused on eHealth literacy because the Internet has been the major platform for disseminating health information during the COVID-19 pandemic, since it is easily available and can instantly update information on, for example, preventive behaviors and access to social and health services. Considering that web-based information is complex and misinformation on the internet has led to inappropriate behaviors [38], those who used the internet for health information but with limited eHealth literacy skills to discern the quality of different information sources were the potential at-risk populations and require more attention. We observed stronger associations between eHealth literacy and education rather than income, which probably reflects the notion that knowledge and skills are more affected by cognitive function than by the available material. The education-related disparity in eHealth literacy was larger than the income-related disparity in our study, which suggest that education plays a more crucial role than income in affecting eHealth literacy. Other studies have also suggested that disparities in eHealth literacy were due to knowledge gaps rather than merely physical barriers to internet access [39]. Furthermore, we noticed that eHealth literacy and seeking of web-based information on COVID-19 have similar risk factors including an older age and a lower SES [13]. eHealth literacy enables seeking of web-based information [8]; further studies can explore the extent to which low eHealth literacy hinders seeking of web-based information among older and low-SES individuals.

Successful control of the COVID-19 pandemic would need universal adherence to preventive behaviors, which have proven very effective in reducing disease spread [40]. Seeking of web-based information on COVID-19 is associated with the adherence to preventive behaviors, suggesting the need to understand the barriers to the low SES group, including low eHealth literacy to use the internet to obtain health information. Our participants with higher eHealth literacy showed high adherence to personal preventive behaviors, which was consistent with the results of previous non–COVID-19 studies that eHealth literacy correlated with health behaviors such as regular physical exercise and balanced diets [24,41,42]. Our study extended those findings to COVID-19 preventive behaviors in the specific context of the COVID-19 pandemic, in which increasing misinformation has been disseminated through the internet. Low eHealth literacy could lead to difficulties in fact-checking and mistrust in peoples' beliefs in coronavirus conspiracies, which would impede the performance of preventive behaviors [43]. Such disparities in eHealth literacy have led to disparities in guidelines on preventive behaviors, their profound consequence being health inequality [44]. Web-based information should be better designed to address the eHealth literacy levels of target users, particularly those with a low SES, to bridge the existing knowledge gap. Further studies are needed to explore how to improve eHealth literacy effectively and the approach involving the use of eHealth literacy to facilitate better health behaviors.

Limitations

Our study has some limitations. First, the cross-sectional data cannot confirm the causal association, although it is unlikely that higher eHealth literacy or seeking of web-based information on COVID-19 would lead to higher education and income. Second, we measured the perceived eHealth literacy instead of actual performance on the internet. Some studies have measured performed eHealth literacy and reported a weak or moderate correlation between perceived and performed eHealth literacy [15,45]. Third, eHEALS, the most commonly used validated scale, was developed at the early stage of internet technology; its fit with Web 2.0-related technologies (social media) was not clear because of the considerable changes on the internet (more participatory and interactive web) [46]. Future studies are needed

to improve the model of eHealth literacy with the evolution of the internet and the COVID-19 pandemic [46,47]. Fourth, we did not collect data on channels of web-based information on COVID-19; hence, further studies should include details of the frequency and channels on the internet for seeking information on COVID-19.

Conclusions

This study provides the first evidence that Chinese adults with a higher SES had higher eHealth literacy and sought web-based information on COVID-19, and that both these factors are associated with high adherence to the guidelines on preventive behaviors during the COVID-19 pandemic. Effective interventions are needed to enhance the low eHealth literacy skills of low-SES individuals to combat the COVID-19 pandemic.

Acknowledgments

We thank the Social Policy Research for conducting the survey and the participants for their responses in the study.

Authors' Contributions

SZZ, JYHW, and MPW conceived and designed the study. ZG, YW, and MPW analyzed the data. ZG, NG, and MPW drafted the manuscript. All authors interpreted the data, critically reviewed the manuscript, and provided their final approval for submission of the manuscript for publication.

Conflicts of Interest

None declared.

Multimedia Appendix 1
eHealth Literacy Scale.
[PDF File (Adobe PDF File), 128 KB - jmir_v23i4e24577_app1.pdf]

Multimedia Appendix 2
Unweighted prevalence of preventive behaviors by online COVID-19 information seeking (N=1501).
[PDF File (Adobe PDF File), 39 KB - jmir_v23i4e24577_app2.pdf]

References

1. Zarocostas J. How to fight an infodemic. Lancet 2020 Feb 29;395(10225):676 [FREE Full text] [doi: 10.1016/S0140-6736(20)30461-X] [Medline: 32113495]
2. Liu PL. COVID-19 Information Seeking on Digital Media and Preventive Behaviors: The Mediation Role of Worry. Cyberpsychol Behav Soc Netw 2020 Oct;23(10):677-682. [doi: 10.1089/cyber.2020.0250] [Medline: 32498549]
3. Xie B, He D, Mercer T, Wang Y, Wu D, Fleischmann K, et al. Global health crises are also information crises: A call to action. J Assoc Inf Sci Technol 2020 Mar 13:1419-1423 [FREE Full text] [doi: 10.1002/asi.24357] [Medline: 32427189]
4. Tasnim S, Hossain MM, Mazumder H. Impact of Rumors and Misinformation on COVID-19 in Social Media. J Prev Med Public Health 2020 May;53(3):171-174 [FREE Full text] [doi: 10.3961/jpmph.20.094] [Medline: 32498140]
5. Freeman D, Waite F, Rosebrock L, Petit A, Causier C, East A, et al. Coronavirus conspiracy beliefs, mistrust, and compliance with government guidelines in England. Psychol Med 2020 May 21:1-13 [FREE Full text] [doi: 10.1017/S0033291720001890] [Medline: 32436485]
6. Norman CD, Skinner HA. eHealth Literacy: Essential Skills for Consumer Health in a Networked World. J Med Internet Res 2006 Jun 16;8(2):e9 [FREE Full text] [doi: 10.2196/jmir.8.2.e9] [Medline: 16867972]
7. Neter E, Brainin E. eHealth literacy: extending the digital divide to the realm of health information. J Med Internet Res 2012 Jan;14(1):e19 [FREE Full text] [doi: 10.2196/jmir.1619] [Medline: 22357448]
8. Wong DK, Cheung M. Online Health Information Seeking and eHealth Literacy Among Patients Attending a Primary Care Clinic in Hong Kong: A Cross-Sectional Survey. J Med Internet Res 2019 Mar 27;21(3):e10831 [FREE Full text] [doi: 10.2196/10831] [Medline: 30916666]
9. Pirisi A. Low health literacy prevents equal access to care. Lancet 2000 Nov 25;356(9244):1828. [doi: 10.1016/S0140-6736(05)73297-9] [Medline: 11117924]
10. Beaunoyer E, Dupéré S, Guitton MJ. COVID-19 and digital inequalities: Reciprocal impacts and mitigation strategies. Comput Human Behav 2020 Oct;111:106424 [FREE Full text] [doi: 10.1016/j.chb.2020.106424] [Medline: 32398890]
11. Morgan-Daniel J, Ansell M, Adkins L. COVID-19 Patient Education and Consumer Health Information Resources and Services. J Consum Health Internet 2020 Sep 01;24(3):302-313 [FREE Full text] [doi: 10.1080/15398285.2020.1792166]

12. Levin-Zamir D, Bertschi I. Media Health Literacy, eHealth Literacy, and the Role of the Social Environment in Context. Int J Environ Res Public Health 2018 Aug 03;15(8):1643 [FREE Full text] [doi: 10.3390/ijerph15081643] [Medline: 30081465]
13. Tennant B, Stellefson M, Dodd V, Chaney B, Chaney D, Paige S, et al. eHealth literacy and Web 2.0 health information seeking behaviors among baby boomers and older adults. J Med Internet Res 2015 Mar 17;17(3):e70 [FREE Full text] [doi: 10.2196/jmir.3992] [Medline: 25783036]
14. Mitsutake S, Shibata A, Ishii K, Oka K. Association of eHealth literacy with colorectal cancer knowledge and screening practice among internet users in Japan. J Med Internet Res 2012 Nov 13;14(6):e153 [FREE Full text] [doi: 10.2196/jmir.1927] [Medline: 23149453]
15. van der Vaart R, van Deursen AJ, Drossaert CH, Taal E, van Dijk JA, van de Laar MA. Does the eHealth Literacy Scale (eHEALS) measure what it intends to measure? Validation of a Dutch version of the eHEALS in two adult populations. J Med Internet Res 2011 Nov 09;13(4):e86 [FREE Full text] [doi: 10.2196/jmir.1840] [Medline: 22071338]
16. Patel JA, Nielsen FBH, Badiani AA, Assi S, Unadkat VA, Patel B, et al. Poverty, inequality and COVID-19: the forgotten vulnerable. Public Health 2020 Jun;183:110-111 [FREE Full text] [doi: 10.1016/j.puhe.2020.05.006] [Medline: 32502699]
17. Thematic Report: Household Income Distribution in Hong Kong. Census and Statistics Department: Hong Kong Special Administrative Region. 2017 Jun. URL: https://www.statistics.gov.hk/pub/B11200962016XXXXB0100.pdf [accessed 2021-04-01]
18. Thematic Household Survey Report No. 69. Census and Statistics Department: Hong Kong Special Administrative Region. 2020 Mar. URL: https://www.statistics.gov.hk/pub/B11302692020XXXXB0100.pdf [accessed 2021-04-01]
19. Kwok KO, Li KK, Chan HHH, Yi YY, Tang A, Wei WI, et al. Community Responses during Early Phase of COVID-19 Epidemic, Hong Kong. Emerg Infect Dis 2020 Jul;26(7):1575-1579 [FREE Full text] [doi: 10.3201/eid2607.200500] [Medline: 32298227]
20. Wang MP, Viswanath K, Lam TH, Wang X, Chan SS. Social determinants of health information seeking among Chinese adults in Hong Kong. PLoS One 2013;8(8):e73049 [FREE Full text] [doi: 10.1371/journal.pone.0073049] [Medline: 24009729]
21. Li J, Zheng H. Online InformationSeeking and Disease Prevention Intent During COVID-19 Outbreak. Journal Mass Commun Q 2020 Sep 28:107769902096151. [doi: 10.1177/1077699020961518]
22. Luk TT, Zhao S, Weng X, Wong JY, Wu YS, Ho SY, et al. Exposure to health misinformation about COVID-19 and increased tobacco and alcohol use: a population-based survey in Hong Kong. Tob Control 2020 Aug 27. [doi: 10.1136/tobaccocontrol-2020-055960] [Medline: 32855353]
23. Zhao SZ, Wong JYH, Wu Y, Choi EPH, Wang MP, Lam TH. Social Distancing Compliance under COVID-19 Pandemic and Mental Health Impacts: A Population-Based Study. Int J Environ Res Public Health 2020 Sep 14;17(18):6692 [FREE Full text] [doi: 10.3390/ijerph17186692] [Medline: 32937929]
24. Mitsutake S, Shibata A, Ishii K, Oka K. Associations of eHealth Literacy With Health Behavior Among Adult Internet Users. J Med Internet Res 2016 Jul 18;18(7):e192 [FREE Full text] [doi: 10.2196/jmir.5413] [Medline: 27432783]
25. Shuai JG, Xiao MY, Yu YS, Dan N, Xue ML, Lu W. Adaptation and evaluation of Chinese version of eHEALS and its usage among senior high school students. Chin J Health Educ 2013;29(2):108.
26. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Struct Equ Modeling 1999 Jan;6(1):1-55. [doi: 10.1080/10705519909540118]
27. Richtering SS, Hyun K, Neubeck L, Coorey G, Chalmers J, Usherwood T, et al. eHealth Literacy: Predictors in a Population With Moderate-to-High Cardiovascular Risk. JMIR Hum Factors 2017 Jan 27;4(1):e4 [FREE Full text] [doi: 10.2196/humanfactors.6217] [Medline: 28130203]
28. Infection prevention and control during health care when novel coronavirus (nCoV) infection is suspected: interim guidance, 25 January 2020. World Health Organization. 2020. URL: https://apps.who.int/iris/handle/10665/330674 [accessed 2021-04-01]
29. Hong Kong Poverty Situation Report 2018. Government of the Hong Kong Special Administrative Region. 2019 Dec. URL: https://www.statistics.gov.hk/pub/B9XX0005E2018AN18E0100.pdf [accessed 2021-04-01]
30. Health Equity Assessment Toolkit (HEAT): Software for exploring and comparing health inequalities in countries. Built-in database edition. Version 2.0. World Health Organization. 2017. URL: https://www.who.int/gho/health_equity/heat_technical_notes.pdf [accessed 2021-04-01]
31. Din HN, McDaniels-Davidson C, Nodora J, Madanat H. Profiles of a Health Information-Seeking Population and the Current Digital Divide: Cross-Sectional Analysis of the 2015-2016 California Health Interview Survey. J Med Internet Res 2019 May 14;21(5):e11931 [FREE Full text] [doi: 10.2196/11931] [Medline: 31094350]
32. Zhao X, Fan J, Basnyat I, Hu B. Online Health Information Seeking Using "#COVID-19 Patient Seeking Help" on Weibo in Wuhan, China: Descriptive Study. J Med Internet Res 2020 Oct 15;22(10):e22910 [FREE Full text] [doi: 10.2196/22910] [Medline: 33001838]
33. Czaja S, Boot W, Charness N, Rogers W. Individual differences. In: Designing for Older Adults. 3rd edition. Boca Raton, FL: CRC Press; 2019.

34. Usage of Information Technology and the Internet by Hong Kong Residents. Census and Statistics Department: The Government of the Hong Kong Special Administrative Region. 2020 Oct. URL: https://www.censtatd.gov.hk/hkstat/sub/sp120.jsp?productCode=FA100109 [accessed 2021-04-01]
35. Jones NM, Thompson RR, Dunkel Schetter C, Silver RC. Distress and rumor exposure on social media during a campus lockdown. Proc Natl Acad Sci USA 2017 Oct 31;114(44):11663-11668 [FREE Full text] [doi: 10.1073/pnas.1708518114] [Medline: 29042513]
36. Rikard RV, Thompson MS, McKinney J, Beauchamp A. Examining health literacy disparities in the United States: a third look at the National Assessment of Adult Literacy (NAAL). BMC Public Health 2016 Sep 13;16:975 [FREE Full text] [doi: 10.1186/s12889-016-3621-9] [Medline: 27624540]
37. Neter E, Brainin E, Baron-Epel O. The dimensionality of health literacy and eHealth literacy. Eur Health Psychol 2015;17(6):280.
38. Jayasinghe R, Ranasinghe S, Jayarajah U, Seneviratne S. Quality of online information for the general public on COVID-19. Patient Educ Couns 2020 Aug 07;103(12):2594-2597 [FREE Full text] [doi: 10.1016/j.pec.2020.08.001] [Medline: 32800422]
39. Viswanath K, Finnegan JR. The Knowledge Gap Hypothesis: Twenty-Five Years Later. Annals of the International Communication Association 2016 May 18;19(1):187-228. [doi: 10.1080/23808985.1996.11678931]
40. Cowling BJ, Ali ST, Ng TWY, Tsang TK, Li JCM, Fong MW, et al. Impact assessment of non-pharmaceutical interventions against coronavirus disease 2019 and influenza in Hong Kong: an observational study. Lancet Public Health 2020 May;5(5):e279-e288 [FREE Full text] [doi: 10.1016/S2468-2667(20)30090-6] [Medline: 32311320]
41. Hsu W, Chiang C, Yang S. The effect of individual factors on health behaviors among college students: the mediating effects of eHealth literacy. J Med Internet Res 2014 Dec 12;16(12):e287 [FREE Full text] [doi: 10.2196/jmir.3542] [Medline: 25499086]
42. Kim S, Son Y. Relationships Between eHealth Literacy and Health Behaviors in Korean Adults. Comput Inform Nurs 2017 Feb;35(2):84-90. [doi: 10.1097/CIN.0000000000000255] [Medline: 27258808]
43. Chong YY, Cheng HY, Chan HYL, Chien WT, Wong SYS. COVID-19 pandemic, infodemic and the role of eHealth literacy. Int J Nurs Stud 2020 Aug;108:103644 [FREE Full text] [doi: 10.1016/j.ijnurstu.2020.103644] [Medline: 32447127]
44. Sentell T, Vamos S, Okan O. Interdisciplinary Perspectives on Health Literacy Research Around the World: More Important Than Ever in a Time of COVID-19. Int J Environ Res Public Health 2020 Apr 26;17(9):3010 [FREE Full text] [doi: 10.3390/ijerph17093010] [Medline: 32357457]
45. Neter E, Brainin E. Perceived and Performed eHealth Literacy: Survey and Simulated Performance Test. JMIR Hum Factors 2017 Jan 17;4(1):e2 [FREE Full text] [doi: 10.2196/humanfactors.6523] [Medline: 28096068]
46. Norman C. eHealth literacy 2.0: problems and opportunities with an evolving concept. J Med Internet Res 2011 Dec 23;13(4):e125 [FREE Full text] [doi: 10.2196/jmir.2035] [Medline: 22193243]
47. Griebel L, Enwald H, Gilstad H, Pohl A, Moreland J, Sedlmayr M. eHealth literacy research-Quo vadis? Inform Health Soc Care 2018 Dec;43(4):427-442. [doi: 10.1080/17538157.2017.1364247] [Medline: 29045164]

Abbreviations

aOR: adjusted odds ratio
eHEALS: eHealth literacy scale
SES: socioeconomic status
SII: slope indices of inequality

Edited by G Eysenbach; submitted 25.09.20; peer-reviewed by E Neter, D He; comments to author 30.10.20; revised version received 19.11.20; accepted 12.12.20; published 14.04.21.

Please cite as:
Guo Z, Zhao SZ, Guo N, Wu Y, Weng X, Wong JYH, Lam TH, Wang MP
Socioeconomic Disparities in eHealth Literacy and Preventive Behaviors During the COVID-19 Pandemic in Hong Kong: Cross-sectional Study
J Med Internet Res 2021;23(4):e24577
URL: https://www.jmir.org/2021/4/e24577
doi:10.2196/24577
PMID:33784240

Original Paper

Reduction in Hospital System Opioid Prescribing for Acute Pain Through Default Prescription Preference Settings: Pre–Post Study

Benjamin Heritier Slovis[1,2], MD, MA; Jeffrey M Riggio[1,3], MD, MS; Melanie Girondo[4], MSN; Cara Martino[4], MSN; Bracken Babula[1,3], MD, MS; Lindsey M Roke[4], PharmD; John C Kairys[1,5], MD

[1]Office of Clinical Informatics, Thomas Jefferson University, Philadelphia, PA, United States

[2]Department of Emergency Medicine, Thomas Jefferson University, Philadelphia, PA, United States

[3]Department of Medicine, Thomas Jefferson University, Philadelphia, PA, United States

[4]Information Services and Technology, Thomas Jefferson University, Philadelphia, PA, United States

[5]Department of Surgery, Thomas Jefferson University, Philadelphia, PA, United States

Corresponding Author:
Benjamin Heritier Slovis, MD, MA
Office of Clinical Informatics
Thomas Jefferson University
833 Chestnut Street Floor 10
Philadelphia, PA, 19107
United States
Phone: 1 (215) 955 6844
Fax: 1 (215) 923 6225
Email: bxs088@jefferson.edu

Abstract

Background: The United States is in an opioid epidemic. Passive decision support in the electronic health record (EHR) through opioid prescription presets may aid in curbing opioid dependence.

Objective: The objective of this study is to determine whether modification of opioid prescribing presets in the EHR could change prescribing patterns for an entire hospital system.

Methods: We performed a quasi-experimental retrospective pre–post analysis of a 24-month period before and after modifications to our EHR's opioid prescription presets to match Centers for Disease Control and Prevention guidelines. We included all opioid prescriptions prescribed at our institution for nonchronic pain. Our modifications to the EHR include (1) making duration of treatment for an opioid prescription mandatory, (2) adding a quick button for 3 days' duration while removing others, and (3) setting the default quantity of all oral opioid formulations to 10 tablets. We examined the quantity in tablets, duration in days, and proportion of prescriptions greater than 90 morphine milligram equivalents/day for our hospital system, and compared these values before and after our intervention for effect.

Results: There were 78,246 prescriptions included in our study written on 30,975 unique patients. There was a significant reduction for all opioid prescriptions pre versus post in (1) the overall median quantity of tablets dispensed (54 [IQR 40-120] vs 42 [IQR 18-90]; $P<.001$), (2) median duration of treatment (10.5 days [IQR 5.0-30] vs 7.5 days [IQR 3.0-30]; $P<.001$), and (3) proportion of prescriptions greater than 90 morphine milligram equivalents/day (27.46% [10,704/38,976; 95% CI 27.02%-27.91%] vs 22.86% [8979/39,270; 95% CI 22.45%-23.28%]; $P<.001$).

Conclusions: Modifications of opioid prescribing presets in the EHR can improve prescribing practice patterns. Reducing duration and quantity of opioid prescriptions could reduce the risk of dependence and overdose.

(J Med Internet Res 2021;23(4):e24360) doi:10.2196/24360

KEYWORDS

informatics; electronic health record; opioids; prescriptions; oxycodone

Introduction

The opioid epidemic is a public health emergency unlike any other [1-3]. The roots of this crisis are founded in over 30 years of influential guidelines, marketing campaigns, and advertising that led to an increase in opioid prescriptions [4-8]. Overprescribing has increased since the early 1990s [9]. In 2016, there were 66.5 opioid prescriptions written for every 100 persons in the United States [10]. Prescription opioid abuse has been associated with progression to use of heroin [11], and the increased availability of both prescription and illicit opioids has led to a rise in the rate of death [12]. Opioids caused over 67% of drug-related deaths in 2017 and are responsible for nearly 400,000 deaths overall since 1999 [12].

Reductions in prescribing rates have become paramount to combating the opioid epidemic. Higher doses and longer durations of opioid therapy have been associated with an increased risk of chronic opioid use, with a significant increase in risk on treatment days 5 and 31 [13-15]. Therefore, the Centers for Disease Control and Prevention (CDC) recommends that caution be exercised when increasing doses to greater than 50 morphine milligram equivalents (MME) per day, and doses greater than 90 MME/day should be avoided. The CDC also states that 3 days or less is often sufficient for acute pain, and more than 7 days is “rarely needed” [16,17].

The HITECH act of 2009 made electronic health records (EHRs) ubiquitous. By 2017, 95% of hospitals in the United States are using EHRs [18]. EHRs and the technologies associated with them have been successfully utilized to combat the opioid epidemic. Electronic prescribing of controlled substances can improve medication safety, with a 2017 study demonstrating that an increasing number of prescribers are prescribing electronically [19]. Prescription drug monitoring programs, which require prescribers to review prior controlled substance prescriptions prior to prescribing, have shown reductions in opioid prescribing rates [20]. Computerized provider order entry (CPOE) systems are components of modern EHRs, and allow for reduction in practice variation and medical error by simplifying the prescribing of medications to an electronic process [21].

In order to curb overprescribing and reduce quantities of tablets, CPOE-linked interventions focus on preset defaulted prescription settings [22-24]. Previous studies have shown that the introduction or modification of opioid prescription presets has demonstrated reductions of tablet quantities and prescription MME for postsurgical patients [22], as well as significant reductions in the number of tablet quantities dispensed in the emergency department (ED) [23,25]. These studies suggest that modification of EHR-linked CPOE settings is a simple, inexpensive, and effective method of reducing the number of tablets associated with opioid prescriptions. However, all previous studies have focused on isolated emergency [23-27] or surgery departments [22]. We are unaware of any research examining the effects of modifications to opioid presets across an entire hospital system.

Slovis et al [27] previously demonstrated a reduction in duration of therapy and number of tablets dispensed when modifying the opioid prescription presets in our ED [27]. The success of this project prompted an enterprise-wide intervention to modify opioid prescription settings in our EHR’s CPOE.

Our institution’s CPOE consists of 4 fields for oral tablet prescription entry: (1) dose (number of tablets or milligrams per dose), (2) frequency (doses per day), (3) duration (number of days), and (4) quantity (number of tablets per prescription). Prior to our intervention dose, frequency and quantity were required for prescribing, but duration was not.

In August 2018, we implemented a number of interventions in our EHR’s CPOE: (1) we made the duration of treatment for an opioid prescription mandatory, (2) we provided a quick button for 3 days’ duration on all opioid prescriptions while removing all other quick buttons, and (3) we set the default quantity to 10 tablets for all oral opioid formulations, and removed any departmental variations. Prior to our intervention there was variability in durations and quantity dispensed for opioids (Multimedia Appendix 1). All interventions were passive in nature, no decision support alerts were part of the design, and there were no provider re-education measures nor large-scale announcement as part of the implementation design.

The purpose of this study was to assess the effects of modifications of opioid prescription default settings across an entire institution. We hypothesize that modifications of these settings can lead to reductions in duration of treatment, tablets dispensed, and proportion of prescriptions greater than 90 MME/day for a hospital system for patients with nonchronic pain. By demonstrating a reduction in prescribing at the hospital system level we have the potential to reduce risk of dependence, overdose, and possibly death.

Methods

Design, Setting, and Participants

In this quasi-experimental retrospective pre–post analysis, we examined the effects on prescribing patterns at our institution before and after modifications of opioid prescription settings. Thomas Jefferson University Hospitals, as an enterprise, has 908 acute care beds and over 2700 physicians and practitioners caring for more than 1.4 million people throughout the inpatient, outpatient, and ED settings [28]. This study was performed at the Center City Division, our urban academic institution comprising a large tertiary-care hospital, a community hospital, and multiple ambulatory clinics. Our study encompasses all outpatient prescriptions for opioids written between September 1, 2017, and August 31, 2019, in our EHR (Epic Systems Corporation) and the patients who received them. We did not include prescriptions for buprenorphine or methadone as these medications are used in medical assisted treatment (MAT) for opioid use disorder [29]. We limited our analysis to only oral capsule and tablet formulations (excluding oral liquid, buccal films, patches, etc.).

Our interest was in the effect of our intervention on opioids prescribed for acute pain. To this end, we excluded patients with chronic pain from our analysis via a modified approach to a validated algorithm [30]. Tian et al [30] achieved an accuracy of 95% for the identification of patients with chronic diseases

via (1) a single International Classification of Diseases, 9th Revision (ICD-9) code [31] "highly likely" to represent chronic pain OR (2) 2 or more ICD-9 codes "likely" to represent chronic pain separated by at least 30 days OR (3) receipt of at least 90 days of opioids OR (4) an ICD-9 code "likely" to represent chronic pain AND 2 or more pain scores greater than or equal to 4.

We mapped ICD-9 codes utilized in the algorithm to the 10th Revision (ICD-10) [32] using the Centers for Medicare and Medicaid Services General Equivalence Mappings [33] via the R package "touch" [34]. ICD code mappings were also manually reviewed. There were 9476 prescriptions in our cohort that did not include an ICD-10 and remained in our analysis as if they did not meet criteria for exclusion. Our data set included unreliable pain score documentation, so we eliminated this step in the algorithm. Pain scores held the lowest positive predictive value for identifying patients with chronic pain in the initial study [30], and we presumed that excluding these patients would only bias our results against our objective, as patients with chronic pain are presumed to have long-term prescriptions. Finally, we excluded all prescriptions written in the ED given the results of our previous intervention.

Variables

Using our analytics software Qlik Sense (QlikTech International), we identified all prescriptions for non-MAT opioid medications written during the study period and extracted a number of variables for each individual prescription and the associated patient.

At the prescription level these variables were medication, number of tablets, dosage unit (ie, mg), route of administration (oral vs rectal), and frequency of administration. Using our previously described method, we also included the calculated duration of therapy [35]. We utilized this value to represent the duration of the prescription as our prior work demonstrated that what is documented in the prescription is at times unreliable [35]. We also extracted precalculated MME/day for each prescription. At the patient level we extracted demographic data including age and sex.

Data Cleaning, Outcomes, and Statistical Analysis

In order to mitigate bias, all information was obtained through the same data query, and all participants were selected in the same way (eg, a script for an opioid medication). Participants were retrospectively recruited from a continuous 24 months.

Participants were divided into 2 cohorts (12 months before and after the intervention). The intervention was introduced on August 24, 2018, so the month of August 2018 was included in the preintervention cohort to avoid influencing postintervention results.

We calculated and compared frequencies of missing values for metrics of interest before and after the intervention. Demographic information was compared to assess similarity between the 2 cohorts.

The total number of prescriptions and unique patients were evaluated before and after the intervention. To ensure our intervention did not inadvertently cause an increase in the frequency of opioid prescribing, we compared the median number of prescriptions per patient per month as well as the odds of a patient receiving more than 1 prescription or refills.

The proportion of prescriptions was calculated for each class of opioid medication. Opioid class was determined by the active ingredient in the compound. We chose not to separate by schedule class as these can change [36,37]. We compared the proportions of opioid classes before and after the intervention and then measured if our intervention changed the quantity, duration, or proportion of prescriptions greater than 90 MME/day for all prescriptions, as well as for each opioid class.

For statistical analysis, Wilcoxon rank-sum test was utilized for nonparametric data, and *t* test for parametric data. Chi-square and Fisher exact test were performed for comparison of categorical values. Confidence intervals were included for all appropriate analysis. Statistics were performed in R statistical software (R Core Team).

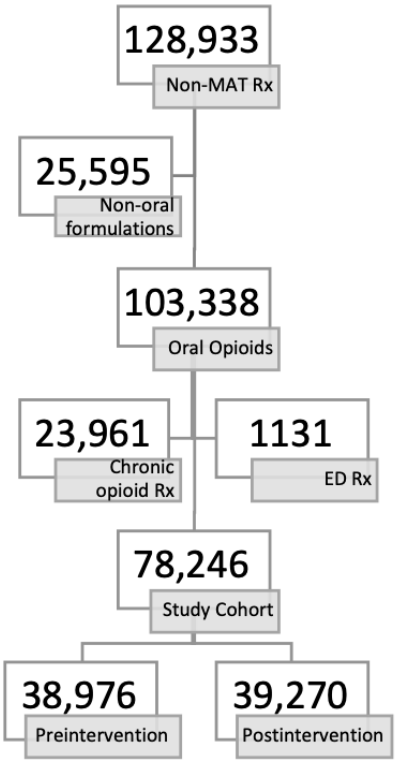
Results

Overview of Prescriptions

There were 128,933 non-MAT opioid prescriptions written during our study period. Of these, 103,338/128,933 (80.15%) were written for oral tablet or capsule formulations. We excluded 23,961/103,338 (23.19%) prescriptions that were written for patients with chronic pain. Of the remaining 79,377, we excluded 1131/79,377 prescriptions (1.42%) that were written by the ED. The median duration for these ED prescriptions was 2.7 days (IQR 2.5-3.3) with a median dispensed quantity of 9 tablets (IQR 9-31).

The remaining 78,246 prescriptions were for 30,975 unique patients. There were 38,976/78,246 (49.81%) prescriptions written for 16,464/30,975 (53.15%) patients in the preintervention period and 39,270/78,246 (50.19%) prescriptions written for 17,399/30,975 (56.17%) patients in the postintervention period. Figure 1 demonstrates the inclusion and exclusion criteria leading to our study cohorts.

Figure 1. Breakdown of our study cohort included excluded groups. ED: emergency department; MAT: medical assisted treatment.



Missing Data

Of the 78,246 prescriptions, 7167 (9.16%; 95% CI 8.96-9.36) were missing a calculated duration. The proportion of missing calculated durations decreased from 9.91% (7758/78,246; (95% CI 9.62-1.02) to 8.41% (6578/78,246; 95% CI 8.14-8.70; *P*<.001) before and after the intervention. In addition, 7156/78,246 (9.15%; 95% CI 8.94-9.35) prescriptions were missing documented quantity dispensed with a reduction from 9.90% (7750/78,246; 95% CI 9.60-10.20) to 8.40% (6570/78,246; 95% CI 8.13-8.68; *P*<.001) before versus after the intervention. Finally, 1139/78,246 (1.46%; 95% CI 1.37-1.54) prescriptions were missing the MME/day field with a significant increase in this proportion before versus after the intervention (915/78,246 [1.17%; 95% CI 1.07-1.28] to 1358/78,246 [1.74%; 95% CI 1.61-1.87]; *P*<.001). Missing values were excluded from their respective analysis.

Patient Demographics

Overall, 17,344/30,975 (55.99%; 95% CI 55.44%-56.55%) of all unique patients were female. There was no significant difference in patient sex before and after the intervention (17,300/30,975 [55.85%; 95% CI 55.10%-56.61%] vs 17,237/30,975 [56.65%; 95% CI 55.91%-57.39%] female; *P*=.14). The median age was 59, which did not change.

Rate of Prescribing

There was no significant change in the median number of prescriptions written per month, before and after the intervention (3254.5 [IQR 3176.25-3331.25] vs 3338 [3157.0-3391.5]; *P*=.59). The median monthly prescriptions per person remained 1 (IQR 1-2). The odds of getting more than 1 opioid prescription after the intervention did not significantly increase (1.003, 95% CI 0.967-1.04) and the odds of being prescribed a refill decreased (0.801, 95% CI 0.747-0.858).

Types of Opioids Prescribed

The majority (48,261/78,246, 61.68%; 95% CI 61.34%-62.02%) of prescriptions during the study period were for oxycodone. Table 1 demonstrates the proportion of individual opioids.

There were small but significant reductions in the proportion of prescriptions for morphine—6.30% (2456/38,976; 95% CI 6.06%-6.55%) to 5.95% (2335/39,270; 95% CI 5.72%-6.19%); *P*=.04)—and oxymorphone—0.37% (143/38,976; 95% CI 0.31%-0.43%) to 0.24% (96/39,270; 95% CI 0.20%-0.30%; *P*=.002). There was no change in the proportion of the other opioids.

Table 1. Proportion of each opioid during the study period with 95% confidence intervals.

Medication	Proportion[a] (N=78,246), n (%)	95% CI
Codeine	379 (0.48)	0.44-0.54
Hydrocodone	13 (0.02)	0.009-0.03
Hydromorphone	3995 (5.11)	4.95-5.26
Meperidine	40 (0.05)	0.04-0.07
Morphine	4791 (6.12)	5.96-6.29
Oxycodone	48,261 (61.68)	61.34-62.02
Oxymorphone	239 (0.31)	0.27-0.35
Tapentadol	172 (0.22)	0.19-0.26
Tramadol	20,356 (26.02)	25.71-26.32

[a]Proportion of the number of that individual opioid prescriptions from the total 78,246 opioid prescriptions.

Dispensing and Duration

There was a significant reduction in the overall median quantity of opioid tablets dispensed before versus after the intervention (54 [IQR 40-120] vs 42 [IQR 18-90]; *P*<.001). There was also a reduction in median duration of treatment before and after the intervention (10.5 days [IQR 5.0-30] vs 7.5 days [IQR 3.0-30]; *P*<.001). Finally, there was a significant reduction in the proportion of prescriptions greater than 90 MME/day (27.46% [10,704/38,976; 95% CI 27.02%-27.91%] vs 22.86% [8979/39,270; 95% CI 22.45%-23.28%]; *P*<.001), despite no change in the median of 45 MME/day per prescription before and after the intervention. These results are displayed in Figure 2.

There was a significant reduction in all metrics for oxycodone. Tramadol and codeine demonstrated a reduction in tablets dispensed, while codeine also had a reduction in the duration of treatment (*P*<.001 for all three metrics). Hydromorphone had a significant reduction (*P*<.001) in MME greater than 90/day despite no change in the other metrics. Table 2 demonstrates the effect of our intervention overall, and on each type of opioid medication before and after the intervention.

Given the majority of prescriptions were for oxycodone, Figure 3 demonstrates monthly numbers of tablets dispensed, median duration of treatment, and proportion of prescriptions greater than 90 MME/day for oxycodone.

Figure 2. Monthly median tablets prescribed pre and post-intervention, monthly median duration (days) pre and post-intervention and proportion of prescriptions >90 MME/day pre and post-intervention for all opioids. MME: morphine milligram equivalents.

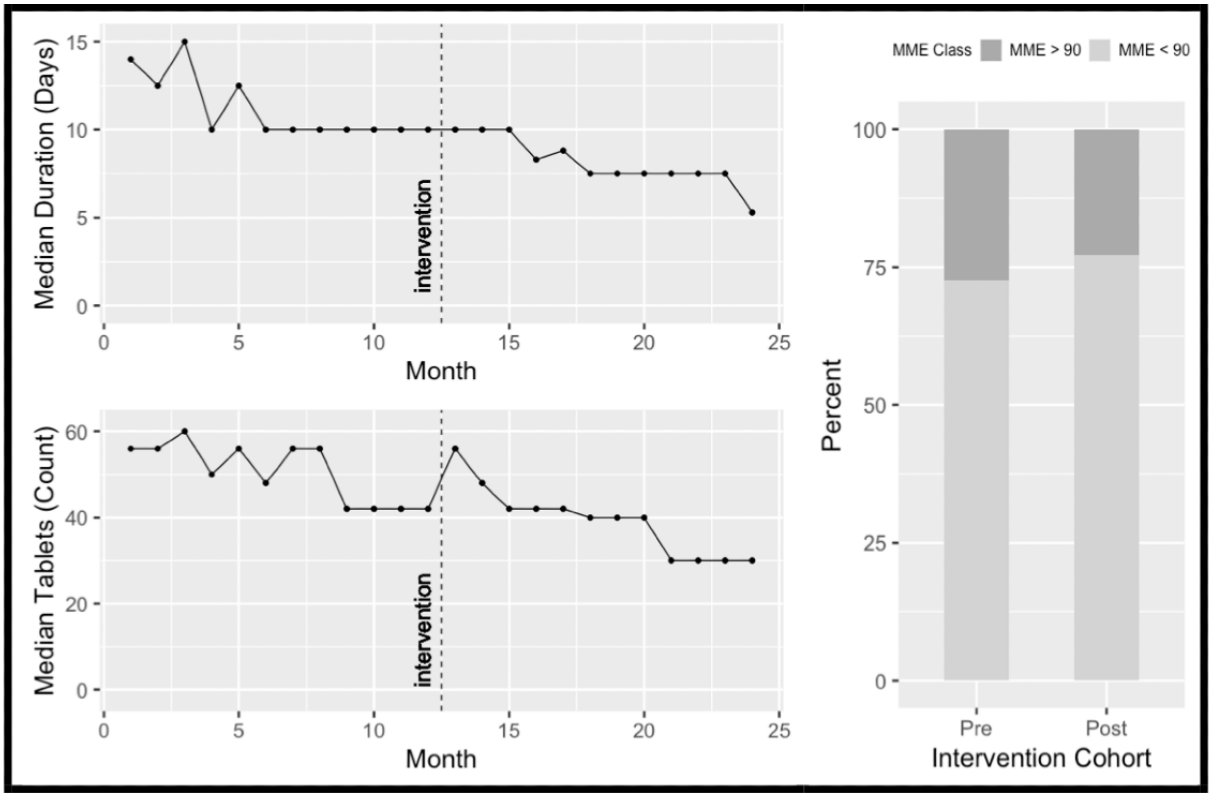
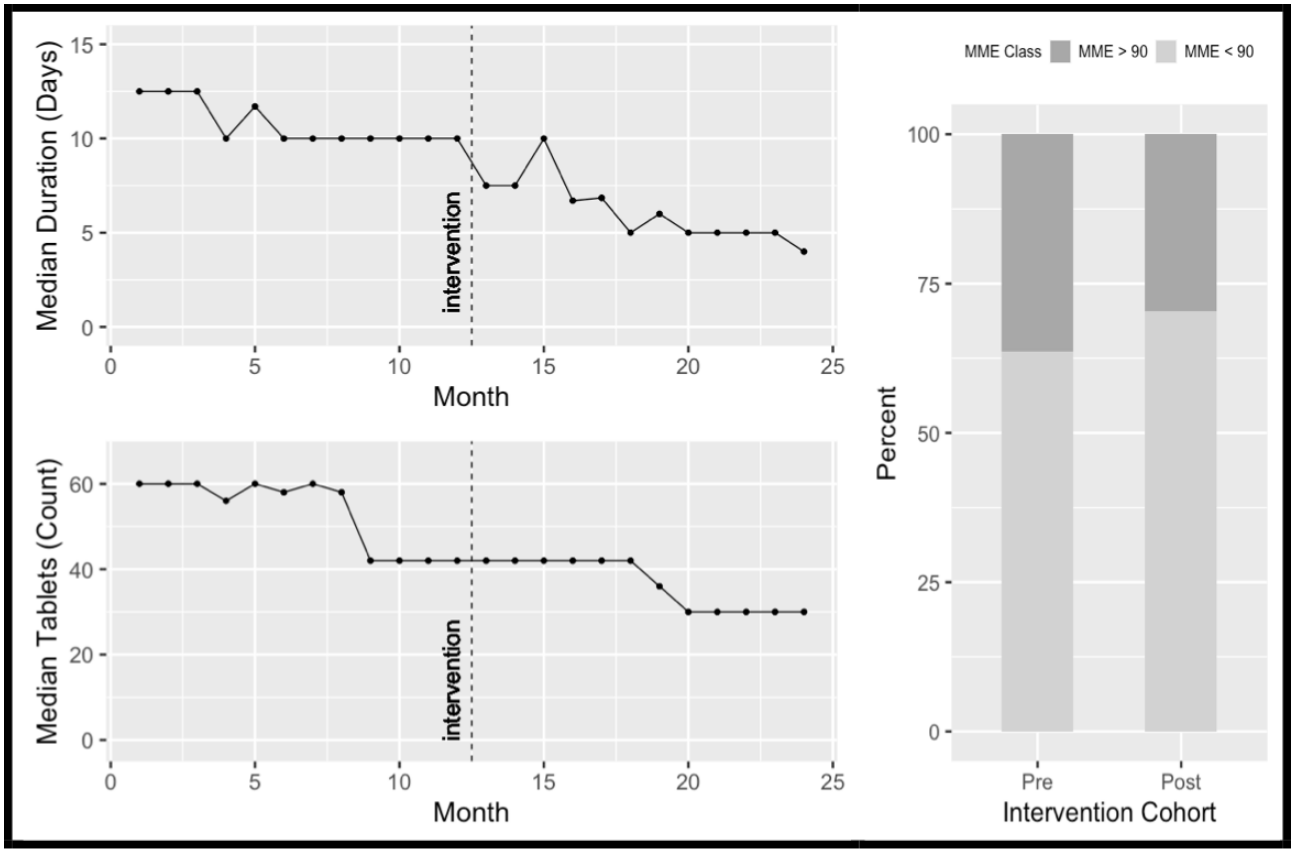


Table 2. Quantity, duration of treatment, and proportion of MME[a]/day >90 for all opioids and each individual opioid before and after the intervention.

Opioid	Before the intervention	After the intervention	*P* value
All opioids			
Quantity in tablets (IQR)	54 (40-120)	42 (18-90)	<.001
Duration in days (IQR)	10.5 (5.0-30.0)	7.5 (3.0-30)	<.001
>90 MME/day, n/N (%); 95% CI	10,704/38,976 (27.46); 27.02-27.91	8979/39,270 (22.86); 22.45-23.28	<.001
Codeine			
Quantity in tablets (IQR)	42 (40-42)	20 (18-42)	<.001
Duration in days (IQR)	5 (5-7.5)	2.5 (1.7-5.0)	<.001
>90 MME/day, n/N (%); 95% CI	0	0	
Hydrocodone			
Quantity in tablets (IQR)	42 (36-51)	45 (30-90)	.91
Duration in days (IQR)	15 (11-22.5)	11.25 (8.75-15.63)	.53
>90 MME/day, n/N (%); 95% CI	1/4 (25.00); 1.32-78.06	4/9 (44.44); 15.34-77.35	>.99
Hydromorphone			
Quantity in tablets (IQR)	60 (42-120)	60 (20-120)	
Duration in days (IQR)	7.5 (5-20)	7.5 (3.3-20)	
>90 MME/day, n/N (%); 95% CI	783/1985 (39.45); 37.29-41.64	664/2010 (33.03); 30.99-35.15	<.001
Meperidine			
Quantity in tablets (IQR)	144 (60-182)	100 (100-120)	.41
Duration in days (IQR)	10.5 (5.0-15.0)	6 (6.0-10.5)	.93
>90 MME/day, n/N (%); 95% CI	0	0	
Morphine			
Quantity in tablets (IQR)	60 (56-90)	60 (60-90)	
Duration in days (IQR)	30 (30-30)	30 (20-30)	
>90 MME/day, n/N (%); 95% CI	1092/2456 (44.46); 42.49-46.46	1048/2335 (44.88); 42.85-46.93	.87
Oxycodone			
Quantity in tablets (IQR)	56 (42-120)	42 (18-90)	<.001
Duration in days (IQR)	10 (5.0-30.0)	6 (2.5-30)	<.001
>90 MME/day, n/N (%); 95% CI	8722/23,962 (36.40); 35.79-37.01	7190/24,299 (29.59); 29.02-30.17	<.001
Oxymorphone			
Quantity in tablets (IQR)	60 (56-90)	62 (60-90)	.18
Duration in days (IQR)	30 (30-30)	30 (30-30)	
>90 MME/day, n/N (%); 95% CI	85/143 (59.44); 50.90-67.47	58/96 (60.42); 49.89-70.10	>.99
Tapentadol			
Quantity in tablets (IQR)	112 (60-166)	101 (60-120)	.29
Duration in days (IQR)	30 (30-30)	30 (30-30)	
>90 MME/day, n/N (%); 95% CI	21/82 (25.61); 16.89-36.65	15/90 (16.67); 9.93-26.32	.21
Tramadol			
Quantity in tablets (IQR)	40 (28-97.5)	30 (12-90)	<.001
Duration in days (IQR)	10 (7.5-20.0)	10 (5-20)	
>90 MME/day, n/N (%); 95% CI	0	0	

[a]MME: morphine milligram equivalents.

Figure 3. Monthly median number of tablets dispensed, median durations (days) of treatment and proportion of oxycodone prescriptions > 90 MME/day pre and post-intervention. MME: morphine milligram equivalents.



Discussion

Principal Findings

In this quasi-experimental pre–post study our intervention resulted in a significant reduction (P<.001) in the duration of treatment, quantity of tablets dispensed, and proportion of prescriptions greater than 90 MME/day for all opioid prescriptions written for patients with acute pain without an increase in the rate of prescribing or MME/day per prescription, while maintaining clinician autonomy. Our postintervention median duration of treatment for all opioids was 7.5 days, which is slightly longer than the CDC's recommendations of 7 days [16], but an improvement over our preintervention duration of 10 days. In addition, we improved compliance with CDC recommendations that opioid dosing should not exceed 90 MME/day [16].

Oxycodone accounted for more than 60% (48,261/78,246, 61.68%) of our institutions prescriptions and was the only opioid to demonstrate reductions in all metrics, while tramadol, codeine, and hydromorphone had reductions in at least one metric. Our results show an improved median duration of treatment for oxycodone of 6 days, which is within CDC's recommendations for acute pain.

Modifications to prescription presets is a relatively simple and effective way to combat the opioid epidemic. Our results are similar to other studies of single departments. Delgado et al [23] studied 2 EDs prescribing oxycodone. They examined the difference in prescribing patterns before and after the implementation of a new EHR, where they previously had no prescribing presets and the new EHR was preset to 10 tablets per prescription demonstrating a decrease in the median number of tablets from 11.3 and 12.6 to 10 and 10.9, respectively, in the 2 departments [23].

Chiu et al [22] looked at opioid prescribing in the outpatient surgical setting. Their study included 3 hospitals examining changes in preset opioid prescription quantities from 30 to 12 tablets. The intervention reduced the median number of tablets prescribed from 30 to 20, with a decrease of 5.22 tablets and 34.41 MME per prescription, and no difference in refill rates [22].

Despite their expectation of decreased tablets dispensed, Zwank et al [26] showed an increase from 15.31 to 15.77 in the mean number of tablets dispensed for hydrocodone and oxycodone in the ED after they removed a 15-tablet preset and required manual entry of a dispense quantity [26]. This may suggest that presets are vital to reduce prescribing quantities; however, this is contradicted by Santistevan et al [24] who demonstrated that removal of a default of 20 tablets for hydrocodone and oxycodone reduced the median number of tablets prescribed from 20 to 15, although their preintervention presets were higher than Zwank et al's.

Montoy et al [25] implemented a block-randomization study at 2 EDs. They examined 6 possible opioid tablet quantity presets: "status quo" 12 and 20, as well as "null," 5, 10, and 15. They

demonstrated that each tablet increase in preset yielded an increase of 0.19 tablets prescribed, and lower default quantities were associated with lower number of pills in 8 of 15 pairwise comparisons [25]. These results support our approach that modifications to prescription presets are an effective way to enact change in opioid prescribing.

Despite a number of publications demonstrating reductions in opioid prescribing through presets [22,23,25], to our knowledge this research is the first to demonstrate reduction in overall opioid prescribing and prescribing of multiple individual opioids for an entire hospital system. Our hospitals and ambulatory clinics are located in one of the most lethal counties in one of the most lethal states associated with the opioid epidemic [38-40]. The CDC estimates that the risk of chronic use of opioids is 13.5% after 8 days of treatment [15] and the World Health Organization estimates the annual rate for opioid-dependent individuals overdosing at 45% and death at 0.65% [41]. We reduced our median duration of treatment from 10.5 to 7.5 (6 days for oxycodone), thus potentially bringing more than half of our institutions' opioid prescriptions below this high-risk threshold.

Examining our hospital's nearly 3300 prescriptions per month and postintervention median reduction of 12 tablets per prescription, we can estimate that our hospital system has reduced nearly 39,600 opioid tablets from being prescribed per month, or 475,200 tablets per year. Extrapolate this to the over 168 million prescriptions written in the United States in 2018 [42], and our simple intervention could potentially reduce the number of tablets being prescribed each year by over 2 billion. Our intervention excluded prescriptions for chronic pain, thus these approximations are likely extreme, but even a small percentage of these estimates is a net large and clinically important reduction in opioid tablets. Significant reductions in the number of opioid tablets prescribed, durations of treatment, and doses greater than 90 MME/day can play an important role in curbing the epidemic, improving quality of care, and ultimately saving lives.

Limitations

Our findings have a number of limitations. First, although a validated algorithm, we did not complete Tian et al's [30] algorithmic approach to identifying patients with chronic pain with EHR data. We believe this is unlikely to significantly bias our results as this was the least sensitive element in the original study and exclusion of this step, if anything, biases our results away from our objective (ie, including more patients with chronic pain with presumably higher doses and longer prescription durations in our data set). In addition, this algorithm was designed for use with ICD-9 codes. We mapped ICD-9 to ICD-10 codes and performed a manual review, but errors may still have persisted. Second, with regard to our results, we did demonstrate that morphine and oxymorphone had slight reductions in proportion prescribed after the intervention; however, this was unlikely to have biased our results. We also should acknowledge that 9.16% (7167/78,246) of our study cohort were missing calculated durations and quantity of tablets dispensed and the percentage of these missing values decreased after the intervention, while 1.46% (1139/78,246) of prescriptions were missing MME/day and this increased after the intervention. We believe the majority of missing calculated durations are likely due to the missing dispense quantities, which are required for the computation of this metric. Missing MME/day is likely due to missing concentrations, doses, or conversion factors of some formulations. Some of these missing values may be due to prescriptions being refilled and bypassing some of the new requirements we implemented. We excluded missing data from our analysis, but it is possible this exclusion influenced our results.

For our MME/day threshold we chose to use the CDC guidelines for opioid prescribing as the CDC is a federal agency that upholds "the health of the people of the United States." [43] It should be noted that individual states may have opioid prescribing guidelines that deviate from those presented by the CDC [44].

It should also be noted that our research was a quasi-experimental pre–post study, and we cannot infer causality with our intervention and the results observed. It is possible that other influential factors could have played a role in the reduction of opioid prescribing at out institution during this 2-year period, as the opioid epidemic has come to the forefront of modern medicine. However, we believe our results are encouraging, and other studies have demonstrated similar outcomes to suggest our intervention likely played at least a role in the reduction of prescribing patterns at our institution. Finally, our region has one of the highest opioid overdose and death rates in the nation, making our results of high value to curbing the epidemic; however, our results may not be generalizable to the rest of the country.

Conclusions

In this quasi-experimental retrospective pre–post analysis, we demonstrated that modifications of the opioid prescribing presets in our hospital system's EHR can improve prescribing practice patterns and reduce the number of pills dispensed, duration of treatment, and proportion of prescriptions greater than 90 MME/day. We have identified a simple and effective way to reduce opioid prescribing, and to our knowledge, we are the first to perform this kind of an intervention throughout an entire hospital system, and not just at the department level. Reduction in opioid prescribing may aid in curbing the opioid epidemic, thus improving quality of care and potentially saving lives.

Acknowledgments

This publication was made possible, in part, by support from the Thomas Jefferson University + Philadelphia University Open Access Fund.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Prior Opioid Settings.
[DOCX File , 24 KB - jmir_v23i4e24360_app1.docx]

References

1. Srivastava AB, Gold MS. Beyond Supply: How We Must Tackle the Opioid Epidemic. Mayo Clin Proc 2018 Mar;93(3):269-272. [doi: 10.1016/j.mayocp.2018.01.018] [Medline: 29502558]
2. US Department of Health and Human Services. The Opioid Epidemic: By the Numbers. 2016. URL: https://www.hhs.gov/sites/default/files/Factsheet-opioids-061516.pdf [accessed 2021-03-26]
3. Rudd RA, Seth P, David F, Scholl L. Increases in Drug and Opioid-Involved Overdose Deaths - United States, 2010-2015. MMWR Morb Mortal Wkly Rep 2016 Dec 30;65(5051):1445-1452 [FREE Full text] [doi: 10.15585/mmwr.mm655051e1] [Medline: 28033313]
4. Rummans TA, Burton MC, Dawson NL. How Good Intentions Contributed to Bad Outcomes: The Opioid Crisis. Mayo Clin Proc 2018 Mar;93(3):344-350. [doi: 10.1016/j.mayocp.2017.12.020] [Medline: 29502564]
5. American Pain Society Quality of Care Committee. Quality improvement guidelines for the treatment of acute pain and cancer pain. JAMA 1995 Dec 20;274(23):1874-1880. [doi: 10.1001/jama.1995.03530230060032] [Medline: 7500539]
6. A consensus statement from the American Academy of Pain Medicine and the American Pain Society. The use of opioids for the treatment of chronic pain. Clin J Pain 1997 Mar;13(1):6-8. [Medline: 9084947]
7. Van Zee A. The promotion and marketing of oxycontin: commercial triumph, public health tragedy. Am J Public Health 2009 Mar;99(2):221-227. [doi: 10.2105/AJPH.2007.131714] [Medline: 18799767]
8. Purdue Pharma. Execs to pay $634.5 million fine in OxyContin case. 2007. URL: https://www.cnbc.com/id/18591525 [accessed 2021-03-26]
9. Nora D, Volkow MD. America's addiction to opioids: heroin and prescription drug abuse. Bethesda, MD: National Institute on Drug Abuse; 2014. URL: https://archives.drugabuse.gov/testimonies/2014/americas-addiction-to-opioids-heroin-prescription-drug-abuse [accessed 2021-03-26]
10. Annual surveillance report of drug-related risks and outcomes. Atlanta, GA: Centers for Disease Control and Prevention; 2017. URL: https://www.cdc.gov/drugoverdose/PDF/pubs/2017-CDC-drug-surveillance-report.pdf [accessed 2021-03-26]
11. Jones CM. Heroin use and heroin use risk behaviors among nonmedical users of prescription opioid pain relievers - United States, 2002-2004 and 2008-2010. Drug Alcohol Depend 2013 Sep 01;132(1-2):95-100. [doi: 10.1016/j.drugalcdep.2013.01.007] [Medline: 23410617]
12. Scholl L, Seth P, Kariisa M, Wilson N, Baldwin G. Drug and Opioid-Involved Overdose Deaths — United States, 2013–2017. MMWR Morb. Mortal. Wkly. Rep 2018 Dec 21;67(5152):1419-1427. [doi: 10.15585/mmwr.mm6751521e1]
13. Bohnert ASB, Valenstein M, Bair MJ, Ganoczy D, McCarthy JF, Ilgen MA, et al. Association between opioid prescribing patterns and opioid overdose-related deaths. JAMA 2011 Apr 06;305(13):1315-1321. [doi: 10.1001/jama.2011.370] [Medline: 21467284]
14. Edlund MJ, Martin BC, Russo JE, DeVries A, Braden JB, Sullivan MD. The role of opioid prescription in incident opioid abuse and dependence among individuals with chronic noncancer pain: the role of opioid prescription. Clin J Pain 2014 Jul;30(7):557-564 [FREE Full text] [doi: 10.1097/AJP.0000000000000021] [Medline: 24281273]
15. Shah A, Hayes CJ, Martin BC. Characteristics of Initial Prescription Episodes and Likelihood of Long-Term Opioid Use - United States, 2006-2015. MMWR Morb Mortal Wkly Rep 2017 Mar 17;66(10):265-269 [FREE Full text] [doi: 10.15585/mmwr.mm6610a1] [Medline: 28301454]
16. Dowell D, Haegerich TM, Chou R. CDC Guideline for Prescribing Opioids for Chronic Pain--United States, 2016. JAMA 2016 Apr 19;315(15):1624-1645 [FREE Full text] [doi: 10.1001/jama.2016.1464] [Medline: 26977696]
17. Dowell D, Haegerich TM, Chou R. CDC Guideline for Prescribing Opioids for Chronic Pain — United States, 2016. MMWR Recomm. Rep 2016 Mar 18;65(1):1-49. [doi: 10.15585/mmwr.rr6501e1]
18. Percent of Hospitals, By Type, that Possess Certified Health IT. Washington, DC: Office of the National Coordinator for Health Information Technology; 2018 Sep. URL: https://dashboard.healthit.gov/quickstats/pages/certified-electronic-health-record-technology-in-hospitals.php [accessed 2021-03-26]
19. Parasrampuria S, Blanco M, Barker W. Electronic Prescribing of Controlled Substances among Office-Based Physicians. Washington, DC: Office of the National Coordinator for Health Information Technology; 2017. URL: https://www.healthit.gov/sites/default/files/page/2019-09/officebasedphysicianelectronicprescribingofcontrolledsubstance2017.pdf [accessed 2021-03-26]
20. Wen H, Hockenberry JM, Jeng PJ, Bao Y. Prescription Drug Monitoring Program Mandates: Impact On Opioid Prescribing And Related Hospital Use. Health Aff (Millwood) 2019 Sep;38(9):1550-1556 [FREE Full text] [doi: 10.1377/hlthaff.2019.00103] [Medline: 31479368]

21. What is computerized provider order entry? Washington, DC: Office of the National Coordinator for Health Information Technology; 2018. URL: https://www.healthit.gov/faq/what-computerized-provider-order-entry [accessed 2018-03-21]
22. Chiu AS, Jean RA, Hoag JR, Freedman-Weiss M, Healy JM, Pei KY. Association of Lowering Default Pill Counts in Electronic Medical Record Systems With Postoperative Opioid Prescribing. JAMA Surg 2018 Nov 01;153(11):1012-1019 [FREE Full text] [doi: 10.1001/jamasurg.2018.2083] [Medline: 30027289]
23. Delgado MK, Shofer FS, Patel MS, Halpern S, Edwards C, Meisel ZF, et al. Association between Electronic Medical Record Implementation of Default Opioid Prescription Quantities and Prescribing Behavior in Two Emergency Departments. J Gen Intern Med 2018 Apr;33(4):409-411 [FREE Full text] [doi: 10.1007/s11606-017-4286-5] [Medline: 29340937]
24. Santistevan JR, Sharp BR, Hamedani AG, Fruhan S, Lee AW, Patterson BW. By Default: The Effect of Prepopulated Prescription Quantities on Opioid Prescribing in the Emergency Department. West J Emerg Med 2018 Mar;19(2):392-397 [FREE Full text] [doi: 10.5811/westjem.2017.10.33798] [Medline: 29560071]
25. Montoy JCC, Coralic Z, Herring AA, Clattenburg EJ, Raven MC. Association of Default Electronic Medical Record Settings With Health Care Professional Patterns of Opioid Prescribing in Emergency Departments: A Randomized Quality Improvement Study. JAMA Intern Med 2020 Apr 01;180(4):487-493. [doi: 10.1001/jamainternmed.2019.6544] [Medline: 31961377]
26. Zwank MD, Kennedy SM, Stuck LH, Gordon BD. Removing default dispense quantity from opioid prescriptions in the electronic medical record. Am J Emerg Med 2017 Oct;35(10):1567-1569. [doi: 10.1016/j.ajem.2017.04.002] [Medline: 28416266]
27. Slovis B, London K, Randolph F, Aini M, Mammen P, Martino C, et al. 336 The Effect of Implementing Electronic Health Record Default Prescribing Preferences on Opioid Prescriptions Written in the Emergency Department. Annals of Emergency Medicine 2018 Oct;72(4):S133. [doi: 10.1016/j.annemergmed.2018.08.341]
28. About Us: Thomas Jefferson University. 2019. URL: https://hospitals.jefferson.edu/about-us.html [accessed 2021-03-26]
29. Medication and Counseling Treatment. Rockville, MD: U.S. Department of Health and Human Services Substance Abuse and Metal Health Services Administration; 2019 May 7. URL: https://www.samhsa.gov/medication-assisted-treatment/treatment#medications-used-in-mat [accessed 2021-03-26]
30. Tian TY, Zlateva I, Anderson DR. Using electronic health records data to identify patients with chronic pain in a primary care setting. J Am Med Inform Assoc 2013 Dec;20(e2):e275-e280 [FREE Full text] [doi: 10.1136/amiajnl-2013-001856] [Medline: 23904323]
31. World Health Organization. International Classification of Diseases (9th Revision). Geneva, Switzerland: World Health Organization; 1978.
32. World Health Organization. The ICD-10 classification of mental and behavioural disorders: Clinical descriptions and diagnostic guidelines. Geneva, Switzerland: World Health Organization; 1992.
33. Diagnosis Code Set General Equivalence Mappings: ICD-10-CM to ICD-9-CM and ICD-9-CM to ICD-10-CM 2009 Version: Documentation and User's Guide. Baltimore, MD: Centers for Medicare and Medicaid Services; 2009. URL: https://www.cms.gov/Medicare/Coding/ICD10/index [accessed 2021-03-26]
34. Wang WLY, Yan J. touch: Tools of Utilization and Cost in Healthcare. 2018. URL: https://CRAN.R-project.org/package=touch [accessed 2021-03-26]
35. Slovis BH, Kairys J, Babula B, Girondo M, Martino C, Roke LM, et al. Discrepancies in Written Versus Calculated Durations in Opioid Prescriptions: Pre-Post Study. JMIR Med Inform 2020 Mar 31;8(3):e16199 [FREE Full text] [doi: 10.2196/16199] [Medline: 32229472]
36. Drug Enforcement Administration, Department of Justice. Schedule of controlled substances: placement of tramadol into schedule IV. Final rule. Fed Regist 2014 Jul 02;79(127):37623-37630 [FREE Full text] [Medline: 25016619]
37. Drug Enforcement Administration, Department of Justice. Schedules of controlled substances: rescheduling of hydrocodone combination products from schedule III to schedule II. Final rule. Fed Regist 2014 Aug 22;79(163):49661-49682 [FREE Full text] [Medline: 25167591]
38. Drug Overdose Mortality by State. Hyattsville, MD: National Center for Health Statistics; 2019. URL: https://www.cdc.gov/nchs/pressroom/sosmap/drug_poisoning_mortality/drug_poisoning.htm [accessed 2021-03-26]
39. Estimated Pennsylvania Drug Overdose Deaths by County, 2017-2018. URL: https://data.pa.gov/stories/s/9q45-nckt/ [accessed 2021-03-26]
40. DEA Intelligence Report: Opioid Threat in Pennsylvania: Drug Enforcement Administration. 2018. URL: https://www.hsdl.org/?view&did=818567 [accessed 2021-03-26]
41. World Health Organization. Information sheet on opioid overdose. Geneva, Switzerland: World Health Organization; 2018. URL: https://www.who.int/substance_abuse/information-sheet/en/ [accessed 2021-03-26]
42. U.S. Opioid Prescribing Rate Maps. Atlanta, GA: Centers for Disease Control and Prevention; 2018. URL: https://www.cdc.gov/drugoverdose/maps/rxrate-maps.html [accessed 2021-03-26]
43. Centers for Disease Control and Prevention. Mission Statement. Atlanta, GA: Centers for Disease Control and Prevention URL: https://www.cdc.gov/about/organization/mission.htm [accessed 2021-03-26]
44. Acute pain phase prescribing recommendations. URL: https://mn.gov/dhs/opip/opioid-guidelines/pain-phase/acute-pain.jsp [accessed 2021-03-26]

Abbreviations

CDC: Centers for Disease Control and Prevention
CPOE: computerized provider order entry
ED: emergency department
EHR: electronic health record
ICD-9: International Classification of Diseases, 9th Revision
MAT: medical assisted treatment
MME: morphine milligram equivalents

Edited by R Kukafka; submitted 16.09.20; peer-reviewed by AM Kaihlanen, A Almoudaris; comments to author 17.11.20; revised version received 14.12.20; accepted 16.03.21; published 14.04.21.

Please cite as:
Slovis BH, Riggio JM, Girondo M, Martino C, Babula B, Roke LM, Kairys JC
Reduction in Hospital System Opioid Prescribing for Acute Pain Through Default Prescription Preference Settings: Pre–Post Study
J Med Internet Res 2021;23(4):e24360
URL: https://www.jmir.org/2021/4/e24360
doi:10.2196/24360
PMID:33851922

Original Paper

Association of Electronic Health Record Vendors With Hospital Financial and Quality Performance: Retrospective Data Analysis

Bradley Beauvais[1*], MA, MBA, PhD; Clemens Scott Kruse[1*], MBA, MHA, MSIT, PhD; Lawrence Fulton[1*], MHA, MS, PhD; Ramalingam Shanmugam[1*], PhD; Zo Ramamonjiarivelo[1*], MBA, PhD; Matthew Brooks[1*], MPH, PhD

School of Health Administration, College of Health Professions, Texas State University, San Marcos, TX, United States

[*]all authors contributed equally

Corresponding Author:
Clemens Scott Kruse, MBA, MHA, MSIT, PhD
School of Health Administration
College of Health Professions
Texas State University
601 University Dr
San Marcos, TX, 78666
United States
Phone: 1 2103554742
Email: scottkruse@txstate.edu

Abstract

Background: Electronic health records (EHRs) are a central feature of care delivery in acute care hospitals; however, the financial and quality outcomes associated with system performance remain unclear.

Objective: In this study, we aimed to evaluate the association between the top 3 EHR vendors and measures of hospital financial and quality performance.

Methods: This study evaluated 2667 hospitals with Cerner, Epic, or Meditech as their primary EHR and considered their performance with regard to net income, Hospital Value–Based Purchasing Total Performance Score (TPS), and the unweighted subdomains of efficiency and cost reduction; clinical care; patient- and caregiver-centered experience; and patient safety. We hypothesized that there would be a difference among the 3 vendors for each measure.

Results: None of the EHR systems were associated with a statistically significant financial relationship in our study. Epic was positively associated with TPS outcomes (R^2=23.6%; β=.0159, SE 0.0079; *P*=.04) and higher patient perceptions of quality (R^2=29.3%; β=.0292, SE 0.0099; *P*=.003) but was negatively associated with patient safety quality scores (R^2=24.3%; β=−.0221, SE 0.0102; *P*=.03). Cerner and Epic were positively associated with improved efficiency (R^2=31.9%; Cerner: β=.0330, SE 0.0135, *P*=.01; Epic: β=.0465, SE 0.0133, *P*<.001). Finally, all 3 vendors were associated with positive performance in the clinical care domain (Epic: β=.0388, SE 0.0122, *P*=.002; Cerner: β=.0283, SE 0.0124, *P*=.02; Meditech: β=.0273, SE 0.0123, *P*=.03) but with low explanatory power (R^2=4.2%).

Conclusions: The results of this study provide evidence of a difference in clinical outcome performance among the top 3 EHR vendors and may serve as supportive evidence for health care leaders to target future capital investments to improve health care delivery.

(J Med Internet Res 2021;23(4):e23961) doi:10.2196/23961

KEYWORDS

electronic health records; medical informatics; hospitals; delivery of health care; financial management; quality of health care; treatment outcome

Introduction

Background

In the first part of the 20th century, health care predominantly revolved around a single health care provider, diagnosing and treating patients within the confines of their office or in the patient's home, and patient medical histories were recorded on paper. However, as the US health care industry has progressed and modernized, the proliferation of information technology has accelerated. In 2013, US health information technology investment totaled US $2.8 billion; however, by 2017, it had reached a staggering US $7.1 billion [1]. However, some reports indicate that hospitals have struggled to remain profitable during this same time frame [2]. Many health care executives are questioning the return on investment in hospital technology and wondering whether their capital outlay will result in improved financial outcomes [3].

Numerous studies have noted the high expense of care delivery yet low health care information technology proliferation in the United States. Citing the advancement of similar technology in other developed countries, researchers have indicated that through increased investment in health information technology, the United States can lower overall health care spending and simultaneously improve quality of care and patient outcomes [4,5]. As a result, the federal government has been heavily involved in the modernization of health information technology. As early as 2004, President George W Bush called for computerized health records in his State of the Union address and offered a strategy to provide Americans access to electronic health records (EHRs); unfortunately, he did not earn sufficient funding to change provider behavior [6]. In 2009, President Barack Obama signed the Health Information Technology for Economic and Clinical Health (HITECH) Act as part of the American Recovery and Reinvestment Act (ARRA) and originally set aside US $27 billion for an incentive program that encouraged hospitals and providers to adopt EHR systems [7,8]. This legislation prompted the adoption of EHRs in 2014 and established time frames for mandated EHR adoption. Furthermore, with the passage of the Medicare Access and CHIP Reauthorization Act, Medicare introduced the Medicare EHR Incentive Program, and the Merit-Based Incentive Payment System to advance the meaningful use of EHRs. As a result, by 2015, 96% of hospitals and 87% of physician practices implemented EHRs [9]. By mid-2016, the total federal government investment in EHR rose to US $35 billion and continued to rise [10].

One could argue that the proliferation of health care technology spending did not occur organically. The hospital and health care industry shifted to the use of EHR systems primarily because of the financial incentives incorporated within the ARRA and HITECH legislation [11]. HITECH provided eligible professionals who demonstrated the meaningful use of an EHR qualified for payments of US $18,000 in the first year; US $12,000 for the second year; US $8000 for the third year; US $4000 for the fourth year; and US $2000 for the fifth year [12]. An eligible professional was generally considered to be a physician. After 2015, physicians who failed to meaningfully use EHRs were subject to reductions in Medicare and Medicaid reimbursement. Meaningful use of an EHR includes 3 components: (1) the EHR must be certified and include e-prescribing capabilities; (2) the technology must provide for the electronic exchange of personal health information with other EHR systems (interoperability); and (3) the system must produce reports utilizing various clinical and quality metrics.

Beginning in 2011, incentive payments were also available for eligible hospitals that showed meaningful use of EHR and that submitted quality metrics based on criteria identified by the US Department of Health and Human Services. Incentive amounts were phased out in 2015 for hospitals that had not implemented a meaningful EHR. In addition, incentive payments were not available for hospitals that were not meaningful EHR users [13]. Despite the robust startup incentives offered by the federal government and continuing support provided via increased Medicare reimbursement, questions remain if the investment in health care information technology is sustainable. The adoption of a comprehensive EHR system can surpass several billion dollars for a large health care system [14]. The sustainability of these systems can approach several hundred million dollars annually, and numerous health care systems report significant implementation and sustainment cost overages [15,16]. Furthermore, despite the prevalent adoption of EHR systems since the passage of the HITECH Act, sharing of health care data and interoperability of information technology remains to be elusive. There remains to be little financial incentive to share and use data to reduce costs or improve the quality of care [17].

Thus, we seek to assess the association between investment in information technology and the dominant EHR platforms each have on the financial and quality outcomes of hospitals in the United States. We hypothesize that there is a difference in outcomes among the 3 vendors for each measure. Although extensive research has focused on the perceived and actual benefits of information technology in health care, this is an area of research that has not been fully evaluated. The body of literature predating the passage of the HITECH and Affordable Care Acts is fairly robust; however, with the passage of these legislative acts and the rate of change in technology, many of these studies are now outdated, particularly with respect to the direct impact of specific EHR vendors [18-20].

Literature Review

Two recent studies examined the relationships between health information technology capital expenditure and both financial and quality outcomes and aligned very closely with our work. Wang et al [21] examined the impact of investment in health information technology on hospital financial performance and productivity. In a later study, Wang and Gibbs [22] offered a framework to compare the performance of EHR systems. We intend to build on these studies in a few key areas [21,22].

First, in the 2018 and 2019 studies, the key financial outcome variable examined was return on assets (ROA). Net revenue per staffed bed was also considered in the 2018 study. Broadly speaking, financial performance can be assessed in 4 main areas: (1) profitability or return on investment, (2) liquidity, (3) leverage, and (4) operating efficiency. Within each category,

there are several variables to consider. Although ROA and revenue per bed are important, we believe that greater operational clarity can be achieved with a specific focus on the income statement and net income.

Second, quality was evaluated in the 2019 study via the Hospital Value–Based Purchasing (HVBP) Total Performance Score (TPS). These quality measures are all important to the field; however, we view the level of utilization of beds as a suboptimal measure of quality that is not well supported in the literature. The number of times a bed turns over speaks more to the volume of services demanded but offers little insight with respect to the quality of those services. We intend to focus more on the dimensions of HVBP and its subdomains. The contributory factors that are evaluated to produce the TPS include patient perceptions of care and measures of patient safety, process of clinical care delivery, and efficiency and cost reduction.

Third, both the 2018 and 2019 studies included important market characteristic variables as controls. These included hospital size, market concentration index (MCI), payer mix (Medicare and Medicaid; ie, the percentage of revenue coming from each of those programs), uncompensated care cost, ownership (governmental, proprietary, and nonprofit), teaching status, geographic classification, and year fixed effects. We suggest that additional variables may be insightful, as several other factors have been shown to influence both financial and quality outcomes [23-25]. Additional control variables worth considering are the level of outpatient services rendered, urban versus rural location, average length of stay, case mix index, wage index, sole community provider status, system membership, and geographic region [26]. These variables can further clarify the strength of the association between the independent variable of interest and our targeted dependent variables, while also serving to diminish any possible omitted variable bias.

Fourth, the 2019 study considers various EHR systems but does not clearly identify which vendors perform better or worse based on the available data. To the authors' credit, they did not want the research to be construed as the promotion of a particular vendor. However, in our view, the health care industry is in an era of evidence-based medicine and management. Thus, we believe that it is appropriate to identify the system in question. This more transparent approach, coupled with more specific outcome data that clearly identifies practitioner actionable evidence, should provide greater practical insight and facilitate improved organizational decision making.

In the following sections, we integrate and broaden the investigation of previous research efforts and evaluate the impact of 3 of the largest EHR providers' performance on measures of finance and quality. Quite simply, we would like to determine which EHR system performs the best and seek to provide health care leaders with an additional evidentiary basis for making EHR adoption decisions. Given the variation in EHR system cost, options, ease of use, training requirements, and on-site and follow-up support, we recognize that this can be a highly complex decision. Although we hypothesize that there is a difference in performance among the 3 EHR vendors in terms of financial and quality performance, we are uncertain a priori where each system will perform the best on the evaluation measures we have selected.

Methods

Data and Sample

The data for this study were extracted from two primary sources: the Definitive Health Care database and the American Hospital Association (AHA) Annual Survey database for 2018. The Definitive Health Care database provided the dependent and independent variables of interest, in addition to most of the control variables for this study. The Definitive Health Care database compiles US hospital data sources including Medicare Cost Reports, commercial claims data, Medicare Standard Analytics Files, Centers for Medicare and Medicaid Services (CMS) Hospital Compare, and many other data elements [27]. The cost report contains provider information such as facility characteristics, utilization data, cost and charges by cost center (in total and for Medicare), Medicare settlement data, and financial statement data. The AHA Annual Survey database provided the remaining data for the geographic region control variables [28]. All variables were linked with the 2 contributing data sources based on the Medicare provider number. Data on a total of 2667 short-term acute care hospitals were accumulated for analysis.

Measures—Dependent Variables

Table 1 shows the full complement of the study variables. Our study included 2 types of dependent variables drawn from the Definitive Health Care data set. The first set of data comes from the hospital income statement: net income scaled in millions of dollars. The second set of dependent variables is drawn from the hospitals' 2018 value-based purchasing scores and includes (1) the TPS, (2) patient experience score, (3) clinical process score, (4) efficiency score, and (5) the safety score. Each hospital's TPS is a weighted measure of performance based on each of the other areas listed, while each subordinate measure is an aggregation of several commonly tracked clinical and administrative criteria, as shown in Figure 1.

Under the CMS HVBP Program, Medicare makes incentive payments to hospitals based on how well they perform on each measure compared with other hospitals' performance during a baseline period and how much they improve their performance on each measure compared with the baseline reporting period [29]. All value-based purchasing variables are unweighted and based on a scale of 0 to 100, with higher scores being better, and have been validated by CMS for validity and reliability [30]. The total number of value-based purchasing participating hospitals provided a size limit to the study. A missingness map using Amelia, a program for missing data developed by Honaker et al [31], revealed approximately 1% missing from 2667 observations and 32 variables (k). The maximum proportion missing from any column was 12.5% and from any row was 10.1%; therefore, these values were conservatively imputed with the median using R Statistical Software [32].

Table 1. Variables and operational definitions.

Variable	Original source	Definition
Cerner	Definitive Health Care	Hospital using the Cerner EHR[a] as its primary EHR platform
Epic	Definitive Health Care	Hospital using the Epic EHR as its primary EHR platform
Meditech	Definitive Health Care	Hospital using the Meditech EHR as its primary EHR platform
TPS[b]	CMS[c]	The TPS is derived from 4 equally weighted domains in financial year 2018: • Clinical care • Patient experience of care • Safety • Efficiency and cost reduction
Patient experience score	CMS	Composite of 9 measures extracted from the hospital consumer assessment of health care providers and systems survey
Clinical care score	CMS	Composite of 3 mortality measures: acute myocardial infarction, heart failure, and pneumonia
Efficiency and cost reduction score	CMS	Medicare Spending Per Beneficiary
Safety score	CMS	Composite of 7 safety related rates: catheter-associated urinary tract infections, central line-associated blood stream infection, clostridium difficile infection, methicillin-resistant staphylococcus aureus, patient safety for selected indicators composite, elective delivery before 39 completed weeks gestation, and surgical site infections
For-profit status	Definitive Health Care	Hospitals operated by investor-owned organizations
Number of beds	Definitive Health Care	Number of staffed beds
Rural status	Definitive Health Care	Hospital located in a nonmetropolitan county or a hospital within a metropolitan county that is far away from the urban center, as defined by the Health Resource Services Administration
Government status	Definitive Health Care	Hospitals operated by local, county, or state government
Teaching status	Definitive Health Care	Hospitals affiliated with universities, colleges, medical schools, or nursing schools
Outpatient service mix	Definitive Health Care	Percent of care delivered in an outpatient setting
Average length of stay	Definitive Health Care	The average number of days that patients spend in hospital, measured by dividing the total number of days stayed by all inpatients during a year by the number of admissions or discharges
Case mix	Definitive Health Care	The case mix index is the average relative diagnosis related group weight of a hospital's inpatient discharges, calculated by summing the Medicare Severity-Diagnosis Related Group weight for each discharge and dividing the total by the number of discharges
Government payer mix	Definitive Health Care	The proportion of hospital reimbursement from governmental sources (Medicare, Medicaid, TRICARE, etc)
Wage Index	Definitive Health Care	A labor market area's wage index value is the ratio of the area's average hourly wage to the national average hourly wage
Sole community hospital	Definitive Health Care	A sole community hospital classified by specific criteria from CMS (distance from other like hospitals, rural, travel time, number of beds, etc)
System member	Definitive Health Care	An entity that owns or has owned 2 or more hospitals. In addition, health systems may also maintain ownership of other postacute or ambulatory sites of care
Market concentration	Definitive Health Care	The Herfindahl-Hirschman Index measure of market concentration was used to determine market competitiveness. It is calculated by squaring the market share of each firm competing in a market and then summing the resulting numbers
Average age of facility	Definitive Health Care	Average age of facility is calculated using the accumulated depreciation (total depreciation) and the depreciation expense (depreciation over a single period)
Occupancy rate	Definitive Health Care	Measure of utilization calculated as (inpatient days of care or bed days available)×100

Variable	Original source	Definition
Region	American Hospital Association	Regions of the United States as defined by the American Hospital Association

[a]EHR: electronic health record.

[b]TPS: total performance score.

[c]CMS: Centers for Medicare and Medicaid Services.

Figure 1. Financial year 2018 Hospital Value–Based Purchasing program measures.

Total Performance Score Domain	Measure Description
Safety	Catheter-Associated Urinary Tract Infections
Safety	Central Line Associated Blood Stream Infection
Safety	Clostridium difficile infection (C. difficile)
Safety	Methicillin-Resistant Staphylococcus aureus bacteremia
Safety	Patient Safety for Selected Indicators (composite)
Safety	Elective Surgery Priot to 39 Completed Weeks Gestation
Safety	Surgical Site Infection
Clinical care	Acute Myocardial Infarction (AMI) 30-day mortality rate
Clinical care	Heart Failure (HF) 30-day mortality rate
Clinical care	Pneumonia (PN) 30-day mortality rate
Efficiency and cost reduction	Medicare Spending per Beneficiary (MSPB)
Experience of care	Hospital Consumer Assessemnt of Healthcare Providers and Systems Survey

Measures—Independent and Control Variables

Our independent variables of interest included the top EHR systems used by hospitals in the United States (ie, Cerner, Epic, Meditech, or other), as reported in the Definitive Health Care database. These variables were included in our analysis as a dichotomous variable for each EHR system of interest ("1" if the system was used, "0" if not). Consistent with previous research, we also included several organizational-level control variables to account for other explanatory factors that could influence financial and quality outcomes. These variables included for-profit ownership status, number of beds, rural or urban geographic location, government ownership, teaching status, outpatient service mix, average length of stay, case mix, government payer percentage, wage index, sole community provider designation, system membership, MCI, occupancy rate, and geographic location by the AHA region. All analyses were performed using collinearity diagnostics. The variance inflation factor exceeded 10 in our analyses. Table 1 shows the data sources and operational definitions of the variables used in this study.

Results

Overview

Descriptive statistics and pairwise correlations were also calculated. The distributions of most of the dependent variables (income, TPS, experience score, clinical score, and safety) were relatively normal; however, the HVBP efficiency score was skewed to the right. Box-Cox analysis of the variable suggested a negative square root transformation, but for interpretability, it was not transformed. All dependent variables were min-max scaled between 0 and 1 for easier interpretation as percentile scores (eg, income percentile). All statistical analyses were performed using the R Statistical Software [32]. In all the analyses, a two-tailed *P* value <.05 was considered statistically significant.

Table 2 provides detailed descriptive statistics for each variable. Participating hospitals had a mean net income of US $15.01 million (SD US $109.04 million); TPS mean of 37.26 (SD 11.16); patient experience mean of 33.36 (SD 18.07); clinical process score mean of 59.45 (SD 19.24); efficiency score mean of 19.39 (SD 24.75); and safety score mean of 53.01 (SD 17.74). In all cases, higher scores on the HVBP variables were better. EHR vendors are represented in the following proportions in our sample: Epic: 39.85% (1063/2667; SD 0.48); Cerner: 23.39% (624/2667; SD 0.42); Meditech: 19.98% (533/2667; SD 0.39); and "other": 16.78% (447/2667; SD 0.39).

Among the numerous organizational characteristics included in our analysis as control variables, we observe 18.00% (480/2667) of the hospitals in the study population are for-profit facilities (SD 0.38), 22.98% (613/2667) are in rural locations (SD 0.42), 45.97% (1226/2667) are teaching facilities (SD 0.49), 72.97% (1946/2667) are affiliated with a health care system (SD 0.45), and that the hospitals are widely distributed across each of the AHA geographic regions.

Table 2. Descriptive statistics.

Variable	Values, mean (SD)
Net income (in millions; US $)	15.01 (109.04)
Total performance score	37.26 (11.17)
Patient experience score (unweighted)	33.36 (18.07)
Clinical process score (unweighted)	59.45 (19.24)
Efficiency score (unweighted)	19.39 (24.75)
Safety score (unweighted)	53.01 (17.74)
EHR[a]-Cerner	0.23 (0.42)
EHR-Epic	0.40 (0.48)
EHR-Meditech	0.20 (0.40)
EHR-other	0.17 (0.39)
For-profit	0.18 (0.40)
Beds	214.75 (185.47)
Rural	0.23 (0.42)
Government	0.13 (0.34)
Teaching	0.47 (0.50)
Outpatient service mix	0.53 (0.15)
Average length of stay	4.30 (0.92)
Case mix index	1.61 (0.28)
Government payer percent	0.71 (0.11)
Wage index	1.00 (0.20)
Sole community provider	0.08 (0.28)
System member	0.73 (0.45)
Market concentration index	0.34 (0.330)
Occupancy rate	0.57 (0.17)
Average age of facility	12.95 (9.23)
Region 1[b] (Connecticut, Maine, New Hampshire, Rhode Island, and Vermont)	0.04 (0.20)
Region 2 (New Jersey, New York, and Pennsylvania)	0.12 (0.32)
Region 3 (Delaware, Kentucky, Maryland, North Carolina, Virginia, West Virginia, and Washington, DC)	0.08 (0.28)
Region 4 (Alabama, Florida, Georgia, Mississippi, South Carolina, Tennessee, and Puerto Rico)	0.17 (0.37)
Region 5 (Illinois, Michigan, Indiana, Ohio, and Wisconsin)	0.17 (0.37)
Region 6 (Iowa, Kansas, Minnesota, Missouri, Nebraska, North Dakota, South Dakota)	0.08 (0.27)
Region 7 (Arkansas, Louisiana, and Texas)	0.13 (0.35)
Region 8 (Arizona, Colorado, Idaho, Montana, New Mexico, Utah, and Wyoming)	0.07 (0.260)
Region 9 (Alaska, California, Hawaii, Nevada, Oregon, and Washington)	0.13 (0.34)

[a]EHR: electronic health record.

[b]The representative geographical region is American Hospital Association Region 1 (Connecticut, Maine, New Hampshire, Rhode Island, and Vermont).

Net Income

Table 3 reflects the results of our regression analyses of hospitals' utilization of the top 3 EHR vendors and the associated hospital financial performance as measured by net income. On the basis of our analysis of net income regressed on EHR vendors (R^2=10.6%), we see no significant results for any of the vendors, when compared with facilities that fall into the "other" category. Thus, we can say, on average, and when controlling for the numerous organizational factors as controls, none of the EHRs are associated with favorable or unfavorable financial outcomes as measured by net income.

Table 3 also shows additional significant variables in our analysis that are associated with hospital net income, including

the number of hospital beds (β=.0001, SE 0.0000; *P*<.001), hospital case mix (β=.0067, SE 0.0032; *P*=.04), hospital wage index (β=−.0200, SE 0.0059; *P*<.001), and several geographic variables. These findings indicate that with a point increase in hospital case mix, net income increases by 0.67%, and with each point increase in the hospital wage index, net income falls by 2%.

Table 3. Analysis results for net income and total performance score.

Variable	Net income (adjusted R^2=10.64%)			Total performance score (adjusted R^2=23.61%)		
	β	SE	Significance (*P* value)	β	SE	Significance (*P* value)
Intercept	.5591	0.0139	<.001	.4027	0.0487	<.001
Cerner	.0007	0.0023	—[a]	−.0021	0.0080	—
Epic	−.0024	0.0023	—	.0159	0.0079	.04
Meditech	.0018	0.0023	—	.0117	0.0079	—
For-profit	.0011	0.0021	—	−.0176	0.6203	.02
Beds	.0001	0.0000	<.001	−.0045	0.0019	<.001
Rural	−.0012	0.0024	—	.0556	0.0084	<.001
Government	−.0017	0.0022	—	−.0125	0.0076	—
Teaching	.0051	0.0016	—	.0129	0.0055	.02
Outpatient service mix	−.0092	0.0073	—	.2137	0.0254	<.001
Average length of stay	−.0015	0.0009	—	−.0215	0.0033	<.001
Case mix	.0067	0.0032	.04	.0254	0.0111	.02
Government payer percent	−.0017	0.0022	—	−.0125	0.0076	—
Wage index	−.0200	0.0059	<.001	.0523	0.0204	.01
Sole community provider	.0044	0.0027	—	.0345	0.0095	<.001
System member	.0002	0.0018	—	.0061	0.0061	—
Market concentration	.0014	0.0028	—	−.0380	0.0098	<.001
Average age of facility	.0000	0.0001	—	.0000	0.0003	—
Occupancy rate	.0042	0.0053	—	.0042	0.0183	—
Region 2[b] (New Jersey, New York, and Pennsylvania)	−.0071	0.0039	—	−.0166	0.0135	—
Region 3 (Delaware, Kentucky, Maryland, North Carolina, Virginia, West Virginia, and Washington, DC)	−.0077	0.0043	—	.0070	0.0151	—
Region 4 (Alabama, Florida, Georgia, Mississippi, South Carolina, Tennessee, and Puerto Rico)	−.0129	0.0042	<.01	−.0211	0.0146	—
Region 5 (Illinois, Michigan, Indiana, Ohio, and Wisconsin)	−.0060	0.0039	—	.0044	0.0135	—
Region 6 (Iowa, Kansas, Minnesota, Missouri, Nebraska, North Dakota, and South Dakota)	−.0086	0.0044	.05	.0200	0.0152	—
Region 7 (Arkansas, Louisiana, and Texas)	−.0132	0.0043	<.01	−.0187	0.0150	—
Region 8 (Arizona, Colorado, Idaho, Montana, New Mexico, Utah, and Wyoming)	−.0057	0.0045	—	−.0138	0.0155	—
Region 9 (Alaska, California, Hawaii, Nevada, Oregon, and Washington)	−.0218	0.0040	<.001	.0167	0.0140	—

[a]Not significant.

[b]Referent geographical region is Region 1 (Connecticut, Maine, New Hampshire, Rhode Island, and Vermont); referent electronic health record is "other."

Total Performance Score

Table 3 also shows the results of our regression analyses of the association of EHRs with HVBP quality measures. On the basis of our analysis of TPS regressed on EHR vendors (R^2=23.6%), Epic reflects only statistically significant results (β=.0159, SE 0.0079; *P*=.04). These results indicate that the Epic EHR is associated with a 1.6% higher performance score when compared with facilities that fall into the "other" category. Neither Meditech nor Cerner were associated with significant results.

Table 3 further indicates several control variables that are significantly associated with TPS performance score. These variables include for-profit ownership (β=−.0176, SE 0.6203; *P*=.02), number of hospital beds (β=−.0045, SE 0.0019; *P*<.001), rural designation (β=.0556, SE 0.0084; *P*<.001), teaching designation (β=.0129, SE 0.0055; *P*=.02), outpatient service mix (β=.2137, SE 0.0254; *P*<.001), average length of stay (β=−.0215, SE 0.0033; *P*<.001), case mix (β=.0254, SE 0.0111; *P*=.02), wage index (β=.0523, SE 0.0204; *P*=.01), sole community provider designation (β=.0345, SE 0.0095; *P*<.001), and the MCI (β=−.0380, SE 0.0098; *P*<.001).

Among other findings, these results imply, on average, for-profit facilities perform 1.7% lower on the TPS measure. In addition, with each day increase in average length of stay, we observed a 2.2% decrease in TPS, and with each point increase in market concentration, we see an associated 3.8% decrease in TPS performance. However, with each point increase in the case mix index, wage index, and outpatient service mix, we observed a 2.5%, 5.2%, and 21.4% increase in TPS outcomes, respectively. Teaching hospitals are also associated with a 1.2% higher level of performance than nonteaching facilities.

Efficiency Score

In Table 4, we also show the evaluation of efficiency performance scores regressed on EHR vendors (R^2=31.9%). Cerner (β=.0330, SE 0.0135; *P*=.01) and Epic (β=.0465, SE 0.0133; *P*<.001) were positively associated with improved efficiency quality scores approximately 3.3% higher (Cerner) and 4.7% higher (Epic) than hospitals in the "other" category. Meditech was not associated with any significant results.

Table 4 also indicates several variables that are significantly associated with hospital efficiency. These variables include the number of hospital beds (β=−.0001, SE 0.0000; *P*=.002), rural designation (β=.0723, SE 0.0142; *P*<.001), outpatient service mix (β=.4831, SE 0.0429; *P*<.001), average length of stay (β=−.0233, SE 0.0055; *P*<.001), case mix (β=−.0698, SE 0.0188; *P*<.001), government payer percent (β=−.2486, SE 0.0375; *P*<.001), wage index (β=.1145, SE 0.0344; *P*<.001), sole community provider designation (β=.0888, SE 0.0160; *P*<.001), occupancy rate (β=.0656, SE 0.0310; *P*=.03), and several regional variables. This implies, on average, there is a statistically significant 2.3% decrease in efficiency score with each day increase in average length of stay, a 6.9% decrease with each point increase in the case mix index, and a 24.9% decrease in efficiency is associated with each point increase in government payer percentage.

Table 4. Analysis results for efficiency score and patient experience score.

Variable	Efficiency score (adjusted R^2=31.98%)			Patient experience score (adjusted R^2=29.3%)		
	β	SE	Significance (*P* value)	β	SE	Significance (*P* value)
Intercept	.0012	0.0822	__[a]	.3441	0.0612	<.001
Cerner	.0330	0.0135	.01	−.0002	0.0100	—
Epic	.0465	0.0133	<.001	.0292	0.0099	.003
Meditech	.0161	0.0133	—	.0157	0.0099	—
For-Profit	−.0119	0.0125	—	−.0542	0.0093	<.001
Beds	−.0001	0.0000	.002	−.0001	0.0000	<.001
Rural	.0723	0.0142	<.001	.0517	0.0105	<.001
Government	−.0065	−0.0065	—	.0044	0.0095	—
Teaching	−.0050	0.0093	—	.0270	0.0069	<.001
Outpatient service mix	.4831	0.0429	<.001	.3091	0.0320	<.001
Average length of stay	−.0233	0.0055	<.001	−.0251	0.0041	<.001
Case mix	−.0698	0.0188	<.001	.0100	0.0140	<.001
Government payer percent	−.2486	0.0375	<.001	.0044	0.0095	<.001
Wage index	.1145	0.0344	<.001	−.0744	0.0256	—
Sole community provider	.0888	0.0160	<.001	.0081	0.0119	—
System member	.0138	0.0103	—	−.0124	0.0077	—
Market concentration	.0158	0.0165	—	−.0586	0.0123	<.001
Average age of facility	.0004	0.0005	—	.0004	0.0003	—
Occupancy rate	.0656	0.0310	.03	−.0618	0.0230	.007
Region 2[b] (New Jersey, New York, and Pennsylvania)	.1062	0.0227	<.001	−.0649	0.0169	<.001
Region 3 (Delaware, Kentucky, Maryland, North Carolina, Virginia, West Virginia, and Washington, DC)	.1627	0.0254	<.001	−.0522	0.0189	.005
Region 4 (Alabama, Florida, Georgia, Mississippi, South Carolina, Tennessee, and Puerto Rico)	.0949	0.0247	<.001	−.0255	0.0184	—
Region 5 (Illinois, Michigan, Indiana, Ohio, and Wisconsin)	.0708	0.0228	.002	−.0254	0.0170	—
Region 6 (Iowa, Kansas, Minnesota, Missouri, Nebraska, North Dakota, and South Dakota)	.2219	0.0257	<.001	−.0472	0.0191	.01
Region 7 (Arkansas, Louisiana, and Texas)	.0481	0.0254	—	−.0066	0.0189	—
Region 8 (Arizona, Colorado, Idaho, Montana, New Mexico, Utah, and Wyoming)	.1702	0.0262	<.001	−.0977	0.0195	<.001
Region 9 (Alaska, California, Hawaii, Nevada, Oregon, and Washington)	.2693	0.0237	<.001	−.0740	0.0176	<.001

[a]Not significant.

[b]Referent geographical region is Region 1 (Connecticut, Maine, New Hampshire, Rhode Island, and Vermont); referent electronic health record is "other."

Patient Experience Score

Table 4 provides the final analysis results of our evaluation of hospitals' patient experience performance scores regressed on EHR vendors. Epic was positively associated with higher patient perceptions of quality scores 2.9% higher than hospitals in the "other" category (R^2=29.3%; β=.0292, SE 0.0099; *P*=.003).

Table 4 provides additional insight pertaining to the significant association between the control variables included in our study and patient experience scores. These variables include for-profit ownership (β=−.0542, SE 0.0093; *P*<.001), number of beds (β=−.0001, SE 0.0000; *P*<.001), rural status (β=.0517, SE 0.0105; *P*<.001), teaching (β=.0270, SE 0.0069; *P*<.001), outpatient service mix (β=.3091, SE 0.0320; *P*<.001), average length of stay (β=−.0251, SE 0.0041; *P*<.001), case mix

(β=.0100, SE 0.0140; *P*<.001), government payer percent (β=.0044, SE 0.0095; *P*<.001), market concentration (β=−.0586, SE 0.0123; *P*<.001), occupancy rate (β=−.0618, SE 0.0230; *P*=.007), and several geographic regions.

These results imply that, on average, the for-profit hospitals in our study scored 5.4% lower on the HVBP patient experience scores. In addition, for each additional day in the hospital, the patient experience scores decreased by 2.5%. Each point increase in market concentration and occupancy rate also reduces patient experience by 5.9% and 6.2%, respectively. Conversely, rural and teaching hospitals are associated with higher patient experience, with associated increased scores of 5.2% and 2.7%, respectively.

Patient Safety Score

Table 5 provides insight into our research on patient safety performance scores regressed on EHR vendors (R^2=24.3%). Epic (β=−.0221, SE 0.0102; *P*=.03) was negatively associated with patient safety quality scores of 2.2% lower than hospitals in the "other" category. Meditech and Cerner scores were not associated with significant results.

Table 5 also provides details regarding the significant associations between the control variables included in our study and patient safety scores. These variables include the number of hospital beds (β=−.0002, SE 0.0000; *P*<.001), rural status (β=.0420, SE 0.0109; *P*<.001), teaching (β=.0168, SE 0.0071; *P*=.02), outpatient service mix (β=.0965, SE 0.0330; *P*=.003), average length of stay (β=−.0143, SE 0.0042; *P*<.001), case mix (β=−.0812, SE 0.0144; *P*<.001), government payer percent (β=.0746, SE 0.0288; *P*=.009), and occupancy rate (β=−.0919, SE 0.0238; *P*<.001). These results indicate that patient safety is negatively impacted by 1.4% for every day increase in average length of stay and also declines by 8.1% for every point increase in the case mix index. Furthermore, with each percent increase in the hospital occupancy rate, patient safety scores declined by 9.2%. Conversely, patient safety scores were positively associated with rural and teaching hospitals by 4.2% and 1.7%, respectively. In addition, with each percentage increase in government payments, patient safety scores improved by 7.5%.

Table 5. Analysis results for patient safety score and clinical process score.

Variable	Patient safety score (adjusted R^2=24.35%)			Clinical process score (adjusted R^2=4.19%)		
	β	SE	Significance (*P* value)	β	SE	Significance (*P* value)
Intercept	.8251	0.0632	<.001	.5974	0.0758	<.001
Cerner	−.0182	0.0104	—[a]	.0284	0.0124	.02
Epic	−.0221	0.0102	.03	.0389	0.0122	.002
Meditech	−.0004	0.0103	—	.0274	0.0123	.03
For-profit	.0022	0.0096	—	.0512	0.0116	<.001
Beds	−.0002	0.0000	<.001	−.0001	0.0000	.02
Rural	.0420	0.0109	<.001	.0028	0.0131	—
Government	−.0083	0.0098	—	−.0264	0.0118	.03
Teaching	.0168	0.0071	.02	.0167	0.0086	.05
Outpatient service mix	.0965	0.0330	.003	.0182	0.0396	—
Average length of stay	−.0143	0.0042	<.001	−.0094	0.0051	—
Case mix	−.0812	0.0144	<.001	.0223	0.0173	—
Government payer percent	.0746	0.0288	.009	−.0839	0.0346	.02
Wage index	−.0314	0.0264	—	.0013	0.0317	—
Sole community provider	.0237	0.0123	—	.0146	0.0148	—
System member	.0082	0.0079	—	.0329	0.0095	<.001
Market concentration	−.0174	0.0127	—	.0323	0.0152	.03
Average age of facility	−.0003	0.0004	—	.0009	0.0004	.04
Occupancy rate	−.0919	0.0238	<.001	−.0057	0.0286	—
Region 2[b] (New Jersey, New York, and Pennsylvania)	.0051	0.0174	—	−.0014	0.0210	—
Region 3 (Delaware, Kentucky, Maryland, North Carolina, Virginia, West Virginia, and Washington, DC)	.0150	0.0196	—	.0212	0.0235	—
Region 4 (Alabama, Florida, Georgia, Mississippi, South Carolina, Tennessee, and Puerto Rico)	−.0056	0.0190	—	.0156	0.0228	—
Region 5 (Illinois, Michigan, Indiana, Ohio, and Wisconsin)	.0200	0.0175	—	.0275	0.0210	—
Region 6 (Iowa, Kansas, Minnesota, Missouri, Nebraska, North Dakota, and South Dakota)	−.0013	0.0197	—	−.0171	0.0237	—
Region 7 (Arkansas, Louisiana, and Texas)	.0055	0.0195	—	−.0126	0.0234	—
Region 8 (Arizona, Colorado, Idaho, Montana, New Mexico, Utah, and Wyoming)	−.0023	0.0201	—	−.0457	0.0242	—
Region 9 (Alaska, California, Hawaii, Nevada, Oregon, and Washington)	.0142	0.0182	—	.0181	0.0219	—

[a]Not significant.

[b]Referent geographical region is Region 1 (Connecticut, Maine, New Hampshire, Rhode Island, and Vermont); referent electronic health record is "other."

Clinical Care Performance Score

Finally, Table 5 shows the results of our evaluation of clinical care performance scores regressed on EHR vendors. All 3 vendors were associated with positive performance with Epic (β=.0388, SE 0.0122; *P*=.002), Cerner (β=.0283, SE 0.0124; *P*=.02), and Meditech (β=.0273, SE 0.0123; *P*=.03), reflecting positively associated higher clinical care performance scores between 2.7% (Meditech) and 3.8% (Epic) higher than hospitals in the "other" category. However, on this dependent variable,

we recognize that the explanatory power of the regressors is very low (R^2=4.2%).

Table 5 also provides insight into the association between the control variables in our study and clinical process outcomes. The statistically significant variables in our analysis included for-profit status (β=.0512, SE 0.0116; *P*<.001), number of hospital beds (β=−.0001, SE 0.0000; *P*=.02), government operated (β=−.0264, SE 0.0118; *P*=.03), teaching (β=.0167, SE 0.0086; *P*=.05), government payer percentage (β=−.0839, SE 0.0346; *P*=.02), system membership (β=.0329, SE 0.0095; *P*<.001), market concentration (β=.0323, SE 0.0152; *P*=.03), and the average age of the facility (β=.0009, SE 0.0004; *P*=.04). These results indicate that government-operated hospitals are associated with 2.6% lower clinical process scores, and for each point increase in government payer percentage, there is also an 8.3% lower score. However, for-profit, teaching, and system-owned hospitals appear to perform better on this measure by 5.1%, 1.6%, and 3.3%, respectively. Hospitals in concentrated markets also appear to perform better than those in less concentrated markets. With each point increase in market concentration, we observed an increase of 3.2%.

Discussion

Principal Findings

In general, our findings were insightful regarding the performance of individual EHR vendors. We did not expect to see a clearly obvious choice of EHR vendor with respect to performance as defined by our financial and quality-focused dependent variables. To this end, we did not ascertain that there is a single EHR that outperforms all other competitors across all of our study measures.

Our findings pertaining to financial outcomes were somewhat interesting in that no single EHR demonstrated a significant and positive association with net income. In most instances, the capital allocation process is predicated on reasonable assurance that there will be some tangible return on investment over a reasonable amount of time. As EHR adoption is a major capital investment and requires a major organizational change and extensive training, some have argued that it may take the organization a few years to see its effect on financial performance. For instance, Collum et al [20] found a statistically significant improvement in the total margin 2 years after EHR adoption in hospitals. However, the authors attributed the observed effect more to HITECH Act incentive payments than operational improvements, primarily because the authors found no significant association with operating margin. Thus, these previous authors' observations, coupled with our own, continue to indicate that EHR return on investment remains an unsettled matter.

In our evaluation of TPS as a dependent variable, we note that Epic was the singular EHR with a positive and significant association with improved scoring. Given that the TPS is a composite of the other variables in our analysis, it prompted us to examine each of the subdomains' performance scores more closely. Although we did not see a positive association for Epic in improved financial performance, this EHR recorded positive and significant associations in clinical care, patient experience, and efficiency scoring. Thus, when considered together, this combination of scores across these 3 quality subdomains appears to provide a performance advantage to Epic and a positive association with TPS scoring.

However, Epic also demonstrated a statistically significant and negative association with patient safety, which was not observed in our other vendors' performance. Although we did not expect to observe a significant and negative association between any EHR vendors and patient safety scoring, upon further research on this topic, we note that our findings appear to be consistent with several previous researchers' results. Bowman [33] captures these areas of concern very well in her synthesis of 64 studies and papers highlighting numerous areas where EHRs can impose undue burdens on health care providers and introduce the possibility of errors. These include the potential for recording erroneous data entry leading to patient safety hazards, system design flaws, improper system use, inappropriate document capture, erroneous application of copy and paste functions within the medical record, rigid application of prepopulated templates, and errors related to clinical decision support systems such as alert fatigue [33]. In recent years, others have pointed to potential problems with EHRs as a vector for increased risk to patient safety with respect to incorrect use [34,35], malfunctions [36,37], interoperability or system interaction [38], and health information technology blackouts or downtime [36,39]. On the basis of our findings, we can reasonably assume that many of these issues persist.

Our research group discussed the possible reasons for these results. In many ways, our results are supported by independent research. The top vendors that we studied are often highlighted in the KLAS, LLC Research for best in class, most user friendly, and holding buyers' attention [40]. It is a user-friendly variable that could contribute to the software's success in efficiency measures. It is possible that all 3 vendors are equally capable of achieving similar efficiency scores, but the fact that users are more familiar with their function and more willing to explore beyond the basic user training that renders Epic more effective in the areas of TPS. One could reasonably assume that the number of years since the hospital adopted the EHR system, and also the stage of adoption, might have had an effect. Hospitals that have adopted the EHR over several years may be more efficient than hospitals that only recently adopted or changed their EHR system. This could have an impact on levels of customer service, capability to integrate modules together, onboarding processes relating to the EHR, and the organization's capacity to facilitate initial and ongoing training.

Finally, another factor contributing to the success of one vendor over the others could be ownership. Epic and Meditech have proudly and defiantly maintained their private status. This factor could make a vendor more agile in its software development life cycle, enabling them to customize to order or correct flaws rapidly.

Practice Implications

Those involved in purchasing decisions surrounding EHR should carefully consider areas of focus in the facility, capital expenditure cycles, strategic direction, and willingness of the

organization to change EHR vendors. The significant decision to switch EHR vendors is a complex process, and many factors need to be aligned to set up the organization for success. Our research may provide an evidentiary basis for vendor selection. For example, an organization that struggles with improving clinical care performance might consider Epic or Meditech with an understanding that there are numerous other factors that might influence outcomes. A similar choice might be considered for facilities desiring to improve patient experience. However, if patient experience or efficiency—in terms of Medicare Spending Per Beneficiary—is an area of weakness, Epic might be a preferred choice. Ultimately, an organization using a vendor other than these 3 might look at the features that these vendors offer that their vendor does not. Is there something their current vendor could offer and increase measures of efficiency? Regrettably, our research does not extend to the module level of the EHR, so we cannot make any recommendations in that regard.

Limitations and Recommendations for Future Research

Our study had several limitations. First, this is a single year of data drawn from the 2018 data pertaining to performance within only short-term acute care HVBP participating facilities. Future studies should consider examining the growth or decline of EHR influence on these outcomes over time. Furthermore, as a single-year study, our analysis also does not account for any changes in EHR systems during the delay between baseline and performance reporting periods from which HVBP scores were determined, nor do we include the length of time the EHR has been in place within the hospitals studied. A more in-depth paired analysis could be considered to match the EHR system with the exact time frame of performance. Additional financial and quality outcome variables might also be considered, which could broaden the study unit of observations beyond the HVBP constraint. Finally, as more granular data becomes available pertaining to the specific modules in use at the hospital level, future studies might examine how specific module use is associated with specific clinical outcomes.

Beyond the extensive implementation of EHRs, health care providers, hospitals, and health care facilities invest heavily in other forms of information technology to provide and enhance care delivery. As the industry progresses toward value-based care, organizations are increasingly investing in imaging, telehealth, precision medicine, artificial intelligence, cloud-based computing or data storage, consumer-facing technologies, and disease management technologies. Furthermore, in the days since the start of the COVID-19 pandemic, telemedicine has been showcased as an indispensable capability of the EHR. Another aspect that should be evaluated in the future is the telemedicine capabilities of these vendors. Future research should carefully examine care delivered through this modality and compare the outcomes across the top vendors.

Conclusions

The return on investment and outcomes associated with EHRs have been a topic of intense focus and debate over the past 2 decades. Up to this point, a research gap has persisted pertaining to the study and transparent disclosure of comparative studies of major EHR vendors. In our analysis of the big 3 vendors—Epic, Cerner, and Meditech—we endeavored to fill that gap. Yet, we can see that clearly answering which system performs the best is complex. The implementation of any of these products can take years, and success is not guaranteed. However, our findings may provide some clarity to health care leaders seeking to develop an evidence base to support future capital investment in EHR systems. If an organization is already considering a switch to a new EHR vendor, the organization can devote sufficient funding for such an undertaking, and if the leadership is willing to lead such a large organizational change, then our study may provide some points of clarity pertaining to the big 3 vendors that might be apt for consideration.

Conflicts of Interest

None declared.

References

1. Morse S. US health IT investment skyrockets to $7.1 billion. Healthcare IT news. 2018. URL: https://www.healthcareitnews.com/news/us-health-it-investment-skyrockets-71-billion [accessed 2020-04-30]
2. Muchmore S. Hospital profitability down as operators lack flexibility to cut costs, Kaufman Hall says. Healthcare Dive. URL: https://www.healthcaredive.com/news/hospital-profitability-down-as-operators-lack-flexibility-to-cut-costs-kau/559705/ [accessed 2020-04-30]
3. Kacik A. Health systems weigh return on investment as they ramp up tech. Modern Healthcare. URL: https://www.modernhealthcare.com/operations/health-systems-weigh-return-investment-they-ramp-up-tech [accessed 2020-04-30]
4. Anderson GF, Frogner BK, Johns RA, Reinhardt UE. Health care spending and use of information technology in OECD countries. Health Aff (Millwood) 2006 May;25(3):819-831. [doi: 10.1377/hlthaff.25.3.819] [Medline: 16684749]
5. Alvarez K, Goldfarb NI. US health system performance: a national scorecard. Am J Med Qual. 2007. URL: https://go.gale.com/ps/anonymous?id=GALE%7CA157943700&sid=googleScholar&v=2.1&it=r&linkaccess=abs&issn=10628606&p=AONE&sw=w [accessed 2021-03-16]
6. Bush GW. Executive order 13335: incentives for the use of health information technology and establishing the position of the national health information technology coordinator. Federal Register. Washington D.C: Government Publishing Office; 2004. URL: http://edocket.access.gpo.gov/cfr_2005/janqtr/pdf/3CFR13335 [accessed 2020-04-30]

7. Schilling B. The federal government has put billions into promoting electronic health record use: how is it going? The Commonwealth Fund. The Commonwealth Fund.: The Commonwealth Fund; 2015. URL: https://www.commonwealthfund.org/publications/newsletter-article/federal-government-has-put-billions-promoting-electronic-health [accessed 2020-04-30]
8. Lukaszewski M. A history of health information technology and the future of interoperability. Bulletin of the American College of Surgeons.: American College of Surgeons; 2017. URL: https://bulletin.facs.org/2017/11/a-history-of-health-information-technology-and-the-future-of-interoperability/# [accessed 2020-04-30]
9. Henry J, Pylypchuk Y, Searcy T, Patel V. Adoption of electronic health record systems among non-federal acute care hospitals: 2008-2015. HealthIT.gov.: Department of Health and Human Services URL: https://dashboard.healthit.gov/evaluations/data-briefs/non-federal-acute-care-hospital-ehr-adoption-2008-2015.php [accessed 2020-04-30]
10. Arndt R. No end in sight: EHRs hit hospitals' bottom lines with uncertain benefits. Modern Healthcare. 2018. URL: https://www.modernhealthcare.com/article/20181013/NEWS/181019945/no-end-in-sight-ehrs-hit-hospitals-bottom-lines-with-uncertain-benefits [accessed 2020-04-30]
11. Adler-Milstein J, Jha AK. HITECH Act drove large gains in hospital electronic health record adoption. Health Aff (Millwood) 2017 Aug 01;36(8):1416-1422. [doi: 10.1377/hlthaff.2016.1651] [Medline: 28784734]
12. HITECH Act of 2009, 42 USC sec 139w-4(0)(2) sec 13301, subtitle B: incentives for the use of Health Information Technology (2009). Federal Register. Washinton, D.C: Government Publishing Office; 2009. URL: https://www.hhs.gov/sites/default/files/ocr/privacy/hipaa/understanding/coveredentities/hitechact.pdf [accessed 2020-04-29]
13. Hudock R, Wagner P. Analysis of the HITECH Act's incentives to facilitate adoption of Health Information Technology. Epstein Becker Green. 2009. URL: https://www.ebglaw.com/content/uploads/2014/06/28043_ClientAlertHITECH.pdf [accessed 2020-04-30]
14. Unpacking hospitals' EHR implementation costs: what's behind the million-dollar price tags? Beckers Health IT. 2016. URL: https://www.beckershospitalreview.com/healthcare-information-technology/unpacking-hospitals-ehr-implementation-costs-what-s-behind-the-million-dollar-price-tags.html [accessed 2020-04-30]
15. Cohen JK. 10 EHR implementations with the biggest price tags in 2017. Becker's Health IT. 2017. URL: https://www.beckershospitalreview.com/ehrs/10-ehr-implementations-with-the-biggest-price-tags-in-2017.html [accessed 2020-04-30]
16. Resneck J. Report of the board of trustees impact of high capital costs of hospital EHRs on the medical staff. 2018. URL: https://www.ama-assn.org/system/files/2019-04/a19-bot32.pdf [accessed 2020-04-30]
17. Marchibroda J. Health Policy Brief. Interoperability. Health Affairs. 2014. URL: https://www.healthaffairs.org/do/10.1377/hpb20140811.761828/full/healthpolicybrief_122.pdf [accessed 2020-04-30]
18. Schmitt KF, Wofford DA. Financial analysis projects clear returns from electronic medical records. Healthc Financ Manage 2002 Jan;56(1):52-57. [Medline: 11806319]
19. Menachemi N, Burkhardt J, Shewchuk R, Burke D, Brooks RG. Hospital information technology and positive financial performance: a different approach to finding an ROI. J Healthc Manag 2006;51(1):40-58. [doi: 10.1097/00115514-200601000-00008]
20. Collum TH, Menachemi N, Sen B. Does electronic health record use improve hospital financial performance? Evidence from panel data. Health Care Manage Rev 2016;41(3):267-274. [doi: 10.1097/hmr.0000000000000068]
21. Wang T, Wang Y, McLeod A. Do health information technology investments impact hospital financial performance and productivity? Int J Account Inform Syst 2018 Mar;28:1-13. [doi: 10.1016/j.accinf.2017.12.002]
22. Wang T, Gibbs D. A framework for performance comparison among major electronic health record systems. Perspect Health Inf Manag 2019;16(Fall):1h [FREE Full text] [Medline: 31908631]
23. Beauvais B, Richter JP, Kim FS, Palmer EL, Spear BL, Turner RC. A reason to renovate: the association between hospital age of plant and value-based purchasing performance. Health Care Manage Rev 2018 Oct 31;46(1):66-74. [doi: 10.1097/hmr.0000000000000227]
24. Beauvais B, Richter JP, Kim FS. Doing well by doing good: evaluating the influence of patient safety performance on hospital financial outcomes. Health Care Manage Rev 2019;44(1):2-9. [doi: 10.1097/HMR.0000000000000163] [Medline: 28445325]
25. Beauvais B, Richter JP, Kim FS, Sickels G, Hook T, Kiley S, et al. Does patient safety pay? Evaluating the association between surgical care improvement project performance and hospital profitability. J Healthc Manag 2019;64(3):142-154. [doi: 10.1097/JHM-D-17-00208] [Medline: 31999263]
26. Kruse CS, DeShazo J, Kim F, Fulton L. Factors associated with adoption of health information technology: a conceptual model based on a systematic review. JMIR Med Inform 2014 May 23;2(1):e9 [FREE Full text] [doi: 10.2196/medinform.3106] [Medline: 25599673]
27. Definitive Healthcare (2020). URL: https://www.defhc.com/home [accessed 2020-04-30]
28. American Hospital Association. 2020. URL: https://www.ahadata.com/ [accessed 2020-04-30]
29. Hospital value based purchasing. Centers for Medicare & Medicaid Services. 2017. URL: https://www.cms.gov/Outreach-and-Education/Medicare-Learning-Network-MLN/MLNProducts/downloads/Hospital_VBPurchasing_Fact_Sheet_ICN907664.pdf [accessed 2020-04-30]

30. Medicare program; hospital inpatient value-based purchasing program. Final rule. Fed Regist 2011 May 06;76(88):26490-26547 [FREE Full text] [Medline: 21548401]
31. Honaker J, King G, Blackwell M. Amelia II: A program for missing data. J Stat Softw. 2011. URL: https://www.jstatsoft.org/article/view/v045i07 [accessed 2021-03-18]
32. R Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2017. URL: https://www.R-project.org/ [accessed 2020-04-30]
33. Bowman S. Impact of electronic health record systems on information integrity: quality and safety implications. Perspect Health Inf Manag 2013;10:1c [FREE Full text] [Medline: 24159271]
34. Tsou AY, Lehmann CU, Michel J, Solomon R, Possanza L, Gandhi T. Safe practices for copy and paste in the EHR. Appl Clin Inform 2017 Dec 20;26(01):12-34. [doi: 10.4338/aci-2016-09-r-0150]
35. Taieb-Maimon M, Plaisant C, Hettinger AZ, Shneiderman B. Increasing recognition of wrong-patient errors through improved interface design of a computerized provider order entry system. Int J Hum Comput Interact 2017 Sep 05;34(5):383-398. [doi: 10.1080/10447318.2017.1349249]
36. Pradhan H, Stokes J. Does your electronic health record system introduce patient safety risks? Washington Patient Safety Coalition. 2015. URL: http://old.wapatientsafety.org/wp-content/uploads/2015/05/EHRFinal-White-PaperApril-2015.pdf [accessed 2020-04-30]
37. Wright A, Hickman TTT, McEvoy D, Aaron S, Ai A, Andersen JM, et al. Analysis of clinical decision support system malfunctions: a case series and survey. J Am Med Inform Assoc 2016 Nov;23(6):1068-1076 [FREE Full text] [doi: 10.1093/jamia/ocw005] [Medline: 27026616]
38. Adams KT, Howe JL, Fong A, Puthumana JS, Kellogg KM, Gaunt M, et al. An analysis of patient safety incident reports associated with electronic health record interoperability. Appl Clin Inform 2017 Dec 21;08(02):593-602. [doi: 10.4338/aci-2017-01-ra-0014]
39. Sax U, Lipprandt M, Röhrig R. The rising frequency of IT blackouts indicates the increasing relevance of IT emergency concepts to ensure patient safety. Yearb Med Inform 2018 Mar 06;25(01):130-137. [doi: 10.15265/iy-2016-038]
40. US Hospital EMR Market Share 2020. KLAS Research. 2020. URL: https://klasresearch.com/report/us-hospital-emr-market-share-2020/1616 [accessed 2020-07-08]

Abbreviations

AHA: American Hospital Association
ARRA: American Recovery and Reinvestment Act
CMS: Centers for Medicare and Medicaid Services
EHR: electronic health record
HITECH: Health Information Technology for Economic and Clinical Health
HVBP: Hospital Value–Based Purchasing
MCI: market concentration index
ROA: return on assets
TPS: total performance score

Edited by R Kukafka; submitted 29.08.20; peer-reviewed by A Vagelatos, KM Kuo; comments to author 18.09.20; revised version received 30.09.20; accepted 02.02.21; published 14.04.21.

Please cite as:
Beauvais B, Kruse CS, Fulton L, Shanmugam R, Ramamonjiarivelo Z, Brooks M
Association of Electronic Health Record Vendors With Hospital Financial and Quality Performance: Retrospective Data Analysis
J Med Internet Res 2021;23(4):e23961
URL: https://www.jmir.org/2021/4/e23961
doi:10.2196/23961
PMID:33851924

Original Paper

Effects of an mHealth App (Kencom) With Integrated Functions for Healthy Lifestyles on Physical Activity Levels and Cardiovascular Risk Biomarkers: Observational Study of 12,602 Users

Rikuta Hamaya[1,2], MD, MSc; Hiroshi Fukuda[3,4], MD, PhD; Masaki Takebayashi[5], PhD; Masaki Mori[6], PhD; Ryuji Matsushima[6], BA; Ken Nakano[6], MEng; Kuniaki Miyake[6], MD, PhD; Yoshiaki Tani[6], MSc; Hirohide Yokokawa[3], MD, PhD

[1]Division of Preventive Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, United States

[2]Department of Epidemiology, Harvard University, Boston, MA, United States

[3]Department of General Medicine, School of Medicine, Juntendo University, Tokyo, Japan

[4]Department of Advanced Preventive Medicine and Health Literacy, Graduate School of Medicine, Juntendo University, Tokyo, Japan

[5]Graduate School of Health Science, Aomori University of Health and Welfare, Aomori, Japan

[6]DeSC Healthcare Inc., Tokyo, Japan

Corresponding Author:
Rikuta Hamaya, MD, MSc
Division of Preventive Medicine
Department of Medicine
Brigham and Women's Hospital and Harvard Medical School
900 Commonwealth Avenue
Boston, MA, 02215
United States
Phone: 1 617 732 4965
Email: rktrocky@gmail.com

Abstract

Background: Mobile health (mHealth) apps are considered to be potentially powerful tools for improving lifestyles and preventing cardiovascular disease (CVD), although only few have undergone large, well-designed epidemiological research. "kencom" is a novel mHealth app with integrated functions for healthy lifestyles such as monitoring daily health/step data, providing tailored health information, or facilitating physical activity through group-based game events. The app is linked to large-scale Japanese insurance claims databases and annual health check-up databases, thus comprising a large longitudinal cohort.

Objective: We aimed to assess the effects of kencom on physical activity levels and CVD risk factors such as obesity, hypertension, dyslipidemia, and diabetes mellitus in a large population in Japan.

Methods: Daily step count, annual health check-up data, and insurance claim data of the kencom users were integrated within the kencom system. Step analysis was conducted by comparing the 1-year average daily step count before and after kencom registration. In the CVD risk analysis, changes in CVD biomarkers following kencom registration were evaluated among the users grouped into the quintile according to their change in step count.

Results: A total of 12,602 kencom users were included for the step analysis and 5473 for the CVD risk analysis. The participants were generally healthy and their mean age was 44.1 (SD 10.2) years. The daily step count significantly increased following kencom registration by a mean of 510 steps/day (P<.001). In particular, participation in "Arukatsu" events held twice a year within the app was associated with a remarkable increase in step counts. In the CVD risk analysis, the users of the highest quintile in daily step change had, compared with those of the lowest quartile, a significant reduction in weight (–0.92 kg, P<.001), low-density lipoprotein cholesterol (–2.78 mg/dL, P=.004), hemoglobin A_{1c} (HbA_{1c}; –0.04%, P=.004), and increase in high-density lipoprotein cholesterol (+1.91 mg/dL, P<.001) after adjustment of confounders.

Conclusions: The framework of kencom successfully integrated the Japanese health data from multiple data sources to generate a large, longitudinal data set. The use of the kencom app was significantly associated with enhanced physical activity, which might lead to weight loss and improvement in lipid profile.

(J Med Internet Res 2021;23(4):e21622) doi:10.2196/21622

KEYWORDS

mHealth; app; cardiovascular disease; physical activity; smartphone; mobile phone

Introduction

Reducing the burden of cardiovascular disease (CVD) is an urgent public health issue. The major aspect of primordial prevention is self-care behavior for the prevention of CVD, such as healthy diet and exercise [1]. Long-term success in lifestyle modification is multifactorial and needs cost-effective, innovative approaches. With the widespread use of smartphones in the last decade, numerous mobile health (mHealth) apps have been developed and considered as potentially powerful tools for this purpose because of their convenience and practicality [2]. The core functions of mHealth apps are self-monitoring, providing tailored information or feedback, and exercise goal setting/reviewing [3,4]. Several apps have undergone scientific evaluation for health effects, and meta-analyses have implicated the favorable effects of mHealth apps on physical activity [5], behavioral change [6], and multiple CVD risk factors [3]. However, previous studies generally targeted small populations and may also have serious selection bias, warranting the need for well-designed epidemiological research [3].

Furthermore, although mHealth apps may positively influence the user's lifestyle on a short-term basis, the attrition rates are commonly high in long-term condition management [7]. Apps must be optimized to increase the adherence to the app-guided healthy lifestyles in order to practically reduce the user's CVD risk factors. For example, avoidance of manual data entry increases app usage [8]. Providing immediate rewards in the form of incentives may be effective to motivate individuals to enhance physical activity [9]. A game-based design or positive peer influence may also be effective to increase user engagement, as is well shown in the Pokémon Go app [10]. To date, only few mHealth apps could demonstrate the positive impact of the user's objective health status as represented by CVD biomarkers, such as blood pressure (BP), lipid profile, and glucose metabolism.

'kencom' is an mHealth app with integrated functions that include providing tailored health information, automatically gathered annual health check-up results, self-monitoring, and feedback service to the users. It is also synchronized to smartphone pedometers to count daily steps. The app is optimized to keep a low attrition rate through the easy-to-use user interface, incentive system, and regularly held events. Most importantly, the app data can be linked to a large-scale Japanese insurance database with detailed prescription information and to Japanese annual health check-up data, thus making a large longitudinal data set. In this study, we aimed to assess the effects of kencom on physical activity levels and CVD risk factors such as obesity, hypertension, dyslipidemia, and diabetes mellitus in a large Japanese population. We then discussed the plausible mechanisms of kencom from the perspectives of behavioral economics.

Methods

Overview and Functions of Kencom

kencom is a service developed by DeSC Healthcare, Inc. that is available as an app on iOS and Android platforms as well as through a web service. The kencom app is free to use and over 90% of the active users (ie, those who logged into kencom at least one time a month) are app users. Individuals can use the app for free if they live in Japan, are at least 19 years of age, and have joined an affiliated society-managed, employment-based health insurance association, one of the major insurers of Japanese universal health coverage. The affiliated society-managed, employment-based health insurance associations pay a subscription fee to DeSC Healthcare Inc. There are 5 core functions of kencom (Textbox 1; also see Figure 1).

Textbox 1. Core functions of kencom.

1. **Displaying annual health check-up data**

   Users can check their own longitudinal, detailed health data based on the government-led annual health check-ups without manual input (Figure 1A).

2. **Monitoring daily health data and goal setting**

   Steps are counted by each smartphone's built-in pedometer and the data can be synchronized to the kencom app with the user's agreement. Users may set the daily goals of physical activity and the kencom app gives feedback to them according to self-checked achievements. Users may also manually input weight, blood pressure, and blood sugar levels in the app (Figure 1B).

3. **Providing tailored health information**

   Users receive tailored health information according to their lifestyle or disease risks, which is aimed to improve their health literacy. Every original article is peer-reviewed by medical doctors or nutrition professionals.

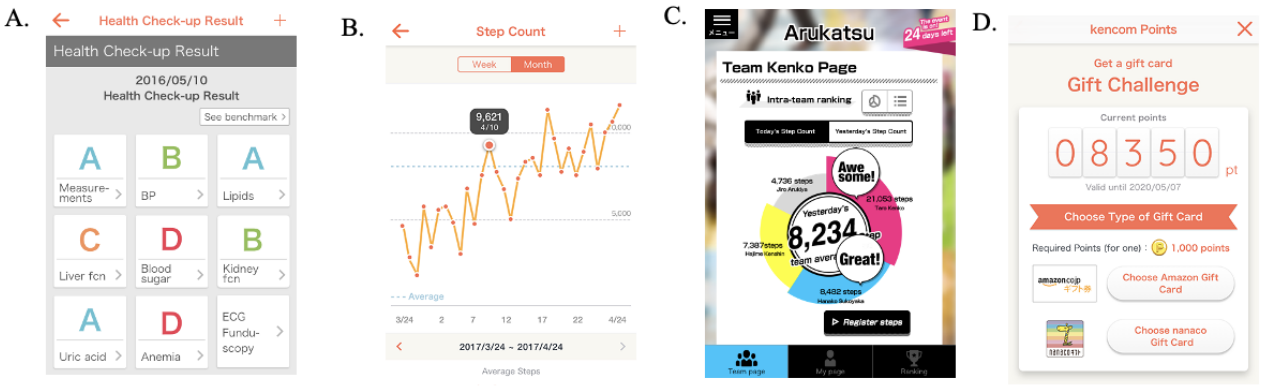
4. **Facilitating physical activity through team-based events**

   kencom regularly runs "Arukatsu" events to facilitate users' physical activity. Up to 10 users form a team and the teams compete with each other in the total step counts for 1 month. Individual step counts are shared with the team members, encouraging each member's physical activity. Game design elements were utilized in the event (Figure 1C).

5. **Incentive system**

   kencom points are awarded for daily logging-in and engaging in various services including Arukatsu. Points were not basically rewarded for daily activities or achieving goals. Users can get gift vouchers from kencom points that can be spent in the market place (Figure 1D).

Figure 1. Sample images of kencom app. A. Annual health check-up results; B. Trends of daily step counts; C. Arukatsu (a game event) team page where individual step counts are shared with the team members, encouraging each member's physical activity; D. Incentive system (kencom point).



Data

This cohort study was based on 3 data sources: Japanese health check-up database, Japanese health insurance claims database, and kencom database. These data were integrated, anonymized, and stored in the affiliated local society-managed, employment-based health insurance associations. DeSC Healthcare Inc. then combined data from different society-managed, employment-based health insurance associations, making it a large, longitudinal database for research purposes. Multimedia Appendix 1 summarizes the variables from each data source. In brief, the Japanese health check-up database consists of the results of questionnaires, physical examinations, measurement of biomarkers, and imaging examinations, which are conducted annually for the majority of adults living in Japan. Japanese health insurance claims database records monthly information about the patient demographics, diagnoses according to the International Classification of Diseases and Related Health Problems, 10th Revision (ICD-10), medical procedures, and medications. The kencom database is mainly about daily physical activity data and the app usage.

The data anonymization was conducted under the "opt-out agreement" between the users and the society-managed, employment-based health insurance associations, in which the users were notified of their data usage and they can propose the deletion of their data. The study complied with the International Society for Pharmacoepidemiology Guidelines for Good Pharmacoepidemiology Practices. This study design was approved by the institutional review board of Juntendo University (approval number 2020045).

Study Population

This analysis targeted previously healthy adults living in Japan. The inclusion criteria were the registration of kencom between April 2016 and June 2018, and the availability of the baseline insurance claims data, health check-up data, and step data before and after kencom registration (12 months each) for the step analysis; and further availability of the follow-up annual health check-ups data for the CVD risk analysis. The baseline health

check-up was defined as the check-up within 1 year before the kencom registration, and the follow-up corresponds to the first check-up conducted after 1-year usage of kencom (Figure 2). These criteria were applied only to iOS users in whom the step data prior to the app installation were automatically transferred into the app. The exclusion criteria included previous medical history of any cancer, end-stage renal disease requiring hemodialysis, and those with too high variance in baseline steps (top one percentile SD: over 7680 steps/day). From 15,363 kencom users with baseline information, 12,602 users met the criteria for the step analysis and 5473 for the CVD risk analysis (Figure 3).

Figure 2. Data collection timelines.

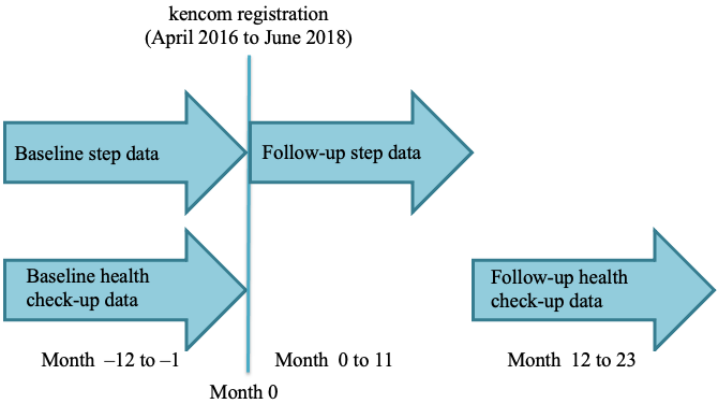
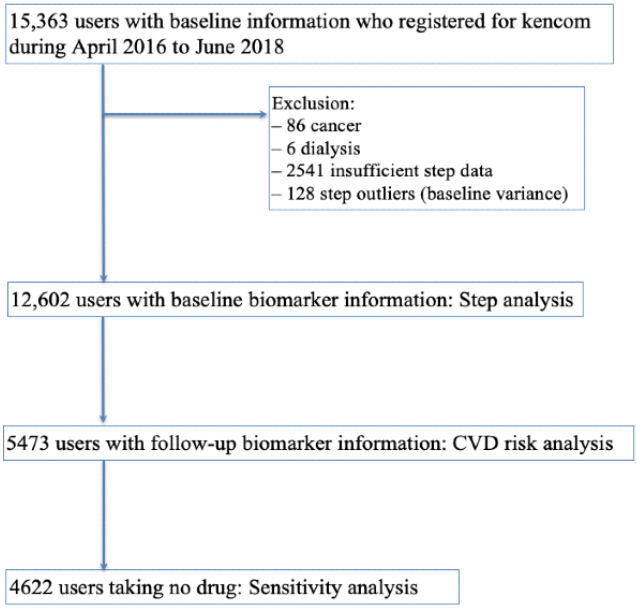


Figure 3. Study flow. CVD: cardiovascular disease.



Step Analysis

Pre- and post-registration daily step counts were defined as the 1-year average step count before and after kencom registration (Figure 2). This definition allowed us to evaluate the change in step (post- minus pre-registration daily step count), minimizing the seasonal variation of step count [11]. The changes in average daily steps following kencom registration (ie, the mean of each user's averaged steps over a year) were assessed for weekdays and the weekend. Furthermore, we compared the average daily step counts over each Arukatsu period, 1 month after the event, 1-2 months after the event, and 2-3 months after the event between the Arukatsu participants and nonparticipants.

CVD Risk Analysis

The changes in CVD biomarkers (weight, systolic/diastolic BP, low-density lipoprotein [LDL]/high-density lipoprotein [HDL] cholesterol, triglyceride, and hemoglobin A_{1c} [HbA_{1c}]) over at least 1-year usage of the app were assessed. Participants were divided according to the quintile of the changes in daily steps after kencom registration, and the changes in CVD biomarkers (follow-up minus baseline check-up data) were evaluated with the lowest quintile as the reference.

Statistical Analysis

Depending on the normality of variables, continuous data were expressed as mean (SD) or median (IQR). The changes in daily step count or biomarker levels following kencom app registration were assessed by unpaired *t* test. We estimated the associations between the changes in CVD biomarkers and the

changes in step count following kencom registration by multiple linear models with the biomarkers as the dependent variables. Multivariable model 1 was adjusted for age (continuous), sex (male/female), BMI (continuous), current smoking (yes/no), and alcohol drinking (yes/no) at baseline health check-up. Intention to improve lifestyle ("not interested," "considering," and "working on") was further adjusted in multivariable model 2, because this question in the Japanese annual health check-ups has not been well scientifically validated. The linear trends in relation to the changes in CVD biomarkers' levels were assessed with the use of quintiles of changes in daily steps following kencom registration as continuous variables in which the median change in the corresponding quintile groups was assigned. Sensitivity analysis was performed among 4622 users not taking any medications for hypertension, dyslipidemia, and diabetes mellitus. All analyses were performed using R 3.6.1 (The R Foundation). Statistical significance was set at a two-tailed $P<.05$ and multiple comparisons were adjusted by the Bonferroni method.

Results

Baseline Characteristics

Baseline cohort characteristics for the step and CVD risk analyses are summarized in Table 1. Among 12,602 users for the step analysis, the mean age was 44.1 (SD 10.2), and 52.25% (6584/12,602) were male. Distribution of age is illustrated in Multimedia Appendix 2. For the intention to improve their lifestyles, 22.9% (1225/5339), 44.6% (2380/5339), and 32.5% (1734/5339) of the users answered "not interested," "considering," and "working on," respectively. The median (IQR) levels of BMI, systolic BP, LDL cholesterol, HDL cholesterol, triglyceride, and HbA_{1c} at baseline were 22.4 (20.4-24.7) kg/m^2, 117 (107-128) mmHg, 119 (100-142) mg/dL, 63 (53-75) mg/dL, 80 (57-119) mg/dL, and 5.4 (5.2-5.6) %, respectively. The users accessed the kencom app for a median of 6.8 (IQR 2.6-14.1) days per month. The characteristics were similar to those of the population for the CVD risk analysis except for the male proportion (63.1% [3451/5473] in the CVD risk analysis).

Table 1. Cohort characteristics.

Characteristics	Step analysis (n=12,602)	CVD[a] risk analysis (n=5473)
Age, year, mean (SD)	44.1 (10.2)	43.8 (10.2)
Male, n (%)	6584 (52.2)	3451 (63.1)
BMI (kg/m^2), median (IQR)	22.4 (20.4-24.7)	22.5 (20.5-24.7)
Current smoking, n/N (%)	2724/12,389 (22.0)	1060/5463 (19.4)
Alcohol drink, n/N (%)	6408/8673 (73.9)	3473/4589 (75.7)
Intention to improve the lifestyle, n/N (%)		
Not interested	1225/5339 (22.9)	685/2895 (23.7)
Considering	2380/5339 (44.6)	1258/2895 (43.5)
Working on	1734/5339 (32.5)	952/2895 (32.9)
Systolic BP[b] (mm/Hg), median (IQR)	117 (107-128)	117 (108-127)
Diastolic BP (mm/Hg), median (IQR)	72 (65-81)	72 (65-81)
LDL[c] cholesterol (mg/dL), median (IQR)	119 (100-142)	119 (100-140)
HDL[d] cholesterol (mg/dL), median (IQR)	63 (53-75)	61 (52-73)
Triglyceride (mg/dL), median (IQR)	80 (57-119)	82 (58-121)
HbA_{1c}[e] (%), median (IQR)	5.4 (5.2-5.6)	5.4 (5.2-5.6)
Average frequency of access per month, median (IQR)	6.8 (2.6-14.1)	6 (1.8-13.2)

[a]CVD: cardiovascular disease.

[b]BP: blood pressure.

[c]LDL: low-density lipoprotein.

[d]HDL: high-density lipoprotein.

[e]HbA_{1c}: hemoglobin A_{1c}

Effects of Kencom on Step Count

Mean step counts before and after kencom registration were 5642 (SD 2686) steps/day and 6152 (SD 2723) steps/day, respectively. The daily step count was significantly increased following kencom registration by a mean of 510 steps/day ($P<.001$). Figure 4A visualizes the change in averaged daily step count in each month before and after kencom registration. The increase in step counts persisted for a year after kencom registration. Physical activity levels were generally higher on

weekdays than on the weekend, and the kencom app positively influenced both step counts (step change: 534 steps/day, $P<.001$ for weekdays; 463 steps/day, $P<.001$ for the weekends; Figure 4B). The spikes in month 1 and month 7 might reflect the Arukatsu event, in which users form teams and each team competes in step counts over a month, thereby exerting a positive peer influence on the participants to improve their physical activity levels. The company promoted kencom when the event was held, which led to increases in the number of users, highlighting the first spike at month 1. The event has been held almost 6 months apart; therefore, the users who experienced Arukatsu at month 1 were likely to experience the next event at month 7, represented as the next spike.

Table 2 describes the average daily step counts during and after each Arukatsu event comparing participants and nonparticipants. On average, the step counts were 1000-2000 steps higher in Arukatsu participants than in the nonparticipants during the event period. The differences in steps became smaller after the event, while generally the participants had higher daily steps than nonparticipants 3 months after the event.

Figure 4. Changes in average daily step count following kencom registration. Dots indicate the mean daily step count of 12,602 users according to the months before/after kencom registration, and bars indicate the standard errors. In total cohort, the 1-year mean daily step count was increased from 5,642 steps/day to 6,152 steps/day following the registration (A). While the average step count was higher on weekdays compared with weekends, the increase in step counts was documented similarly for weekdays and weekends (B).

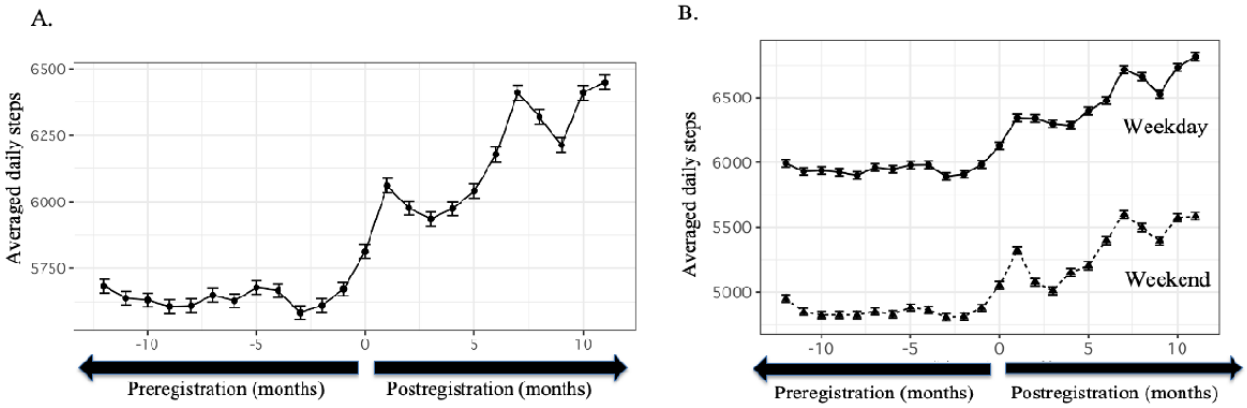


Table 2. Association of Arukatsu event participation and average daily steps.

Event	N	During the event, mean (SD)	Event to month 1[a], mean (SD)	Month 1 to month 2[a], mean (SD)	Month 2 to month 3[a], mean (SD)
2016 event					
Participants	161	8103 (3943)	7067 (3020)	6693 (2807)	7105 (3019)
Nonparticipants	2032	6575 (2553)	6433 (2576)	6160 (2534)	6549 (2677)
2017 First event					
Participants	2017	8056 (3639)	7051 (2772)	6815 (2713)	6761 (2740)
Nonparticipants	3374	6296 (2764)	6385 (2792)	6133 (2716)	5926 (2639)
2017 Second event					
Participants	3083	7607 (4332)	6727 (3277)	6410 (3136)	6353 (2962)
Nonparticipants	5306	6321 (2989)	6238 (2781)	5972 (2662)	6170 (2941)
2018 First event					
Participants	6898	7030 (3890)	6274 (3289)	6063 (3305)	5966 (3109)
Nonparticipants	5434	6317 (2786)	6415 (3129)	6190 (2921)	6046 (2915)
2018 Second event					
Participants	5308	8295 (4205)	7242 (3471)	6865 (3409)	7203 (3595)
Nonparticipants	6610	6327 (3156)	6095 (3040)	5876 (2998)	6193 (3149)

[a]Month X refers to the X month after the corresponding Arukatsu event.

CVD Risk Analysis

Baseline/follow-up CVD biomarkers were available for 5473 users. They were classified according to the quintile of the change in daily steps (<–578, <–122, <248, <864, and ≥864) and the changes in CVD biomarkers following kencom registration were assessed. Baseline characteristics were similar across the quintile groups except for “intention to improve the

lifestyle" (Multimedia Appendix 3). Higher increase in daily step count was associated with a higher prevalence of users already working on improving their lifestyles and higher average frequency of access to kencom.

Increase in step counts was significantly associated with the improvement in weight, HDL cholesterol, and HbA_{1c} (Figure 5; $P<.001$ for all 3; the adjusted threshold for the P value was .007). In adjusted linear regression models, compared with the lowest quintile, the highest quintile in daily step change was associated with a significant reduction in weight (–0.92 kg, $P<.001$), LDL cholesterol (–2.78 mg/dL, $P=.004$), HbA_{1c} (–0.04%, $P=.004$), and increase in HDL cholesterol (+1.91 mg/dL, $P<.001$; Table 3). Further consideration of "intention to improve lifestyle" attenuated the improvement in LDL cholesterol and HbA_{1c} ($P=.10$ and .49, respectively), but changes in weight and HDL cholesterol remained significant ($P<.001$ each, respectively). The adjusted threshold for the P value was .007.

Figure 5. Changes in CVD biomarkers among users according to the changes in daily step count. Changes in CVD biomarkers following kencom registration (follow-up data minus baseline data) were compared according to the quintile of the changes in the user's step count (corresponding to Q1 to Q5 in the x-axis). Bars represent the average changes in the biomarker in each unit and the error bars indicate the standard errors. P values are for the linear trend. BP: blood pressure; CVD: cardiovascular disease; HbA_{1c}: hemoglobin A_{1c}; HDL: high-density lipoprotein; LDL: low-density lipoprotein.

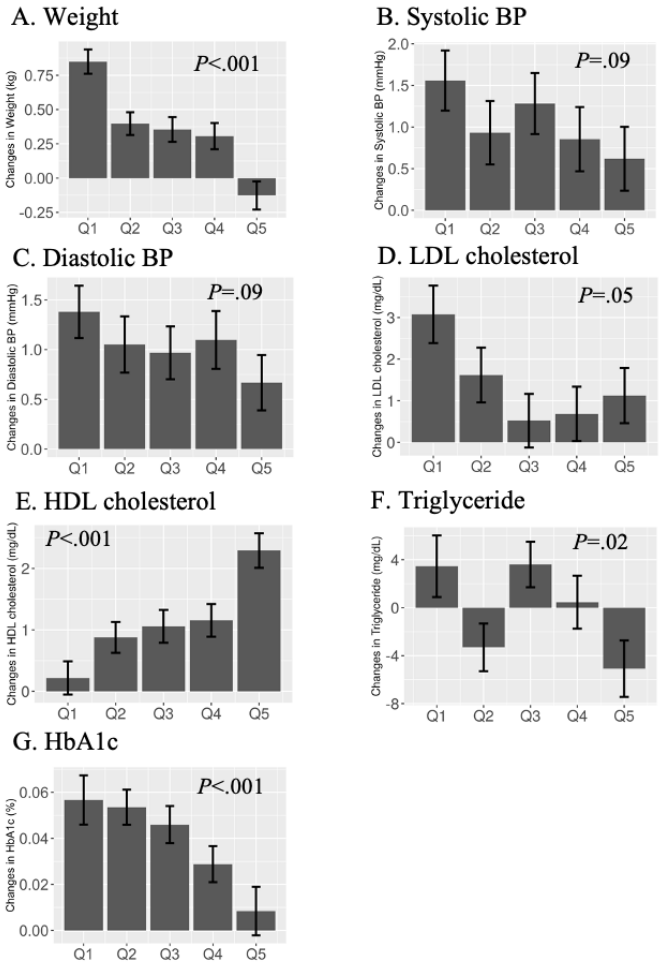


Table 3. Association between the changes in step count and the changes in cardiovascular disease risk biomarkers among 5473 users whose biomarker information was available[a,b].

Cardiovascular disease risk biomarkers	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5	*P* value[c]
Weight (kg)						
Model 1[d]	—	–0.43	–0.45	–0.53	–0.92	<.001
Model 2[e]	—	–0.39	–0.32	–0.50	–0.84	<.001
Systolic blood pressure (mmHg)						
Model 1	—	–0.66	–0.14	–0.35	–0.83	.22
Model 2	—	–0.97	–0.62	0.01	–1.01	.40
Diastolic blood pressure (mmHg)						
Model 1	—	–0.36	–0.08	0.24	–0.73	.21
Model 2	—	–0.26	–0.34	0.50	–0.51	.66
Low-density lipoprotein cholesterol (mg/dL)						
Model 1	—	–1.33	–2.85	–2.48	–2.78	.004
Model 2	—	–0.37	–1.92	–1.55	–1.87	.10
High-density lipoprotein cholesterol (mg/dL)						
Model 1	—	0.48	0.35	0.90	1.91	<.001
Model 2	—	0.47	0.38	0.73	1.97	<.001
Triglyceride (mg/dL)						
Model 1	—	–6.37	1.74	–3.95	–8.14	.04
Model 2	—	–4.82	7.80	–1.35	–8.10	.17
HbA_{1c} (%)						
Model 1	—	0.00	0.00	–0.01	–0.04	.004
Model 2	—	–0.01	–0.01	–0.01	–0.01	.49

[a]A total of 5473 users are divided into 5 groups according to the quintile of the change in the daily step count following kencom registration.

[b]Values are the mean changes compared with quintile 1.

[c]Values are for the linear trend.

[d]Model 1 was adjusted for age, sex, BMI, smoking, and alcohol drinking.

[e]Model 2 was further adjusted for the intention to improve lifestyle (“not interested,” “considering,” and “working on”).

Sensitivity Analysis

Sensitivity analyses were performed among 4622 healthy participants who did not take any medications for hypertension, dyslipidemia, and diabetes mellitus. They were divided into 5 groups according to their changes in step counts following kencom registration (Table 4). After full adjustment for the confounders, the highest quintile was associated with a significant reduction in weight (–0.75 kg, *P*<.001) and increase in HDL cholesterol (+1.84 mg/dL, *P*=.002) compared with the lowest quintile.

Table 4. Association between the changes in step count and the changes in cardiovascular disease risk biomarkers among 4622 users not taking any drugs for hypertension, dyslipidemia, and diabetes mellitus[a,b]

Cardiovascular disease risk biomarkers	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5	*P* value[c]
Weight (kg)						
Model 1[d]	—	–0.45	–0.38	–0.49	–0.79	<.001
Model 2[e]	—	–0.45	–0.26	–0.46	–0.75	<.001
Systolic blood pressure (mmHg)						
Model 1	—	–0.49	–0.29	–0.20	–0.79	.26
Model 2	—	–1.23	–1.06	0.01	–0.98	.51
Diastolic blood pressure (mmHg)						
Model 1	—	–0.25	–0.22	0.02	–0.82	.10
Model 2	—	–0.46	–0.67	0.12	–0.64	.48
Low-density lipoprotein cholesterol (mg/dL)						
Model 1	—	–0.98	–2.53	–2.05	–2.72	.005
Model 2	—	–0.10	–1.55	–1.29	–2.25	.06
High-density lipoprotein cholesterol (mg/dL)						
Model 1	—	0.48	0.42	0.87	1.66	<.001
Model 2	—	0.51	0.40	0.60	1.84	.002
Triglyceride (mg/dL)						
Model 1	—	–7.44	0.07	–4.78	–7.11	.10
Model 2	—	–6.43	6.97	–2.82	–6.64	.30
HbA_{1c}[f] (%)						
Model 1	—	–0.01	0.00	–0.02	–0.03	.008
Model 2	—	–0.02	0.00	–0.02	–0.02	.26

[a]A total of 4622 users taking no drugs for hypertension, dyslipidemia, and diabetes mellitus are divided into 5 groups according to the quintile of the change in the daily step count following kencom registration.

[b]Values are the mean changes compared with quintile 1.

[c]Values are for the linear trend.

[d]Model 1 was adjusted for age, sex, BMI, smoking, and alcohol drinking.

[e]Model 2 was further adjusted for the intention to improve lifestyle (“not interested,” “considering,” and “working on”).

[f]HbA_{1c}: hemoglobin A_{1c}

Discussion

Principal Findings

This is the first study to evaluate the effects of kencom, an mHealth app with integrated functions for healthy lifestyles, on physical activity levels and CVD biomarkers. The use of the kencom app was associated with increased daily step counts by a mean of 510 steps/day, and the higher increase in step counts was related to subsequent weight loss and improvement of lipid profile, in particular HDL cholesterol. The participations in Arukatsu events consistently had increased daily steps. Therefore, commitment to the kencom app might be associated with improved CVD risks through increased physical activity. Notably, this cohort study was conducted within the framework of kencom, in which the Japanese health data from multiple data sources were integrated; it may potentially become a much larger platform of anonymized health data where various epidemiological studies could be implemented.

Association of Kencom and Steps

The present result of an averaged increase of 510 steps/day was not outstanding among incentive-based interventions, as a recent meta-analysis of 12 randomized controlled trials (RCTs) demonstrated an average increase of 607 steps/day by such interventions [11]. However, the observed changes in the activity level were after just registration of the app, unlike after the interventions of the RCTs. Furthermore, the meta-analysis had high heterogeneity due to the constituent small-sized RCTs and was also likely to be confounded by publication bias [11]. The present association of kencom registration and increase in daily step counts was robust because the analysis based on almost all-comers was the comparison of the average step counts over a year, separately in weekdays and weekends, thus accounting

for the selection bias and well-established confounding by holiday period [4,12] and by seasons [13]. Although mHealth apps have typically high attrition rate that attenuates the effects [14], we observed that the increase in daily step counts after the registration of the kencom app was sustained at least one year, supporting the long-term efficacy of this app.

Insight of Behavioral Economics Into the Mechanisms

Theories of behavioral economics may offer the mechanistic explanation as to why the kencom app can positively influence the user's behavior [15]. The dual process theory describes 2 different systems that trigger human behavior: intuition or emotion-linked unconscious reasoning (System 1) and rule-based, analytic conscious reasoning (System 2) [16]. System 1 is susceptible to cognitive biases that can lead to "irrationality" dictated by the subjective reality, often precluding people's health-oriented behaviors. The kencom app has been designed to modify the behaviors of individuals in contemplation stage in which status quo bias (preference for the current state) [17], present bias (tendency toward a small present reward compared with a big future reward), and overoptimism (unrealistic belief that one will less likely encounter negative events) play major roles in their decision makings. The app leverages functions that can work on System 1 and preclude the biases toward sedentary behaviors while accelerating positive biases such as peer bias. This characteristic satisfies the NUDGES framework as summarized in Table 5 [18].

Table 5. NUDGES framework of kencom.[a,b]

Framework components	Explanation	Corresponding features of kencom (examples)
i**N**centives	Incentives encourage the action	kencom points for daily logging-in and entry/winning of Arukatsu events; users can get gift vouchers from kencom points that can be spent in the market place
Understanding mappings	Understandings of the relations between choice and welfare prevent dropouts	Easy and detailed description of Arukatsu events to prevent dropouts; tailored articles and visualization of physical activity may appeal to the users' emotion that healthy behaviors may lead to mental, physical, and financial benefit
Defaults	Better default settings improve the usability	Positive, big images like savory dishes or beautiful landscapes are displayed in the initial screen after logging-in (priming effect); no need to input step or health data by users
Giving feedbacks	Positive feedback keep one's motivation	Setting of the daily goals of physical activity (declaration of commitment) and positive feedback according to the self-checked achievements; comment function in Arukatsu to encourage the team members with each other
Expecting errors	Prevention of expecting errors enhances one's commitment	One-click entry for Arukatsu to avoid procrastination; synchronization with smartphone accelerometers to keep daily record and minimize measurement errors
Structure complex choices	Avoidance of complex options and optimization of the structures lead to less stressful experiences	The simple design allows anyone including those not familiar with a smartphone to use the app immediately; optimized user experience according to the frequency of user's decision

[a]The features of kencom that satisfy NUDGES framework are summarized [18].

[b]The adherence to the NUDGES framework may lead to a better chance of altering one's behaviors.

Some nudges have only short-term effects that inevitably limit the health benefits. The structure of kencom aimed to overcome this limitation by (1) providing multiple interventions in appropriate timings and (2) offering health education that may work on System 2 simultaneously with the nudge interventions. The following points illustrate the detailed interventions, potential effects, and the mechanisms of kencom:

- Incentives are offered for each daily login to kencom, aiming at habituating the behavior.
- Positive, big images such as savory dishes or beautiful landscapes are displayed on the initial screen after login, consistent with priming effects. The user's positive feeling activated by the stimulus may sustain during the exploration of the app.
- The tailored health information is designed to be read like a newsfeed during commute or lunch time, when the users are not yet exhausted in a day and maintain their self-control. The contents generally focus on the "feasible action (plan)," with the intention to shifting the user's fixed mindset into a growth mindset.
- The tailored health information is evidence based, which may improve the user's health literacy.
- Arukatsu events are announced through the app and workplace. The participation may be prompted by the user's fostered growth mindset, incentives, and peer influence from the colleagues.
- Arukatsu events leverage the use of game design elements in nongame contexts (competition of total step counts of the teams) [19], sharing of the step counts with team members, and comment function. These features may facilitate the participants' physical activity through positive peer influence.

Comparison With Other mHealth Apps

Other interventions leveraging cognitive biases have been evaluated for their efficacies in improving physical activities, including the Shape Up Rhode Island state-wide campaign [20], the SHAPE-UP program at a Northeastern university [21], and more recently, the CARADA app [22]. The concept of an event in the CARADA app was similar to Arukatsu, while the evaluation was conducted in much fewer participants (n=175)

by one-time intervention, without any analysis on the postintervention periods [22]. Arukatsu events have been repeatedly conducted almost by half a year, which could foster the users' participation and commitment to the events and apps, eventually contributing to the consistently increasing trend of steps after kencom registration.

The structure of kencom systematically differs from that of popular, scientifically evaluated mHealth apps such as Sweatcoin [4] and Carrot Rewards [23,24], in which incentive systems are leveraged as the main mechanism to stimulate the users' physical activity; in Sweatcoin, digital currency was provided according to how many steps the users take; and in Carrot Rewards, users can earn loyalty points by achieving daily step goals. The "direct" incentives of these apps may efficiently enhance daily activity levels as the observational studies demonstrated an increase of over 1000 steps/day by using these apps. However, caution is needed to interpret these findings; the study on Sweatcoin could have serious selection bias because only 5892 out of more than 30 million committed users were analyzed [4]; and that of Carrot Rewards defined baseline steps based on the steps taken by the users during the 2-week induction period, potentially sensitive to the daily or seasonal variation [23], compared with the 1-year baseline evaluation in our study. In comparison, our analyses robustly support the association of kencom usage and enhanced physical activity with less bias, indicative of a novel mHealth approach toward population immobility without "direct" incentives.

Implications on CVD Risks

The enhanced activity following kencom app registration was associated with improvement in CVD biomarkers, especially in weight and HDL cholesterol. This observation is in accordance with previous studies showing the positive relationship between walking exercise and HDL cholesterol [25,26]. Such an objective improvement in CVD risk factors further underscores the effectiveness of the use of the kencom app. Although the associations were consistent even after controlling for the semiquantitative intention to improve their lifestyles, the effect sizes were not clinically remarkable; this may highlight the characteristics of generally healthy, young participants and the variations in the users' commitment on the app. Longitudinal analyses with longer follow-up data from multiple perspectives may confer the more precise and detailed effects and public health implications of the kencom app usage.

Limitations

The strengths of our study are the large sample size, minimum exclusion criteria, rigorous definitions of baseline/follow-up step counts by averaging over 1-year data, and assessments of serial CVD biomarkers. Our study does have several important limitations. Study participants were mainly healthy Japanese and as such, our results may not be generalizable to other populations. Step data could be biased because the data solely from iPhone users were used, whereas there may be little difference in demographics between iPhone and Android users [27]. We cannot entirely rule out the possibility of unmeasured confounding factors, whereas the homogeneity of this study population minimized them. Finally, the current analysis was among users of the kencom app and a comparison with controls was not conducted. An RCT is warranted to infer causal effects of kencom app use on health outcomes.

Conclusions

The use of an mHealth app "kencom" was associated with enhanced physical activity, which was related to the weight loss and increase of HDL cholesterol. This cohort study was conducted within the framework of kencom, in which health data from multiple data sources were anonymized and integrated, implicating the future usage to build a much larger database to provide reliable evidence.

Acknowledgments

We thank Akiko Tanaka, MSc, for assisting with the handling and interpretation of the health insurance claims data. This study was funded by DeSC Inc.

Conflicts of Interest

RH and HF received consultancy fees from DeSC Inc., including fees for this study. MT received consultancy fees from DeSC Inc. outside of this project. MM, RM, KN, and YT are employees of DeSC Inc. KM is a representative director (medical doctor) of DeSC Inc.

Multimedia Appendix 1
Summary of the variables from each data source.
[DOCX File , 18 KB - jmir_v23i4e21622_app1.docx]

Multimedia Appendix 2
Age distribution.
[PNG File , 64 KB - jmir_v23i4e21622_app2.png]

Multimedia Appendix 3
Baseline characteristics across the quintiles of changes in daily steps following kencom registration.

[DOCX File , 20 KB - jmir_v23i4e21622_app3.docx]

References

1. Riegel B, Moser DK, Buck HG, Dickson VV, Dunbar SB, Lee CS, American Heart Association Council on CardiovascularStroke Nursing; Council on Peripheral Vascular Disease;Council on Quality of CareOutcomes Research. Self-Care for the Prevention and Management of Cardiovascular Disease and Stroke: A Scientific Statement for Healthcare Professionals From the American Heart Association. J Am Heart Assoc 2017 Aug 31;6(9):e006997 [FREE Full text] [doi: 10.1161/JAHA.117.006997] [Medline: 28860232]
2. Burke LE, Ma J, Azar KMJ, Bennett GG, Peterson ED, Zheng Y, et al. Current Science on Consumer Use of Mobile Health for Cardiovascular Disease Prevention: A Scientific Statement From the American Heart Association. Circulation 2015 Sep 22;132(12):1157-1213. [doi: 10.1161/CIR.0000000000000232] [Medline: 26271892]
3. Coorey GM, Neubeck L, Mulley J, Redfern J. Effectiveness, acceptability and usefulness of mobile applications for cardiovascular disease self-management: Systematic review with meta-synthesis of quantitative and qualitative data. Eur J Prev Cardiol 2018 Mar;25(5):505-521. [doi: 10.1177/2047487317750913] [Medline: 29313363]
4. Elliott M, Eck F, Khmelev E, Derlyatka A, Fomenko O. Physical Activity Behavior Change Driven by Engagement With an Incentive-Based App: Evaluating the Impact of Sweatcoin. JMIR Mhealth Uhealth 2019 Jul 08;7(7):e12445 [FREE Full text] [doi: 10.2196/12445] [Medline: 31287064]
5. Flores MG, Granado-Font E, Ferré-Grau C, Montaña-Carreras X. Mobile Phone Apps to Promote Weight Loss and Increase Physical Activity: A Systematic Review and Meta-Analysis. J Med Internet Res 2015;17(11):e253 [FREE Full text] [doi: 10.2196/jmir.4836] [Medline: 26554314]
6. Payne HE, Lister C, West JH, Bernhardt JM. Behavioral functionality of mobile apps in health interventions: a systematic review of the literature. JMIR Mhealth Uhealth 2015;3(1):e20 [FREE Full text] [doi: 10.2196/mhealth.3335] [Medline: 25803705]
7. Whitehead L, Seaton P. The Effectiveness of Self-Management Mobile Phone and Tablet Apps in Long-term Condition Management: A Systematic Review. J Med Internet Res 2016;18(5):e97 [FREE Full text] [doi: 10.2196/jmir.4883] [Medline: 27185295]
8. Seo WK, Kang J, Jeon M, Lee K, Lee S, Kim JH, et al. Feasibility of using a mobile application for the monitoring and management of stroke-associated risk factors. J Clin Neurol 2015 Apr;11(2):142-148 [FREE Full text] [doi: 10.3988/jcn.2015.11.2.142] [Medline: 25851892]
9. O'Malley G, Baker P, Francis D, Perry I, Foster C. Incentive‐based interventions for increasing physical activity and fitness. Cochrane Database of Systematic Reviews. 2012. URL: https://doi.org//10.1002/14651858.CD009598 [accessed 2021-03-31]
10. Khamzina M, Parab KV, An R, Bullard T, Grigsby-Toussaint DS. Impact of Pokémon Go on Physical Activity: A Systematic Review and Meta-Analysis. Am J Prev Med 2020 Feb;58(2):270-282. [doi: 10.1016/j.amepre.2019.09.005] [Medline: 31836333]
11. Mitchell MS, Orstad SL, Biswas A, Oh PI, Jay M, Pakosh MT, et al. Financial incentives for physical activity in adults: systematic review and meta-analysis. Br J Sports Med 2020 Nov;54(21):1259-1268. [doi: 10.1136/bjsports-2019-100633] [Medline: 31092399]
12. Booth M. Assessment of physical activity: an international perspective. Res Q Exerc Sport 2000 Jun;71 Suppl 2:114-120. [doi: 10.1080/02701367.2000.11082794] [Medline: 25680021]
13. Matthews CE, Freedson PS, Hebert JR, Stanek EJ, Merriam PA, Rosal MC, et al. Seasonal variation in household, occupational, and leisure time physical activity: longitudinal analyses from the seasonal variation of blood cholesterol study. Am J Epidemiol 2001 Jan 15;153(2):172-183. [doi: 10.1093/aje/153.2.172] [Medline: 11159163]
14. Becker S, Miron-Shatz T, Schumacher N, Krocza J, Diamantidis C, Albrecht U. mHealth 2.0: Experiences, Possibilities, and Perspectives. JMIR Mhealth Uhealth 2014;2(2):e24 [FREE Full text] [doi: 10.2196/mhealth.3328] [Medline: 25099752]
15. Samson A. The Behavioral Economics Guide 2020 (with an Introduction by Colin Camerer). 2020. URL: https://www.behavioraleconomics.com [accessed 2020-12-15]
16. Kahneman D, Frederick S. Representativeness revisited: Attribute substitution in intuitive judgment. Heuristicsbiases: The psychology of intuitive judgment. New York, NY. Cambridge, MA: Cambridge University Press; 2002:49-81.
17. Kahneman D, Knetsch JL, Thaler RH. Anomalies: The Endowment Effect, Loss Aversion, and Status Quo Bias. Journal of Economic Perspectives 1991 Feb 01;5(1):193-206. [doi: 10.1257/jep.5.1.193]
18. Flanagan S. Nudge: Improving decisions about health, wealth, and happiness. Planning for Higher Education 2010;38(2):57+.
19. Patel MS, Small DS, Harrison JD, Fortunato MP, Oon AL, Rareshide CAL, et al. Effectiveness of Behaviorally Designed Gamification Interventions With Social Incentives for Increasing Physical Activity Among Overweight and Obese Adults Across the United States: The STEP UP Randomized Clinical Trial. JAMA Intern Med 2019 Sep 09:1-9 [FREE Full text] [doi: 10.1001/jamainternmed.2019.3505] [Medline: 31498375]
20. Leahey TM, Kumar R, Weinberg BM, Wing RR. Teammates and social influence affect weight loss outcomes in a team-based weight loss competition. Obesity (Silver Spring) 2012 Jul;20(7):1413-1418 [FREE Full text] [doi: 10.1038/oby.2012.18] [Medline: 22310234]

21. Zhang J, Brackbill D, Yang S, Becker J, Herbert N, Centola D. Support or competition? How online social networks increase physical activity: A randomized controlled trial. Prev Med Rep 2016 Dec;4:453-458 [FREE Full text] [doi: 10.1016/j.pmedr.2016.08.008] [Medline: 27617191]
22. Hamamatsu Y, Ide H, Kakinuma M, Furui Y. Maintaining Physical Activity Level Through Team-Based Walking With a Mobile Health Intervention: Cross-Sectional Observational Study. JMIR Mhealth Uhealth 2020 Jul 03;8(7):e16159 [FREE Full text] [doi: 10.2196/16159] [Medline: 32618576]
23. Mitchell M, Lau E, White L, Faulkner G. Commercial app use linked with sustained physical activity in two Canadian provinces: a 12-month quasi-experimental study. Int J Behav Nutr Phys Act 2020 Feb 25;17(1):24 [FREE Full text] [doi: 10.1186/s12966-020-00926-7] [Medline: 32098625]
24. Rondina Ii R, Pearson EK, Prapavessis H, White L, Richard S, Mitchell MS. Bright spots, physical activity investments that (almost) worked: Carrot Rewards app, driving engagement with pennies a day. Br J Sports Med 2020 Aug;54(15):927-929. [doi: 10.1136/bjsports-2019-100969] [Medline: 31712256]
25. Chiang T, Chen C, Hsu C, Lin Y, Wu H. Is the goal of 12,000 steps per day sufficient for improving body composition and metabolic syndrome? The necessity of combining exercise intensity: a randomized controlled trial. BMC Public Health 2019 Sep 03;19(1):1215 [FREE Full text] [doi: 10.1186/s12889-019-7554-y] [Medline: 31481039]
26. Schneider PL, Bassett DR, Thompson DL, Pronk NP, Bielak KM. Effects of a 10,000 steps per day goal in overweight adults. Am J Health Promot 2006;21(2):85-89. [doi: 10.4278/0890-1171-21.2.85] [Medline: 17152246]
27. Götz FM, Stieger S, Reips U. Users of the main smartphone operating systems (iOS, Android) differ only little in personality. PLoS One 2017;12(5):e0176921 [FREE Full text] [doi: 10.1371/journal.pone.0176921] [Medline: 28467473]

Abbreviations

BP: blood pressure
CVD: cardiovascular disease
HbA_{1c}: hemoglobin A_{1c}
HDL: high-density lipoprotein
LDL: low-density lipoprotein
mHealth: mobile health
RCT: randomized controlled trial

Edited by R Kukafka; submitted 19.06.20; peer-reviewed by M Elliott, AS Aslam; comments to author 19.07.20; revised version received 24.09.20; accepted 21.03.21; published 26.04.21.

Please cite as:
Hamaya R, Fukuda H, Takebayashi M, Mori M, Matsushima R, Nakano K, Miyake K, Tani Y, Yokokawa H
Effects of an mHealth App (Kencom) With Integrated Functions for Healthy Lifestyles on Physical Activity Levels and Cardiovascular Risk Biomarkers: Observational Study of 12,602 Users
J Med Internet Res 2021;23(4):e21622
URL: https://www.jmir.org/2021/4/e21622
doi:10.2196/21622
PMID:33900203

Viewpoint

Theory Integration for Lifestyle Behavior Change in the Digital Age: An Adaptive Decision-Making Framework

Chao Zhang[1], PhD; Daniël Lakens[1], PhD; Wijnand A IJsselsteijn[1], PhD

Human-Technology Interaction Group, Department of Industrial Engineering & Innovation Sciences, Eindhoven University of Technology, Eindhoven, Netherlands

Corresponding Author:
Chao Zhang, PhD
Human-Technology Interaction Group, Department of Industrial Engineering & Innovation Sciences, Eindhoven University of Technology
PO Box 513
Eindhoven, 5600 MB
Netherlands
Phone: 31 624749479
Email: chao.zhang87@gmail.com

Related Article:

This is a corrected version. See correction statement: https://www.jmir.org/2021/4/e29629

Abstract

Despite the growing popularity of digital health interventions, limitations of traditional behavior change theories and a lack of theory integration hinder theory-driven behavior change applications. In this paper, we aim to review theories relevant to lifestyle behavior change from the broader psychology literature and then integrate these theories into a new theoretical framework called adaptive decision-making to address two specific problems. First, our framework represents lifestyle behaviors at two levels—one of individual daily decisions (action level) and one of larger behavioral episodes (reflection level)—to more closely match the temporal characteristics of lifestyle behaviors and their associated digital data. Second, the framework connects decision-making theories and learning theories to explain how behaviors and cognitive constructs dynamically influence each other, making it a suitable scaffold for building computational models. We map common digital intervention techniques onto the behavioral and cognitive processes in the framework and discuss possible contributions of the framework to both theory development and digital intervention design.

(J Med Internet Res 2021;23(4):e17127) doi:10.2196/17127

KEYWORDS

behavior change; health behavior; digital health intervention; decision-making; learning; self-control; habits; theoretical framework

Introduction

Background

Digital intervention systems have been considered as promising tools to change people's unhealthy lifestyle behaviors, such as eating fast food, not exercising, or having suboptimal dental routines [1-4]. In realizing the potential of such systems, many researchers have advocated the role of behavior change theories, especially their translation to digital lifestyle interventions [3,5-8]. Ideally, behavior change theories and digital lifestyle interventions should inform each other. Good theories, when applied appropriately, are generally expected to increase the effectiveness of interventions. They can be used to identify behavioral determinants as intervention targets, translate general behavior change techniques (BCTs) to fine-tuned features in digital systems, and predict intervention outcomes. On the other hand, the vast amount of behavioral data collected by digital systems could potentially contribute to theory evaluation [9-11]. Compared with data from traditional behavioral experiments, digital behavioral data can have larger and more diverse samples, greater ecological validity, and higher temporal resolution.

Despite these expectations, the synergy between theory development and intervention practice is far from ideal [9]. The role of behavior change theories in digital interventions is not as prominent as hoped. Several surveys indicate that the

application rate of theories in digital intervention trials and commercial eHealth apps ranges between 19% and 52% [6,12-19], and when theories are applied, only 3-5 classical theories dominate the applications [15,17,19-21]. Moreover, although some reviews and meta-analyses have suggested that applying theories has benefits [22,23], other reviews found no clear evidence [24-26] and questioned the value of applying theories in real-world applications [27]. Finally, as for theory development, data collected by digital systems are commonly used to evaluate the effectiveness of specific interventions but are rarely used to examine predictions derived from theoretical models [25].

One factor contributing to this *theory-interventiongap* is the lack of theory integration in the field of behavior change research [28], especially integrations that are tailored to digital lifestyle interventions. Even for a phenomenon as complex as behavior change, the large number of individual theories pertaining to behavior change (83 according to a systematic review [20]) clearly suggests that some integration and unification is probably beneficial for theory development in the field. The sheer number of behavior change theories can be overwhelming for intervention designers who want to grasp the literature and selectively apply theories to their designs. Perhaps the difficulty of orienting oneself with respect to the literature can explain why only a limited set of theories are applied [20]. Many basic theoretical ideas in psychology, despite being highly relevant, are underrepresented in applied research, such as decision-making, reinforcement learning, self-control, and habit formation.

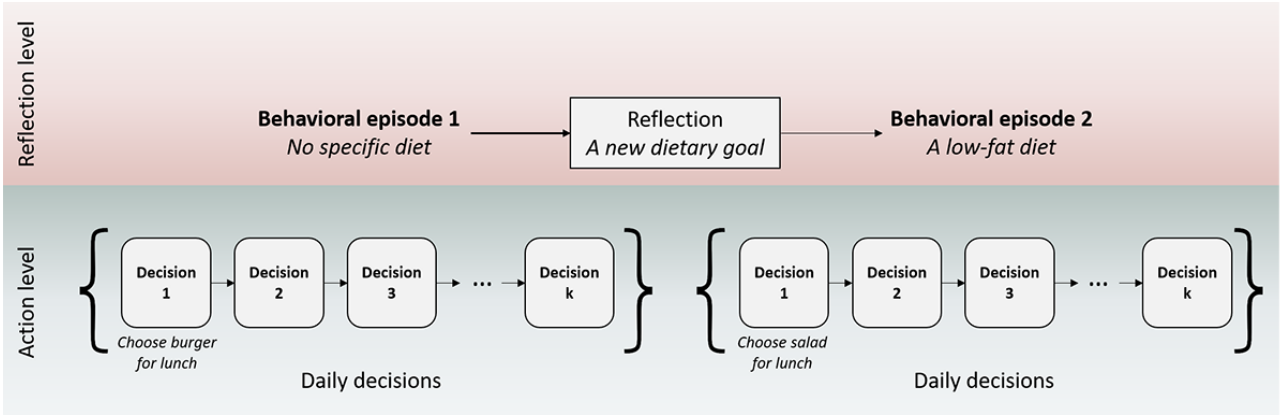
The lack of impact of theories on interventions also raises the question whether the current knowledge about lifestyle behavior change is too limited to be fully useful. Two specific reasons have been proposed to explain why traditional behavior change theories are inadequate in the digital age [6]. First, many prominent traditional theories are static rather than dynamic, in the sense that they provide *snapshot* explanations of what factors determine behavior. Temporal aspects are not taken into consideration. Second, even when time is included in the theories, there is often a mismatch between traditional theories and digital interventions in terms of at what temporal scale behaviors are represented. These two limitations are evident in three of the most applied theories [15,17,19-21]. In the Theory of Planned Behavior (TPB) [29], neither temporal dynamics of the behavioral determinants in the model nor any mechanisms to account for the reciprocal influences of behavior on its determinants are specified. The influential Transtheoretical Model (TTM) includes the temporal aspect of behavior change stages [30]. However, although a healthy-eating app may intervene in its users' daily dietary choices, the TTM only describes the stages of behavior change in terms of months. If a theory represents behavior at a coarse temporal scale, processes at finer scales are overlooked and time-intensive digital interventions cannot be informed. Other theories, such as the Social Cognitive Theory (SCT) [31], consider more rapid interactions between behavior and behavior determinants (eg, self-efficacy); however, the dynamic interaction is theorized only at a very abstract conceptual level without explicitly modeling the role of time or how the interaction works [32]. Both criticisms also coincide with the recent advocates of paying more attention to temporal aspects in health psychology to improve theories and their translation into practice [32,33].

The two aforementioned points should not be considered as criticism of the original theories. Although these theories have been applied in the context of lifestyle behavior change and digital intervention, they were either not meant to explain lifestyle behaviors initially (eg, TPB) or developed at a time when technologies for continuously monitoring lifestyle behaviors were unavailable (eg, TTM or SCT). To advance theories, it is useful to consider some of the often-overlooked characteristics of lifestyle behaviors through the lens of modern digital technologies. Lifestyle behaviors, such as eating, exercising, or toothbrushing, are performed very frequently, as part of everyday habits and routines, and on each occasion, they are fast decisions that are not extensively deliberated. This type of decisions (eg, choosing what to eat for dinner) may be relatively inconsequential; however, they can form larger behavioral *episodes* (eg, following a diet), which may affect one's health significantly over time. This characteristic of hierarchical organization sets lifestyle behaviors apart from single-time health behaviors or decisions, such as cancer screening or vaccination. Moreover, unlike single-time decisions, as lifestyle behaviors are repeated frequently, learning and adaptation through repetitions plays a very significant role in lifestyle changes and interventions. This requires the inclusion of temporal dynamics in behavior change theories.

Objectives

On the basis of the aforementioned rationale, we propose a new integrative theoretical framework, called *adaptive decision-making*, which specifically focuses on lifestyle behaviors and incorporates temporal dynamics. In doing so, the new framework represents lifestyle behaviors at two temporal levels: a lower level (*action level*) that matches the daily individual decisions and the time-intensive interventions realized by digital systems and a higher level (*reflection level*) that matches the episodes of repeated decisions (Figure 1). In addition, both decision-making processes (how behaviors are determined or decisions are made) and learning processes (how earlier behaviors or decisions influence later ones through cognitive variables) at each level will be included in the framework. The goal is to incorporate both traditional and more recent theoretical ideas about behavior change in a single framework and reinterpret these ideas in light of a fine-grained temporal perspective. We hope this effort will facilitate a more integrated approach for developing more precise theories (eg, computational models) and intelligent intervention systems.

Figure 1. A two-level representation of lifestyle behavior (change).



In the remainder of this paper, we first review important theoretical ideas relevant to lifestyle behavior changes from a broad psychology literature. To facilitate theory integration, individual theories are compared in terms of their temporal scales and their emphasis on learning or decision-making. Next, the adaptive decision-making framework is introduced by integrating the relevant but disparate theoretical ideas into a 2-level representation of lifestyle behavior changes presented earlier (Figure 1). Afterward, we relate the framework to intervention practice by mapping common BCTs used in digital systems to the behavioral processes in the framework. The paper concludes with a general discussion on the added value of the framework to behavior change theorists and digital intervention designers.

Review of Individual Theories Relating to Lifestyle Behavior Change

Overview

There are two distinct and complementary traditions for explaining human behavior—a learning tradition and a decision-making tradition [34]. The learning tradition, as its name suggests, focuses on the time course of learning a behavior—in particular, the interdependence among behavioral occasions in a sequence rather than the exact determinants of each occasion. In contrast, researchers in the decision-making tradition care more about what factors determine a behavior on specific occasions and what information is processed at such moments but much less on how repeated decisions are interrelated. As both learning and decision-making aspects are crucial for developing a dynamic framework, this review is organized based on the roots of theories in either tradition. After the review, we briefly discuss whether each theory focuses primarily on explaining individual daily decisions (action level) or episodic behavioral processes (reflection level).

Theories in the Learning Tradition

Reinforcement Learning Theory

Reinforcement learning, or learning by outcomes, is a fundamental form of learning discovered in the early years of modern psychology [35] and is still influential in today's behavioral and brain sciences [36] and artificial intelligence research [37]. Humans and other organisms are theorized to adapt their behaviors through their interactions with changing environments to survive and thrive. If a behavior results in goodness to an organism, the frequency of performing the same behavior increases; conversely, if a bad outcome follows, the behavior will be performed less often in the future. This is summarized as the *law of effect* [38].

Reinforcement theory becomes more complex when one also considers the *law of exercise* [38]. The aforementioned response-outcome learning, or goal-directed learning, is accompanied by stimulus-response learning, also known as a process of habit learning [36]. The distinction between goal-directed learning and habit learning has been demonstrated in instrumental learning experiments where animals or humans are trained to acquire reward-generating responses (eg, pressing a lever to receive food): when a response is overly trained, it persists to be triggered by the corresponding stimulus even when the reward becomes goal irrelevant (eg, when a rodent is satiated) [39]. The recent resurgence of interest in habit formation in social and health psychology also follows the theory of defining habits as mental associations between behaviors and environmental cues [40-42]. When a behavior becomes strongly habitual, goal-related determinants of behavior, such as attitude and intention, cease to influence behavior [43].

Control Theory of Self-regulation

The classical reinforcement learning theory focuses on the role of external immediate rewards in controlling behavior but neglects the role of distal behavioral outcomes that may be cognitively represented. Following criticism of this limitation [44], the control theory of self-regulation assumes that people can mentally represent distal outcomes of goals, and the regulation of behavior is generally toward reducing the discrepancies between the goals and people's current status [45,46]. When a behavior leads to a reduced discrepancy, the reduction itself becomes a reinforcer of the behavior, similar to external rewards. This discrepancy-reduction mechanism is analogous to feedback control systems in engineering, where discrepancies between perceived states and a reference value are constantly monitored to maintain homeostasis.

The control theory also hierarchically represents goals and self-regulation. A 9-level hierarchical control system was proposed by Carver and Scheier [45], in which a behavior output from a higher level serves as the goal reference to the next lower

level. For lifestyle behaviors, it is sufficient to consider three levels: long-term goals (eg, improving health), short-term goals (eg, walking 10,000 steps a day), and actions (eg, taking a specific walk). Taking actions leads to the fulfillment of short-term goals, which in turn brings a person closer to the long-term objectives. Self-regulation operates most frequently at the action level (ie, making daily decisions); however, people's attention can be shifted to higher or lower levels. Downward shifting occurs when lower-level motor control, which is normally highly automated, becomes temporarily impeded during action executions (eg, when learning a new motor skill or when a dysfunctional action has to be inhibited [47]). Upward shifting can be understood as self-reflective moments when a person reconsiders the attainability of a higher-level goal, which is more difficult to predict (Psarra [48]).

Social Cognitive Theory

SCT, proposed by Albert Bandura, is one of the most cited and applied theories in behavior change research [20,21]. The theory encompasses three key concepts: *social learning* [49], *self-efficacy* [50], and *proactive control* [31]. First, based on research on children's learning behaviors [51], *Social Learning Theory* posits that behaviors or attitudes are acquired not only through direct reinforcement but also by observing the behaviors and their corresponding consequences to others [49]. For many health-related behaviors, long-term health consequences are often learned by observing other people's behavioral outcomes. Second, based on organizational decision-making research [52], it was found that subjective belief in one's ability to perform a behavior was closely related to actual performance. According to the control theory mentioned earlier, this self-efficacy belief can be understood as a cognitive mechanism that simulates a series of future actions (eg, dinner choices every day) in an extended episode of goal pursuit (eg, adherence to a diet). If the mentally simulated actions fail to bring sufficient progress, a person may decide to abandon the goal pursuit altogether. Third, Bandura [31] was among the earliest scholars to discuss a discrepancy-production process called proactive control, in which a person sets higher goals to further motivate behavior. Thus, it complements the discrepancy-reduction mechanism at the core of the control theory. The idea that goals are susceptible to changes also allows the possibility of adjusting an unattainable goal downward to reduce its discrepancy to the current status. Altogether, the three concepts contribute to extending reinforcement learning and control theory by incorporating flexibilities in complex human behaviors.

Theories in the Decision-Making Tradition

Expected Utility Theory

Across behavioral sciences (eg, psychology and economics), many mathematical models have been developed to describe how people make choices, given a fixed set of alternatives (options, eg, fries or salad) and attributes (eg, healthiness or tastiness). A fundamental theoretical idea behind many models is the expected utility theory. The theory assumes that people integrate multiple attributes of choice alternatives (their potential for satisfying different personal goals) into a unidimensional construct called expected utility and then choose the alternative with the highest utility [53]. Formally, the expected utility is computed as  , where $V(x_{jn})$ is the subjective value function for the nth possible value of attribute j, and $P(x_{jn})$ is the probabilistic belief that attribute j takes that value [54,55]. The equation implies that the expected utility of one choice alternative increases when choosing the alternative is likely to produce certain outcomes (large $P(x_{jn})$) and when the outcomes are highly valuable (large $V(x_{jn})$). For example, whether people choose salad over fries depends on both their beliefs about their respective benefits for health and their valuations of good health. The theory does not imply that people always consciously follow the equation to compute utilities but rather reflects key neural mechanisms that underlie decision-making [56]. In reality, conscious and deliberative computations are more common for single-time important decisions (eg, comparing different health insurance policies) than for fast daily lifestyle decisions.

Sequential Sampling Models

Empirical data from choice experiments have repeatedly shown that people are less rational than those suggested by classical choice models [57]. People are prone to be influenced by information that is seemingly irrelevant, for example, the addition of an inferior choice option [58] or framing of losses versus gains [59]. To account for these anomalies, a sequential sampling approach has been developed to dynamically model the cognitive process of decision-making, such as the *multialternative decision field theory* [60] and the *associative accumulation model* [57]. The new models share the idea that preferences for different choice alternatives are accumulated over time (eg, a few seconds) and the choice that is made is the choice whose preference signal is first to exceed a decision threshold. At each time step, the preference signals of choice alternatives fluctuate according to a process of utility comparison based on one [57] or multiple attributes (as in drift diffusion models [61]). The stochastic property of sequential sampling models enables them to explain the sensitivity of choices to subtle changes in choice sets and to predict decision time [56]. Finally, sequential sampling models suggest a mechanism for habitual choices, where repeatedly choosing an alternative may shift its starting position of preference accumulation toward a decision threshold at the baseline ([60], Zhang et al, unpublished data, 2021).

Reasoned Action Approach

Influenced by the expected utility theory [62] but with a strong focus on application, the reasoned action approach [63] has produced some of the most applied theories in behavior change research, such as the TPB [29,64] and the Health Belief Model [65]. From a decision-making perspective, this approach categorizes attributes in certain choice situations into a smaller set of behavioral determinants that are generalizable to a wide range of behaviors and measurable by self-report. For example, in the TPB, regardless of the specific alternatives and attributes considered, factors affecting choices are categorized into three determinants, namely attitude, social norm, and perceived behavioral control [64]. When a specific behavior is considered (eg, dinner choice), attitude toward a choice alternative is further determined by many attributes [29], such as taste, nutrition, and

price, whereas social norms are influenced by the perceived social consequences of choosing an alternative (eg, presenting oneself to be environmentally friendly). Perceived behavioral control, similar to self-efficacy, measures one's confidence in maintaining certain choices in the future. TPB was explicitly considered by Prochaska and DiClemente [29] as a model for behavioral prediction rather than for explaining the processes underlying overt behaviors or decisions or how such processes can be influenced.

The reasoned action approach also makes a strong assumption on the intentionality of behavior [66]. For example, behavioral intention is a prerequisite for actual behavior in the TPB [64]. Thus, this approach considers behaviors as *planned* or *intended*, resulting from careful deliberations on the pros and cons of certain behaviors. Such a theoretical position is reasonable because the reasoned action approach was developed to mainly deal with single-time decisions or the planning of behavioral episodes, rather than *small* daily decisions. When applied to lifestyle behaviors, this approach relies on aggregated behavior representations over a substantive period [67].

Dual-Processing Models

A recurrent idea in psychology is that humans possess two distinct modes or systems for processing information and making decisions. Although different dual-system models use different terminologies [68], it is widely accepted that one system is fast, impulsive, and largely automatic, and the other system is slow, reflective, and deliberate [69].

The Reflective-Impulsive Model [70] is a representative of this approach, and it has been explicitly applied to health-related behaviors [71,72]. The reflective system hosts various higher-order mental operations that rely on controlled processes and symbolic representations, including deliberate judgments, planning for goal pursuit, and inhibition of prepotent responses. In contrast, the impulsive system operates fast on associative clusters in long-term memory that group stimuli, affective states, and behavioral responses together. At the moment of a specific decision, the success of self-control depends on the relative ability of the processes in the two systems to activate the corresponding behavioral schemas. Several boundary conditions have been proposed to moderate the relative strengths of the two systems [72]. For example, people are believed to behave more impulsively when their behaviors are highly habitual, when their cognitive loads are high, and when their moods are positive.

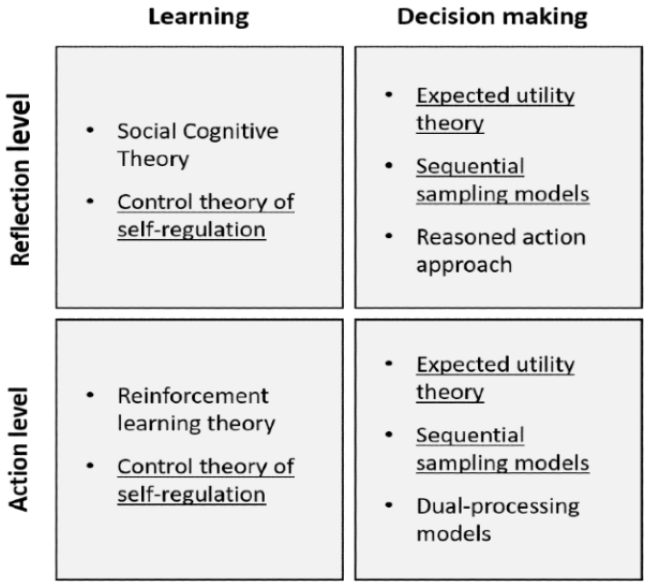
Temporal Scales Used in the Aforementioned Theories

Figure 2 summarizes the learning and decision-making theories based on the temporal scales of behavior representation. A similar distinction was made by Karoly [66], where theories at the action level were called *online* theories and theories at the reflection level were called *offline* theories.

In the learning tradition, the reinforcement learning theory clearly represents behavior at the action level, as the outcome of each specific action or decision is modeled to have concrete impacts on the frequency of repeating the same action in the future. Reinforcement learning experiments also involve repeated trials within a relatively short period (eg, a few hours). The control theory of self-regulation, because of its hierarchical structure, covers both behavioral processes at the reflection level and the action level. SCT and its processes of self-efficacy and proactive control apply mainly to behaviors at the reflection level. Although the two processes may have counterparts at a lower level, as in the control theory, Bandura's [31] focus was clearly on voluntary and deliberative human behaviors.

In the decision-making tradition, mathematical models as part of the expected utility theory and the sequential sampling approach can be equally applied to decisions at both temporal scales, as long as decisions with clearly defined choice sets are considered. As discussed earlier, theories in the reasoned action approach deal mainly with decisions at the reflection level because of its assumption of intentionality. In contrast, dual-process models are mainly intended to account for *small* daily decisions, for which both reflective and impulsive processes play a role. There is a difference between *reflective* used in dual-processing theories and what we mean by *reflection level*, which will become clear after theory integration.

Figure 2. Categorization of reviewed theories based on their theoretical traditions and temporal scales (theories that apply to both temporal scales are underlined).



Theory Integration: An Adaptive Decision-Making Framework

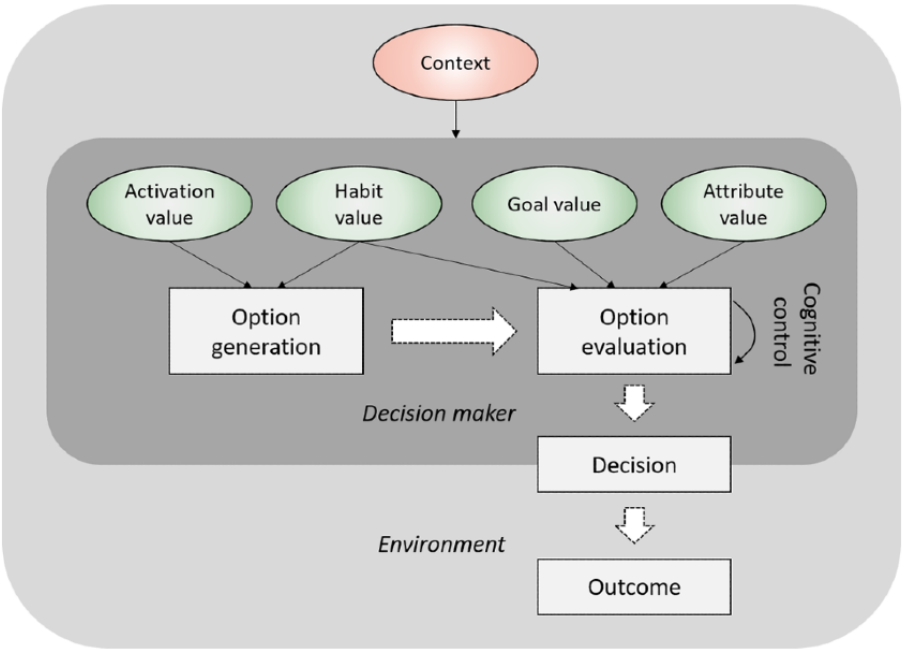
Overview

To reiterate, our goal of theory integration is to develop a unified framework that identifies and connects all relevant decision-making and learning processes at both the action and reflection levels of lifestyle behavior change. Most processes in the framework came directly from the theories reviewed earlier, but efforts were made to unify different terminologies from different theories to form a coherent framework and to tailor the framework to lifestyle behaviors. Taking dietary behavior as a primary example, the framework should explain not only how daily meal choices are made and how each decision outcome influences future choices but also how the goal of adhering to a specific diet is made and how such goals are evaluated. The following sections introduce the adaptive decision-making framework in four parts: action-level decision-making, action-level learning, reflection-level decision-making, and reflection-level adaptation.

Action-Level Decision-Making: Daily Meal Choices

Daily lifestyle decisions, such as daily meal choices, can be modeled as a two-step process—*option generation* and *option evaluation* (Figure 3). The framework assumes that when choosing a meal, different meal options must be generated or recalled by a decision maker first, before evaluations of a few options can be made to inform a final choice [73,74]. The notion of option generation has not been examined in any of the decision-making theories reviewed, probably because most of the theories are based on laboratory choice experiments, where options are simply provided by the experimenters. For lifestyle behaviors in daily environments, how choice alternatives are generated is an important question.

Figure 3. A two-step model of daily lifestyle decision making.



In general, behavioral options can be generated using three different methods. First, if an option is habitual, it will be activated when the associated cues are encountered, such as location and time (eg, lunch at the office) or a combination of contextual cues (eg, a busy Wednesday evening). Second, options may be remembered at the right moments because people intentionally try to maintain them in their prospective memory (ie, not to forget to do something in the future [74,75]). This usually happens when there is a salient goal guiding daily decisions, such as the goal of adhering to a low-carb diet. People may also intentionally associate important options with external cues so that encountering cues is likely to trigger the options [76]. Third, options can be triggered by direct external suggestions at the decision moments, for example, a coaching message from a mobile health app that recommends healthy foods [73]. Through these means, behavioral options that are sufficiently activated (eg, by passing an activation threshold) will be evaluated.

Option evaluation can be modeled as a process of comparing several options and then choosing the one with the highest goal-satisfying value. The exact computation of utilities can follow either classical expected utility models or more dynamic sequential sampling models. Here, it is sufficient to identify three main cognitive variables in the evaluation process. First, when multiple personal goals are relevant for a daily decision, these goals can be regarded as more or less important by a decision maker, thus entailing higher or lower *goal values*. For example, between the goals of living a healthy life and enjoying delicious food, a person who regards the former goal as more valuable would be more likely to choose food options for meals that satisfy their health goals.

Second, for each personal goal, a behavioral option has its *perceived attribute value* relating to that goal, which determines the total utility of the option. These attribute values are subjective beliefs held by people about the causal relationships or contingencies between choosing certain behavioral options and the realizations of personal goals. Although goal values are relatively more stable within-person, attribute values are more context-dependent and prone to changes through learning and experience. For example, the perceived taste of a particular meal option may depend on a person's momentary appetite, and it may change over time through repeated tasting of the food (ie, habituation [77]).

There is a particular challenge for making healthy decisions, as usually two distinct types of attributes are considered: an immediate hedonic aspect such as tastiness and a long-term consideration of health consequences. From a decision-making perspective, this challenge is essentially a problem of self-control [78]. According to the idea of temporal discounting in decision-making theories [79,80], as any reward from potential health improvements is delayed in time when compared with the immediate hedonic aspects, the value of the attribute healthiness is discounted before it is integrated in option evaluation [81,82]. Another reason why health aspects are often weighted less than hedonic aspects in actual decisions is that the former are more abstract concepts so they might be more difficult or take longer to be processed [83-86]. Finally, from a dual-processing perspective (eg, [54]), dietary self-control may sometimes succeed because people can voluntarily exert top-down cognitive control on the option evaluation process, especially if a momentary preference for a meal option conflicts strongly with a diet goal. It has been shown experimentally that cognitive control may either modulate the valuation process to be more in favor of healthiness rather than tastiness [87] or filter people's attention away from hedonic attributes in the early stage of option evaluation [88]. Effective top-down control depends on many contextual variables, such as motivation [89], mental fatigue [90], stress level [91], and daily affective states [92].

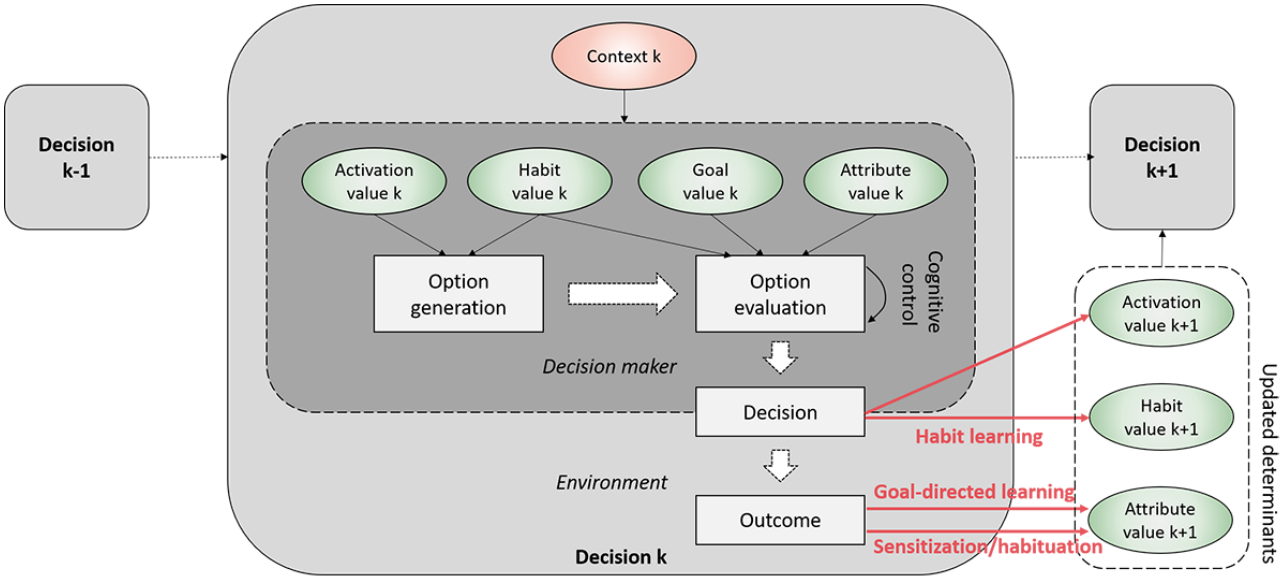
Third, *habit values* or habit strengths, which represent the history of choosing certain behavioral options, may influence the evaluation of options. As mentioned earlier, learning

experiments have shown convincingly that even when two options are provided to decision makers, habitual options are more likely to be chosen than nonhabitual options [36,39]. In sequential sampling models, the influence of habits on the dynamic process of option evaluation can be understood as positively biasing the baseline preferences for habitual options ([60], Zhang et al, unpublished data, 2021). As an intuitive example, if someone often chose fast food in the past, fast food is by default more favorable than other options when no additional deliberations are made.

Action-Level Learning and Adaptation: Developments of Eating Habits

Action-level learning processes can be added to the framework by integrating the ideas of goal-directed learning and habit learning from the reinforcement learning theory to the two-step decision-making model proposed earlier (Figure 4). First, feedback from decision outcomes to perceived attribute values represents goal-directed learning. For example, when a new canteen is opened at a workplace, employees may have initial but very uncertain beliefs about the tastes and calories of different lunch options; however, after a few weeks of trying them out, they gradually form more accurate perceptions about the options. Computationally, the updates of perceived attribute values can be done through model-based and model-free reinforcement learning algorithms (eg, temporal difference learning [37]) or Bayesian belief update [93]. For health-related attributes, because concrete decision outcomes are infrequent except for extreme cases (eg, food poisoning), it is less clear how direct learning from experience works, if it is possible at all (Gershman and Daw [94]). People's beliefs about the health consequences of different foods are more susceptible to social learning and education.

Figure 4. Action-level learning processes added to the decision-making model.



Second, there is direct feedback from decisions to habit values, as in the process of habit formation or habit learning. Although daily lunch decisions in a new canteen are driven primarily by goal-related attribute values, through repeated decisions, mental associations between frequently chosen food options and environmental cues (eg, the physical setting of the canteen or lunchtime) are gradually strengthened. These associations, as habit values, influence future decisions through both the option generation and option evaluation processes, as discussed earlier. The exact mechanism of habit learning is beyond the scope of this paper; however, it has been modeled in the literature [48,72,95,96].

Third, there is also a direct link between decisions and the activation values of options, which has been discussed much less in the learning literature. When a decision is made and the corresponding behavior is executed, the behavior execution increases the activation level of the behavioral option in memory, although such an increase has been shown to be very small empirically [72]. As discussed earlier, the dynamics of activation values are primarily memory processes and are mostly affected by physical and social stimuli in the environment.

Fourth, with repeated daily actions and exposure to the sensory outcomes of the actions, people's neurological responses to the same sensory stimuli, depending on the parameters, may either intensify (sensitization) or fade away (habituation) [97]. This action-level adaptation is especially important for eating behavior, for which the exact same food becomes less palatable over time and accordingly its consumption will decrease [77]. This explains why the large variety in the modern food industry is considered a contributor to obesity [98]. In our framework, sensitization or habituation can be represented as an additional mechanism of sensory feedback from decision outcome to attribute value, in addition to the more cognitive process of goal-directed learning.

Reflection-Level Decision-Making: Dietary Goal Setting

Action-level decision-making and learning processes cover a substantial part of what people do in their daily lives. They depict lifestyle behaviors as repeated decisions without much purpose. However, people also have moments when they reflect on their health status, contemplate possible improvements, and make concrete plans. According to the control theory [45], these

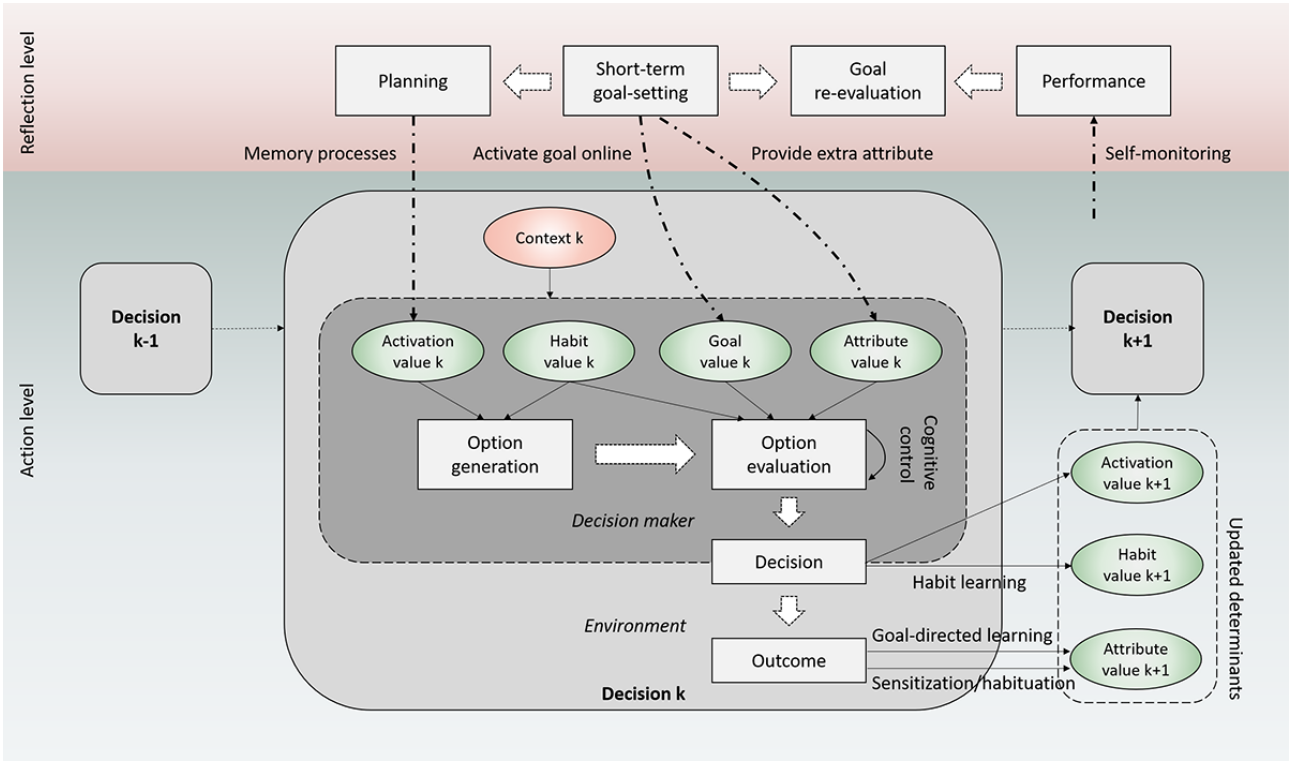
reflective processes require short-term goals that bridge people's abstract long-term goals (ie, what they pursue in their lives) and action-level daily decisions. Short-term goal setting can also be understood as a process of decision-making, albeit at the reflection level rather than the action level. The decisions made are commitments to goals that guide future daily decisions rather than overt behaviors that trigger motor programs.

Therefore, the two-step model of action-level decisions also applies to the setting of short-term goals. In selecting a dietary plan, for example, people first search for diet options that serve their long-term goals and then evaluate the options on relevant attributes, such as taste, ease of preparation, and expenses. At the reflection level, these attributes are often categorized into a few determinants, such as attitude, social norm, and perceived behavior control [29]. Nonetheless, goal setting differs from action-level decisions in some respects. First, because goal setting is less frequent than daily decisions, strong habits are unlikely to be formed to influence decision-making processes. Potential biases by habits are further reduced because people are more careful and take more time to generate goal options and evaluate them. Second, because people set goals for an extensive period of time in the future, they may form a more abstract mental construal [99], which is detached from direct sensory information and visceral attributes, such as effort and tastiness. Thus, the self-control problem for daily lifestyle decisions is less prevalent in goal setting. Third, self-efficacy plays an important role in goal setting [50]. People may carefully consider the feasibility of different diet goals by mentally simulating a series of daily dietary choices in the future.

Motivating Functions of Short-Term Goals

When short-term goals are generated, they can influence daily lifestyle decisions through both option generation and option evaluation (Figure 5). First, setting up a short-term goal can increase the activation values of desirable behavioral options through a process termed *planning*. Planning can be done through two mechanisms discussed earlier: an effortful prospective memory process (eg, rehearsing eating salads [75]) or an *implementation intention* process, that is, mentally associating a behavioral option with certain environmental cues (eg, eating an apple when watching television [76]). Second, compared with long-term goals, short-term goals are more concrete; therefore, complying with these goals brings immediate satisfaction [100]. Goal-compliance satisfaction functions as an additional attribute that competes with other hedonic attributes in the option evaluation.

Figure 5. A full representation of the adaptive decision-making framework (reflection-level processes and interactions between the 2 levels added to the previous decision-making model).



Reflection-Level Learning and Adaptation: Self-monitoring and Re-evaluation of Dietary Goals

Short-term goals must be re-evaluated in reflection moments periodically to change the goals that are too difficult, too easy, or no longer relevant. Such reflection-level adaptation processes are well described by the control theory [45] and SCT [31]. Goal re-evaluation first requires inputs from action-level processes through *self-monitoring*. Through repeated daily dietary choices, past choices and their outcomes are stored in episodic memory and are later retrieved and integrated into a mental representation of overall past performance (Figure 5). Next, discrepancy between a goal reference (eg, a dietary plan) and a performance representation is computed and used to inform the reflection-level adaptation. Depending on the size of discrepancies and other contextual factors, people may motivate themselves further to make healthy daily dietary choices to reduce the goal performance discrepancies. However,

when discrepancies are deemed too large, people may instead lower their goal standards (eg, be less strict on calorie intake) or abandon their goals altogether (eg, give up a diet). Instead, when performance matches or even exceeds current goal standards, they may proactively adjust their goal standards upward to further improve their health [31].

Mapping Digital Intervention Techniques to the Framework

We define *digital intervention techniques* as BCTs or behavior change methods that target lifestyle behaviors and are implemented in digital systems (eg, web, mobile, or wearable systems). BCTs, in turn, are generally defined as the active ingredients of interventions that can influence behaviors in desirable ways [34,101,102]. Mapping BCTs to theoretical constructs is an important exercise when trying to evaluate and enhance the effectiveness of the techniques; such mappings have been carried out previously [102,103]. Focusing on their implementations in digital intervention systems, our mapping exercise is the first to connect BCTs to theoretical constructs in a single unified framework. Specifically, we categorize and interpret digital BCTs according to targeted behavioral processes and cognitive variables in the framework. The strengths and limitations of some techniques are discussed based on the implications of the framework.

Digital Intervention Techniques Targeting Action-Level Decision-Making

As digital systems are prevalent in people's daily lives, they are well positioned to influence people's daily lifestyle decisions at the time of the decision. The ability to target action-level decisions is considered by many as a promising direction for digital lifestyle interventions, as reflected in research on *ecological momentary interventions* [1] and *just-in-time adaptive interventions* (JITAIs) [104-108]. According to our framework, there are many different ways in which digital intervention systems can influence online decision-making processes, depending on whether the techniques target option generation or option evaluation and which cognitive variables are targeted; four main categories can be distinguished.

Option-Based Techniques

Option-based techniques make certain desirable behavioral options salient but leave the evaluation of options to users themselves. When a desirable behavior is obvious but may not be constantly salient to users, digital systems can simply prompt users to actively make decisions to engage in that behavior, for example, to take breaks when overly sedentary behaviors are detected by the system [109,110]. Otherwise, it might be possible to provide users with new options that are better than those known by users themselves [73]. Finding such *attractive* options relies on a system's sensor network and smart algorithms, which potentially make it more knowledgeable than its users in a given behavioral domain or context. For example, a smart system was developed to recommend new commuting routes to users in situ to increase physical activities, based on automatic detection of users' habitual routes and Google Maps data [111].

Attribute-Based Techniques

Attribute-based techniques aim to change users' beliefs about the attribute values of options by providing health-related knowledge or facts. They are referred to as *providing information about behavior-health links* or *providing information about consequences* in the taxonomy of BCTs [101]. Given the common assumption that humans are rational decision makers, providing information about attribute values is a logical approach to behavior change and has been used extensively in traditional health education campaigns. However, attribute-based techniques alone do not guarantee successful behavior change, as attribute value is only one of many factors that influence decision-making. It is also questionable whether digital systems are better tools for providing such information when compared with human experts (eg, lifestyle coaches). Nonetheless, information about attribute values can be provided to justify the recommendations of behavioral options whenever appropriate in digital systems (eg, calorie information for different meal choices).

Goal-Based Techniques

As goal values modulate attribute values in option evaluation, activating health-related goals in the decision moments provides yet another type of intervention technique. When implemented in digital systems, they link the suggestions of concrete behavioral options with the reminder of associated short-term or long-term goals. For example, when a mobile app prompts a user to take a lunch walk, the user's goal of walking 10,000 steps a day (and the achieved steps) can be presented along with the option of taking a lunch walk.

Structure-Based Techniques

Structure-based techniques differ from previous types because they neither change the availability of options nor alter users' existing beliefs about attribute values. As they require less processing effort and are less susceptible to reactance from users than other techniques, structure-based techniques have attracted significant research interest [112,113], usually under the name of *nudging* or *choice architecture* [114,115]. For example, people can become more likely to choose the desirable options when they are presented as default options [116] or an additional option is introduced to change their perceptions of choice sets (context effects [57]). Lee et al [117] adopted the default technique to promote healthy snacking in an online environment by making healthier options the default choices. Zhang et al [118] built on a context effect called the *compromise effect* to promote physical exercise at work using a mobile app. Intensive exercise options were added to make moderate exercise options appear more achievable and thus more attractive [118].

A challenge for all JITAIs is that most lifestyle behaviors are *physical* rather than *digital* in nature. When making decisions about snacking, exercising, or toothbrushing, people do not naturally come to digital applications. In contrast, in e-commerce, for example, people are accustomed to shopping online; therefore, e-commerce sites such as Amazon never miss the opportunity to influence consumers in their decision moments. To influence lifestyle decisions at critical moments, interfaces between the information in digital systems and

people's spontaneous behaviors in the physical world need to be created. Current approaches include predicting users' spontaneous decision-making moments using sensor networks (eg, predicting *about-to-eat* moments [119]) and initiating decisions when interventions are predicted by the system to be most valuable (eg, predicting stressful moments [120]). This challenge will continue to stimulate new intelligent digital solutions and at the same time debates on the associated ethical implications [121].

Digital Intervention Techniques Targeting Action-Level Learning

Digital intervention techniques targeting action-level learning processes operate in between rather than at decision moments. The goal is to support either goal-directed learning or the formation of healthy habits. If these techniques are effective, the cognitive variables that influence decision-making will be in a health-promoting state, so that users are expected to maintain the learned healthy behaviors without continuous intervention by digital systems.

A main challenge for lifestyle behavior change is learning the causal relationships between one's behaviors and health consequences, as these consequences are usually delayed. As discussed, researchers have speculated on the role of episodic memory in tracking internal and external events to support this type of learning [94]. In this regard, the self-tracking function of many digital systems can support learning by externalizing the user's memory systems [98]. Behavioral and contextual data can be objectively recorded and reviewed later by users when consequential health events occur. As self-tracking studies mostly focus on the effectiveness of the technology as a whole, evidence regarding its specific role in supporting goal-directed learning is lacking [122]. Some interview data indicated that users of self-tracking systems believed that they acquired knowledge about behavior-health links through self-tracking technology [123,124].

Instead of directly supporting the learning of health consequences, another popular approach is to provide extra rewards that may reinforce desirable behaviors. In the so-called gamification systems, the most common extra rewards are virtual rewards, such as points, badges, or rankings in leaderboards [125-127]. These virtual rewards are expected to steer users to healthy behavioral options by competing with the inherent hedonic values of many unhealthy behaviors.

Despite its popularity, the effectiveness of virtual rewards in their simplest forms is questionable, as empirical studies found no positive effects in several health domains, such as physical activity [128] and sexual protection behavior [129]. Moreover, users in one study perceived such virtual rewards implemented in an exercise-promoting app as *not motivating* or even *unnecessary* [130]. Our framework implies that the problem with virtual rewards is not in the learning of the contingencies between behavioral responses and rewards but in the corresponding goal values of these rewards: the goal values of virtual rewards are often low, when compared with other hedonic attributes, such as tastiness and reduced effort. Future research on gamification should focus on making virtual rewards more goal relevant and meaningful [131], for example, by embedding them as a game mechanic that users care about [132], using tangible rather than intangible rewards [133,134], or making the rewards socially meaningful [135,136].

Another technique in this category is habit formation support, usually by reminding users about a new and desirable behavioral option. This is especially valuable at the beginning of habit formation when new options are not always remembered by users themselves. Unlike the technique of suggesting options at decision moments, reminders that support habit formation are sent offline and according to time-based schedule (eg, once every morning). They do not persuade users to act immediately but to increase the activation values of certain options so that they are more likely to be generated when decision moments arrive. Reminders have been widely used and have been shown to be effective in domains where forgetting is the main obstacle for behavior change [137,138]. More research is warranted to understand its value in changing more complex lifestyle behaviors when the activation value is one of several cognitive variables.

Digital Intervention Techniques Targeting Reflection-Level Decision-Making

Setting up a short-term goal as a reflection-level decision-making process is often the starting point of self-directed behavior change [139]. Without external interventions, goal setting can be triggered under specific conditions, for example, when someone has learned new health-related knowledge (eg, become aware of the risk of smoking) or has experienced a sudden change in their health status (eg, being diagnosed with diabetes). Thus, a straightforward intervention technique is to proactively prompt users to set up new goals to improve their lifestyles. In many digital systems, following a goal-setting prompt, a user can choose a goal and then record it in the system, which allows the system to remind the user of the goal when needed.

As goal setting is a decision-making process, most techniques discussed in the section on targeting action-level decision-making also apply to goal setting, including option-, attribute-, and structure-based techniques. As a particularly promising direction, digital systems may use their data-gathering power and artificial intelligence to recommend novel and attractive options for short-term goals [73]. To address the subtlety and complexity of goal setting in the health domain, the systems need to personalize options based on users' abilities [140] and based on their unique life experiences and personal context [141-144]. In the future, the difficult task of setting up challenging, motivating, yet realistic goals may indeed be transferred from people to intelligent intervention systems, at least in part.

After the goal-setting step, digital systems can go further to support the planning phase that connects short-term goals to daily decisions in the future. A simple technique is to prompt users to make concrete plans in the system, for example, by adding activities to a calendar. Data provided by users allow digital systems to check user adherence and send reminders when necessary. In addition to this time-based planning technique, digital systems may encourage users to use the event-based planning technique of implementation intention

[76]. Implementation intention has been shown to be effective in the health domain [145,146], and it has also been implemented in digital interventions where no human instructions are required [147,148]. A recent system even uses sensor data to automatically generate *if-then* rules adapted to the living contexts of individual users [149].

Digital Intervention Techniques Targeting Reflection-Level Adaptation

At the level of reflection-based adaptation, providing behavioral feedback to users to support self-monitoring is the most commonly used BCT in digital systems [19,150-153]. Technically, with the development of increasingly powerful sensors, digital systems are able to track lifestyle behaviors and related variables more accurately and in greater detail than people's own memories. Moreover, these systems can transform the rich raw data into numerical or visual information (eg, weekly summary of step count) to facilitate better comparison with short-term goal references [122].

Although self-monitoring as a general BCT has been identified as effective [154], the evaluation of this technique in digital systems has yielded mixed results [153,155]. The evaluation is also impeded by the lack of high-quality studies and a lack of focus on self-monitoring per se [122]. It is evident that the abundance of self-tracking devices has not solved the problem of lifestyle behavior changes. From an evolutionary perspective, as people's natural self-monitoring function has existed long before the existence of digital systems and quantitative data, it is not self-evident that technology-enhanced information would lead to better functioning. A recent study indicates that some self-tracking users may have an exaggerated focus on numeric feedback as the replacement of bodily experience as feedback, potentially leading to negative consequences such as rumination [156]. The bottom line is that even if technology-enhanced self-monitoring is beneficial to some extent, our framework implies that it is only one step in reflection-level adaptation. Future research should investigate how digital systems can also support reflective processes that immediately follow self-monitoring, including the comparison between goal references and monitored performance and the adjustments of goals and behaviors.

Discussion

Overview

Understanding and changing lifestyle behaviors in the digital age require a theoretical perspective that combines decision-making and learning and a representation of behavior at the level of both daily decisions and episodic reflections. These two requirements have guided our review of individual theories and their integration, and the outcome is temporally fine-grained, dynamic, and process-oriented theoretical framework of lifestyle behavior change. Through a mapping exercise, we also linked common digital intervention techniques to behavioral processes and cognitive constructs in the framework.

Theoretical Contributions and Comparisons With Previous Integration Works

The primary objective of developing an adaptive decision-making framework is to address the mismatch between theory and digital intervention in terms of temporal granularity [6]. This was done by considering lifestyle behaviors at two different timescales, one representing the individual daily decisions or actions and one grouping the repeated daily decisions into a larger episode and incorporating self-regulatory processes. This 2-level representation adds value over previous integration attempts that were based on the stage model of change [30], such as the computerized behavior intervention (COMBI) model [157,158] and the i-Change Model [159]. Although the COMBI and i-Change Model postulate a more general process of behavior change (eg, through contemplation, preparation, action, and maintenance), our framework allows one to zoom to a finer level of granularity to explain how repeated daily actions, with the help of reflection-level regulatory processes, lead to maintenance. With time-intensive behavioral monitoring becoming more accessible, our framework compliments earlier work by motivating future intelligent systems that update users' behavioral and cognitive states after every daily lifestyle decision (eg, computing self-efficacy [160]). In our own work, we proposed a system that computes users' habit strengths of toothbrushing based on sensor-measured behaviors [161].

There are previous frameworks that are more similar to our adaptive decision-making framework when it comes to behavior representation. Both PRIME (Plan, Responses, Impulses, Motives, and Evaluations) theory [162] and Temporal Self-Regulation Theory [163] model behavior change as a continuous process rather than a series of discrete stages. However, our framework explicitly distinguishes between the two distinct levels of lifestyle behaviors and the different timescales involved. Note that similar to many other dichotomies used in psychology (eg, impulsive vs reflective, unconscious vs conscious), although our 2-level dichotomy is a simplification of a potential continuum of processes, it is still useful for developing new theories, empirical research, and applications. The conceptual distinction is important because it enables our framework to represent lifestyle behaviors at the same temporal granularity with time-intensive behavioral data while incorporating cognitive processes that are detached from daily decisions (eg, goal setting or planning).

A second limitation in the current literature is the lack of dynamic processes in traditional behavior change theories [6]. By integrating theories from both learning and decision-making tradition, our framework depicts a dynamic bidirectional relationship between behaviors and cognitive variables that influence behaviors. The framework complements previous frameworks that focused exclusively on learning processes, such as the framework of evolutionary learning processes [97] and Action Change Theory [164]. More broadly, we believe that the need to capture the complexities of lifestyle behaviors for designing better digital interventions provides a strong and timely motivation to integrate decision-making and learning theories in basic psychological research [165].

Furthermore, this study may stimulate some rethinking about the popular dual-processing models, in which health behaviors are assumed to be driven by two distinct forces, one reflective and one impulsive [72]. In our view, such dichotomous categorization of diverse processes and constructs might be too coarse for a full understanding of the dynamic lifestyle behavior change process. The adaptive decision-making framework suggests several dualities. First, there is a contrast between long-term health benefits and immediate hedonic rewards in option evaluation, where effortful cognitive control is required to battle impulses. Second, goal-directed evaluation based on attributes (both long-term and short-term) competes with the influences of habits. This has been discussed extensively in the learning literature as a dual action control by goals and habits [166]. Third, the faster processes at the action level can certainly be contrasted with the more deliberative processes at the reflection level, which also operate on a much slower timescale. Note that this distinction between action and reflection levels certainly reminisces the classical distinction between motivation and volition [167] and the related Rubicon model of action phases [168].

Finally, the process-oriented nature of our framework makes it an ideal scaffold for developing new dynamic computational models envisioned by many researchers [6,104,169,170]. The processes and mechanisms in the framework are described at a level of specificity that allows their transformation into computational models by introducing additional assumptions and formal algorithms. For example, key cognitive variables are defined for option evaluation, but the exact computational process of how these variables are integrated to produce decisions is left open to different modeling possibilities [44,171]. Similarly, although the framework acknowledges the joint influence of habit and goal-directed control, the exact arbitration between the two is subjected to different computational accounts [96,172,173].

Added Value to the Synergy Between Theory and Digital Intervention

The adaptive decision-making framework was developed with the aim of narrowing the gap between behavior change theories and digital intervention applications. As a first step, the framework provides a summary of the main theoretical ideas in psychology relevant to applied behavior change research. As such, it can be used as a reference if practitioners want to read more about specific theories and computational models. The integration of both traditional and more cutting-edge theories in the framework should help practitioners to see familiar theories (eg, self-efficacy and goal setting) in a new light and hopefully motivate them to experiment with new ideas (eg, habit formation or sequential sampling models of decision-making). What is especially promising is the prospect of designing intelligent close-loop intervention systems based on computational models informed by our framework. In the current best practice, behavior change determinants and corresponding BCTs are identified and then carefully translated into functions in digital applications; however, their effects on behavior are simply speculated (open loop). By relying on computational models, intelligent systems can monitor behaviors that are targeted and use behavioral data to update the cognitive variables of users to deliver JITAIs for each individual user [107,174]. Examples of such close-loop systems can be found in recent works by other researchers [160] and ourselves [161].

Moreover, our framework's emphasis on behavioral processes and their corresponding digital intervention techniques should contribute to the identification, implementation, and evaluation of intervention techniques. First, similar to an intervention mapping approach [102], our framework makes a clear distinction between people's behavioral processes and the techniques that may influence these processes, which is not always made in other coding taxonomies. For example, *habit formation* has been considered a BCT [101]; however, it is essentially a behavioral process that also operates without interventions and is driven by multiple lower-level processes. It is more informative and actionable for system designers if they are informed about the specific processes underlying habit formation and how they can be changed rather than simply implementing a technique called habit formation. Second, by mapping digital intervention techniques to behavioral processes in our framework, it should become clear that a technically well-defined function often targets multiple distinct behavior processes. For example, self-tracking may increase users' knowledge about behavior-health links but may also support self-monitoring [122]. We argue that evaluation research (eg, review and meta-analysis) should focus more on the effects that specific intervention techniques have on individual processes rather than the effectiveness of broadly defined categories of technologies (eg, *feedback system* [155]), to gain a better understanding of how and why certain intervention techniques work. Third, when combining multiple intervention techniques within a single digital system, our framework can inform designers about whether the techniques target complimentary processes and constructs or the same process and construct. In the latter case, the combination of techniques as a package may not necessarily be more effective than its components, and more careful analysis is needed. For example, as implementation intention and *just-in-time* reminder both increase the activation values of desirable options, it is questionable whether combining them would yield better results (Luszczynska et al [147]).

The adaptive decision-making framework may also help to encourage the use of digital lifestyle intervention systems to advance basic human behavior sciences. We have emphasized in this paper the unique opportunities of *bringing psychology laboratories to the real world* [175,176] and exploiting the time-intensive and ecologically valid behavior data generated by digital systems [9-11]. The integration of fundamental psychological processes in the framework, such as reinforcement learning and sequential sampling models of decision-making, can increase the awareness of scientists in these fields to the practical value of their research and the great potential value of using digital systems in the field as data collection tools for *fundamental* social science research. We also hope that by summarizing and structuring the theoretical landscape for digital system designers, they can find their potential collaborations with behavioral scientists more efficiently. Finally, the clear mapping between intervention techniques and the processes and constructs in the framework makes it easier to search for the required data and manipulations to be used for theory testing.

Scope and Limitation of the Framework

Given the ambitious goal of the framework to incorporate a wide range of theoretical traditions and to connect to the full gamut of digital intervention techniques, it is important to discuss the scope and limitations of our framework. First, the adaptive decision-making framework is not a new theory in itself or a model to be directly tested or falsified in a strong empirical sense. It is a framework that integrates existing theoretical ideas into a novel representation of lifestyle behaviors. In other words, it identifies relevant explananda in the course of lifestyle behavior change and provides explanations based on the most recent theoretical advances available. The usefulness of the framework should be judged by whether it succeeds in informing new computational models and intelligent intervention systems in the future, and predictions derived from the framework and its associated computational models should be rigorously tested using empirical data.

Second, proposing an integrated framework is not meant to discourage the use of individual theories in digital interventions. The adaptive decision-making framework is a framework of basic behavioral and cognitive processes in lifestyle behavior changes that are generalizable to a wide range of behavioral domains. However, there is also large heterogeneity across different behavioral domains and target populations, in terms of which processes in the framework are more critical and which variables or parameters are more changeable. For this reason, specialized theories are always needed, even if future theoretical advances might allow a single unified theory of basic psychological mechanisms. Therefore, it makes perfect sense for digital intervention systems to focus on one or a few processes or to target only a small set of variables for change based on domain-specific theories. The theoretical scope of the framework itself is focused on explaining individual lifestyle behaviors. It does not address interactions between individuals or larger socioeconomic processes.

Third, the adaptive decision-making framework is a theoretical framework of behavior change but not a framework of digital intervention systems. We have discussed the educational and heuristic value of the framework for digital intervention designers to help them understand and apply theories. It can potentially also motivate and facilitate the development of intelligent intervention systems that model user behaviors and cognitive states [161,174]. However, the framework should not be considered as a *cookbook* in the sense of prescribing specific design choices or requirements in specific interventions.

Conclusions

We developed an adaptive decision-making framework in the hope that it will benefit behavior change theorists and digital system designers and, most importantly, facilitate better communication between the two communities. A stronger synergy will potentially help bring us closer to a future where digital systems live up to their potential to promote healthy lifestyles at scale. In the meantime, a wider adoption of more effective and theory-driven digital interventions will offer ample opportunities for building and testing new theories of human behavior.

Conflicts of Interest

None declared.

References

1. Heron KE, Smyth JM. Ecological momentary interventions: incorporating mobile technology into psychosocial and health behaviour treatments. Br J Health Psychol 2010 Feb;15(Pt 1):1-39 [FREE Full text] [doi: 10.1348/135910709X466063] [Medline: 19646331]
2. IJsselsteijn WI, De Kort Y, Midden C, Eggen B, van den Hoven E. Persuasive technology for human well-being: setting the scene. In: Persuasive Technology. Berlin: Springer; 2006:18-19.
3. Kaplan RM, Stone AA. Bringing the laboratory and clinic to the community: mobile technologies for health promotion and disease prevention. Annu Rev Psychol 2013 Jan 03;64(1):471-498. [doi: 10.1146/annurev-psych-113011-143736] [Medline: 22994919]
4. Klasnja P, Pratt W. Healthcare in the pocket: mapping the space of mobile-phone health interventions. J Biomed Inform 2012 Feb;45(1):184-198 [FREE Full text] [doi: 10.1016/j.jbi.2011.08.017] [Medline: 21925288]
5. Michie S, West R. Behaviour change theory and evidence: a presentation to Government. Health Psychol Rev 2013 Mar;7(1):1-22. [doi: 10.1080/17437199.2011.649445]
6. Riley WT, Rivera DE, Atienza AA, Nilsen W, Allison SM, Mermelstein R. Health behavior models in the age of mobile interventions: are our theories up to the task? Transl Behav Med 2011 Mar;1(1):53-71 [FREE Full text] [doi: 10.1007/s13142-011-0021-7] [Medline: 21796270]
7. Patel MS, Asch DA, Volpp KG. Wearable devices as facilitators, not drivers, of health behavior change. J Am Med Assoc 2015 Feb 03;313(5):459-460. [doi: 10.1001/jama.2014.14781] [Medline: 25569175]
8. Walsh JC, Groarke JM. Integrating behavioral science with mobile (mHealth) technology to optimize health behavior change interventions. Eur Psychol 2019 Jan;24(1):38-48. [doi: 10.1027/1016-9040/a000351]
9. Dunton GF, Atienza AA. The need for time-intensive information in healthful eating and physical activity research: a timely topic. J Am Diet Assoc 2009 Jan;109(1):30-35. [doi: 10.1016/j.jada.2008.10.019] [Medline: 19103320]

10. Moller AC, Merchant G, Conroy DE, West R, Hekler E, Kugler KC, et al. Applying and advancing behavior change theories and techniques in the context of a digital health revolution: proposals for more effectively realizing untapped potential. J Behav Med 2017 Feb;40(1):85-98 [FREE Full text] [doi: 10.1007/s10865-016-9818-7] [Medline: 28058516]
11. Rothman AJ. "Is there nothing more practical than a good theory?": why innovations and advances in health behavior change will arise if interventions are used to test and refine theory. Int J Behav Nutr Phys Act 2004 Jul 27;1(1):11 [FREE Full text] [doi: 10.1186/1479-5868-1-11] [Medline: 15279674]
12. Prestwich A, Webb TL, Conner M. Using theory to develop and test interventions to promote changes in health behaviour: evidence, issues, and recommendations. Curr Opin Psychol 2015 Oct;5:1-5. [doi: 10.1016/j.copsyc.2015.02.011]
13. Al-Durra M, Torio M, Cafazzo JA. The use of behavior change theory in internet-based asthma self-management interventions: a systematic review. J Med Internet Res 2015 Apr 02;17(4):e89 [FREE Full text] [doi: 10.2196/jmir.4110] [Medline: 25835564]
14. Conroy DE, Yang C, Maher JP. Behavior change techniques in top-ranked mobile apps for physical activity. Am J Prev Med 2014 Jun;46(6):649-652. [doi: 10.1016/j.amepre.2014.01.010] [Medline: 24842742]
15. Dahlke DV, Fair K, Hong YA, Beaudoin CE, Pulczinski J, Ory MG. Apps seeking theories: results of a study on the use of health behavior change theories in cancer survivorship mobile apps. JMIR Mhealth Uhealth 2015 Mar 27;3(1):e31 [FREE Full text] [doi: 10.2196/mhealth.3861] [Medline: 25830810]
16. Monteiro-Guerra F, Rivera-Romero O, Fernandez-Luque L, Caulfield B. Personalization in real-time physical activity coaching using mobile applications: a scoping review. IEEE J Biomed Health Inform 2020 Jun;24(6):1738-1751. [doi: 10.1109/jbhi.2019.2947243]
17. Cho YM, Lee S, Islam SMS, Kim SY. Theories applied to m-Health interventions for behavior change in low- and middle-income countries: a systematic review. Telemed J E Health 2018 Oct;24(10):727-741 [FREE Full text] [doi: 10.1089/tmj.2017.0249] [Medline: 29437546]
18. Chib A, Lin SH. Theoretical advancements in mHealth: a systematic review of mobile apps. J Health Commun 2018 Nov 19;23(10-11):909-955. [doi: 10.1080/10810730.2018.1544676] [Medline: 30449261]
19. Winter SJ, Sheats JL, King AC. The use of behavior change techniques and theory in technologies for cardiovascular disease prevention and treatment in adults: a comprehensive review. Prog Cardiovasc Dis 2016;58(6):605-612 [FREE Full text] [doi: 10.1016/j.pcad.2016.02.005] [Medline: 26902519]
20. Davis R, Campbell R, Hildon Z, Hobbs L, Michie S. Theories of behaviour and behaviour change across the social and behavioural sciences: a scoping review. Health Psychol Rev 2015;9(3):323-344 [FREE Full text] [doi: 10.1080/17437199.2014.941722] [Medline: 25104107]
21. Webb TL, Joseph J, Yardley L, Michie S. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. J Med Internet Res 2010 Feb 17;12(1):e4 [FREE Full text] [doi: 10.2196/jmir.1376] [Medline: 20164043]
22. Glanz K, Bishop DB. The role of behavioral science theory in development and implementation of public health interventions. Annu Rev Public Health 2010;31:399-418. [doi: 10.1146/annurev.publhealth.012809.103604] [Medline: 20070207]
23. Taylor N, Conner M, Lawton R. The impact of theory on the effectiveness of worksite physical activity interventions: a meta-analysis and meta-regression. Health Psychol Rev 2012 Mar;6(1):33-73. [doi: 10.1080/17437199.2010.533441]
24. Prestwich A, Sniehotta FF, Whittington C, Dombrowski SU, Rogers L, Michie S. Does theory influence the effectiveness of health behavior interventions? Meta-analysis. Health Psychol 2014 May;33(5):465-474. [doi: 10.1037/a0032853] [Medline: 23730717]
25. Garnett C, Crane D, Brown J, Kaner E, Beyer F, Muirhead C, et al. Reported theory use by digital interventions for hazardous and harmful alcohol consumption, and association with effectiveness: meta-regression. J Med Internet Res 2018 Feb 28;20(2):e69 [FREE Full text] [doi: 10.2196/jmir.8807] [Medline: 29490895]
26. Gardner B, Wardle J, Poston L, Croker H. Changing diet and physical activity to reduce gestational weight gain: a meta-analysis. Obes Rev 2011 Jul;12(7):602-620. [doi: 10.1111/j.1467-789X.2011.00884.x] [Medline: 21521451]
27. Hagger MS, Weed M. DEBATE: do interventions based on behavioral theory work in the real world? Int J Behav Nutr Phys Act 2019 Apr 25;16(1):36 [FREE Full text] [doi: 10.1186/s12966-019-0795-4] [Medline: 31023328]
28. Gainforth HL, West R, Michie S. Assessing connections between behavior change theories using network analysis. Ann Behav Med 2015 Oct 22;49(5):754-761. [doi: 10.1007/s12160-015-9710-7] [Medline: 26002108]
29. Prochaska JO, DiClemente CC. Transtheoretical therapy: toward a more integrative model of change. Psychotherapy: Theory, Res, and Practice 1982;19(3):276-288. [doi: 10.1037/h0088437]
30. Ajzen I. The theory of planned behavior. Organ Behav Hum Decis Process 1991 Dec;50(2):179-211. [doi: 10.1016/0749-5978(91)90020-T]
31. Bandura A. Human agency in social cognitive theory. Am Psychol 1989;44(9):1175-1184. [doi: 10.1037/0003-066X.44.9.1175]
32. Scholz U. It's time to think about time in health psychology. Appl Psychol Health Well Being 2019 Jul 10;11(2):173-186. [doi: 10.1111/aphw.12156] [Medline: 30972951]
33. Rothman AJ. It's about time: answering the call for greater precision in research and practice. Appl Psychol Health Well Being 2019 Jul 10;11(2):191-197. [doi: 10.1111/aphw.12158] [Medline: 30972957]

34. Michie S, West R, Spring B. Moving from theory to practice and back in social and health psychology. Health Psychol 2013 May;32(5):581-585. [doi: 10.1037/a0030205] [Medline: 23646841]
35. Postman L. The history and present status of the law of effect. Psychol Bull 1947 Nov;44:489-563. [doi: 10.1037/h0057716] [Medline: 20273420]
36. Yin HH, Knowlton BJ. The role of the basal ganglia in habit formation. Nat Rev Neurosci 2006 Jun;7(6):464-476. [doi: 10.1038/nrn1919] [Medline: 16715055]
37. Sutton RS, Barto AG. Reinforcement learning: an introduction. Cambridge, Mass: MIT press; 2018:A.
38. Thorndike EL. The fundamentals of learning. New York: Teachers College Bureau of Publications; 1932.
39. Dickinson A. Actions and habits: the development of behavioural autonomy. Phil Trans R Soc Lond B 1997 Jan 13;308(1135):67-78. [doi: 10.1098/rstb.1985.0010]
40. Wood W, Neal DT. A new look at habits and the habit-goal interface. Psychol Rev 2007;114(4):843-863. [doi: 10.1037/0033-295x.114.4.843]
41. Wood W, Rünger D. Psychology of habit. Annu Rev Psychol 2016;67:289-314. [doi: 10.1146/annurev-psych-122414-033417] [Medline: 26361052]
42. Verplanken B. Psychology of habit. Cham, Switzerland: Springer; 2018.
43. Gardner B. A review and analysis of the use of 'habit' in understanding, predicting and influencing health-related behaviour. Health Psychol Rev 2015 Jan 21;9(3):277-295 [FREE Full text] [doi: 10.1080/17437199.2013.876238] [Medline: 25207647]
44. Kanfer FH, Karoly P. Self-control: a behavioristic excursion into the lion's den. Behavior Therapy 1972 Jul;3(3):398-416. [doi: 10.1016/s0005-7894(72)80140-0]
45. Carver CS, Scheier MF. Control theory: a useful conceptual framework for personality–social, clinical, and health psychology. Psychol Bull 1982;92(1):111-135. [doi: 10.1037/0033-2909.92.1.111]
46. Powers WT. Feedback: beyond behaviorism. Science 1973 Jan 26;179(4071):351-356. [doi: 10.1126/science.179.4071.351] [Medline: 4682961]
47. Norman DA, Shallice T. Attention to action. In: Consciousness and self-regulation. Boston: Springer; 1986:1-18.
48. Psarra I. A bounded rationality model of short and long-term dynamics of activity-travel behavior (Bouwstenen 213 ed). Eindhoven, NL: Technische Universiteit Eindhoven; 2016.
49. Bandura A. Social learning theory. Morristown, NJ: General Learning Press; 1971. URL: https://www.worldcat.org/title/social-learning-theory/oclc/691872 [accessed 2021-03-29]
50. Bandura A. Self-efficacy mechanism in human agency. Am Psychol 1982;37(2):122-147. [doi: 10.1037/0003-066X.37.2.122]
51. Bandura A, McDonald FJ. Influence of social reinforcement and the behavior of models in shaping children's moral judgment. J Abnorm Soc Psychol 1963;67(3):274-281. [doi: 10.1037/h0044714] [Medline: 14054361]
52. Bandura A, Wood R. Effect of perceived controllability and performance standards on self-regulation of complex decision making. J Pers Soc Psychol 1989 May;56(5):805-814. [doi: 10.1037//0022-3514.56.5.805] [Medline: 2724068]
53. Oppenheimer DM, Kelso E. Information processing as a paradigm for decision making. Annu Rev Psychol 2015 Jan 03;66(1):277-294. [doi: 10.1146/annurev-psych-010814-015148] [Medline: 25559114]
54. Savage LJ. The foundations of statistics. New York: Wiley; 1954.
55. von Neumann J, Morgenstern O. Theory of games and economic behavior, 2nd rev. Princeton, NJ: Princeton University Press; 1947.
56. Busemeyer JR, Townsend JT. Decision field theory: a dynamic-cognitive approach to decision making in an uncertain environment. Psychol Rev 1993;100(3):432-459. [doi: 10.1037//0033-295x.100.3.432]
57. Bhatia S. Associations and the accumulation of preference. Psychol Rev 2013 Jul;120(3):522-543. [doi: 10.1037/a0032457] [Medline: 23607600]
58. Tversky A. Elimination by aspects: a theory of choice. Psychol Rev 1972;79(4):281-299. [doi: 10.1037/h0032955]
59. Tversky A, Kahneman D. The framing of decisions and the psychology of choice. Science 1981 Jan 30;211(4481):453-458. [doi: 10.1126/science.7455683] [Medline: 7455683]
60. Roe RM, Busemeyer JR, Townsend JT. Multialternative decision field theory: a dynamic connectionst model of decision making. Psychol Rev 2001;108(2):370-392. [doi: 10.1037//0033-295x.108.2.370]
61. Ratcliff R, Rouder JN. Modeling response times for two-choice decisions. Psychol Sci 2016 May 06;9(5):347-356. [doi: 10.1111/1467-9280.00067]
62. Fishbein M, Ajzen I. Belief, attitude, intention, and behavior: an introduction to theory and research. Boston: Addison-Wesley; 1975:A.
63. Head KJ, Noar SM. Facilitating progress in health behaviour theory development and modification: the reasoned action approach as a case study. Health Psychol Rev 2014 Mar 18;8(1):34-52. [doi: 10.1080/17437199.2013.778165] [Medline: 25053006]
64. Ajzen I, Madden TJ. Prediction of goal-directed behavior: attitudes, intentions, and perceived behavioral control. J Exp Soc Psychol 1986 Sep;22(5):453-474. [doi: 10.1016/0022-1031(86)90045-4]
65. Janz NK, Becker MH. The health belief model: a decade later. Health Educ Q 1984 Jan 01;11(1):1-47. [doi: 10.1177/109019818401100101] [Medline: 6392204]

66. Karoly P. Mechanisms of self-regulation: a systems view. Annu Rev Psychol 1993 Jan;44(1):23-52. [doi: 10.1146/annurev.ps.44.020193.000323]
67. De Bruijn GJ, Kremers SPJ, De Vet E, De Nooijer J, Van Mechelen M, Brug J. Does habit strength moderate the intention–behaviour relationship in the Theory of Planned Behaviour? The case of fruit consumption. Psychol Health 2007 Dec;22(8):899-916. [doi: 10.1080/14768320601176113]
68. Evans JSBT. Dual-processing accounts of reasoning, judgment, and social cognition. Annu Rev Psychol 2008 Jan;59(1):255-278. [doi: 10.1146/annurev.psych.59.103006.093629] [Medline: 18154502]
69. Kahneman D. A perspective on judgment and choice: mapping bounded rationality. Am Psychol 2003;58(9):697-720. [doi: 10.1037/0003-066x.58.9.697]
70. Strack F, Deutsch R. Reflective and impulsive determinants of social behavior. Pers Soc Psychol Rev 2004 Dec 21;8(3):220-247. [doi: 10.1207/s15327957pspr0803_1] [Medline: 15454347]
71. Hofmann W, Friese M, Wiers RW. Impulsive versus reflective influences on health behavior: a theoretical framework and empirical review. Health Psychol Rev 2008 Sep;2(2):111-137. [doi: 10.1080/17437190802617668]
72. Hofmann W, Friese M, Strack F. Impulse and self-control from a dual-systems perspective. Perspect Psychol Sci 2009 Mar;4(2):162-176. [doi: 10.1111/j.1745-6924.2009.01116.x] [Medline: 26158943]
73. Kamphorst B, Kalis A. Why option generation matters for the design of autonomous e-coaching systems. AI & Soc 2014 Jan 17;30(1):77-88. [doi: 10.1007/s00146-013-0532-5]
74. Tobias R. Changing behavior by memory aids: a social psychological model of prospective memory and habit development tested with dynamic field data. Psychol Rev 2009 Apr;116(2):408-438. [doi: 10.1037/a0015512] [Medline: 19348548]
75. McDaniel MA, Einstein GO. Strategic and automatic processes in prospective memory retrieval: a multiprocess framework. Appl Cognit Psychol 2001;14(7):127-144. [doi: 10.1002/acp.775]
76. Gollwitzer PM. Implementation intentions: strong effects of simple plans. Am Psychol 1999;54(7):493-503. [doi: 10.1037/0003-066X.54.7.493]
77. Epstein LH, Temple JL, Roemmich JN, Bouton ME. Habituation as a determinant of human food intake. Psychol Rev 2009 Apr;116(2):384-407 [FREE Full text] [doi: 10.1037/a0015074] [Medline: 19348547]
78. Berkman ET, Hutcherson CA, Livingston JL, Kahn LE, Inzlicht M. Self-control as value-based choice. Curr Dir Psychol Sci 2017 Oct 09;26(5):422-428 [FREE Full text] [doi: 10.1177/0963721417704394] [Medline: 29335665]
79. Frederick S, Loewenstein G, O'donoghue T. Time discounting and time preference: a critical review. J Econ Lit 2002 Jun;40(2):351-401. [doi: 10.1257/jel.40.2.351]
80. Green L, Myerson J. A discounting framework for choice with delayed and probabilistic rewards. Psychol Bull 2004 Sep;130(5):769-792 [FREE Full text] [doi: 10.1037/0033-2909.130.5.769] [Medline: 15367080]
81. Chapman GB. Temporal discounting and utility for health and money. J Experimental Psychol: Learning, Memory, and Cognition 1996;22(3):771-791. [doi: 10.1037//0278-7393.22.3.771]
82. Story GW, Vlaev I, Seymour B, Darzi A, Dolan RJ. Does temporal discounting explain unhealthy behavior? A systematic review and reinforcement learning perspective. Front Behav Neurosci 2014 Mar 12;8:76 [FREE Full text] [doi: 10.3389/fnbeh.2014.00076] [Medline: 24659960]
83. Maier SU, Beharelle AR, Polanía R, Ruff CC, Hare TA. Dissociable mechanisms govern when and how strongly reward attributes affect decisions. Nat Hum Behav 2020 Sep;4(9):949-963. [doi: 10.1038/s41562-020-0893-y] [Medline: 32483344]
84. Sullivan NJ, Huettel SA. Dietary self-control depends on the latency and rate of information accumulation. Cold Spirng Harbor Laboratory 2018:1 [FREE Full text] [doi: 10.1101/465393]
85. Sullivan N, Hutcherson C, Harris A, Rangel A. Dietary self-control is related to the speed with which attributes of healthfulness and tastiness are processed. Psychol Sci 2015 Feb 16;26(2):122-134 [FREE Full text] [doi: 10.1177/0956797614559543] [Medline: 25515527]
86. Zhang C, Willemsen M, Lakens D. Can mouse-tracking reveal attribute processing speeds in dietary self-control? Commentary on Sullivan et al. (2015) and Lim et al. (2018) with a simulation study. Psyarxiv Prints 2018:1-19. [doi: 10.31234/osf.io/725vp]
87. Hare TA, Camerer CF, Rangel A. Self-control in decision-making involves modulation of the vmPFC valuation system. Science 2009 May 01;324(5927):646-648 [FREE Full text] [doi: 10.1126/science.1168450] [Medline: 19407204]
88. Harris A, Hare T, Rangel A. Temporally dissociable mechanisms of self-control: early attentional filtering versus late value modulation. J Neurosci 2013 Nov 27;33(48):18917-18931. [doi: 10.1523/jneurosci.5816-12.2013]
89. Inzlicht M, Schmeichel BJ. What is ego depletion? Toward a mechanistic revision of the resource model of self-control. Perspect Psychol Sci 2012 Sep 05;7(5):450-463. [doi: 10.1177/1745691612454134] [Medline: 26168503]
90. van der Linden D, Frese M, Meijman TF. Mental fatigue and the control of cognitive processes: effects on perseveration and planning. Acta Psychologica 2003 May;113(1):45-65. [doi: 10.1016/s0001-6918(02)00150-6]
91. Chajut E, Algom D. Selective attention improves under stress: implications for theories of social cognition. J Pers Soc Psychol 2003 Aug;85(2):231-248. [doi: 10.1037/0022-3514.85.2.231] [Medline: 12916567]
92. Zhang C, Smolders KC, Lakens D, IJsselsteijn WA. Two experience sampling studies examining the variation of self-control capacity and its relationship with core affect in daily life. J Res Pers 2018 Jun;74:102-113. [doi: 10.1016/j.jrp.2018.03.001]
93. Russo DJ, Van Roy B, Kazerouni A, Osband I, Wen Z. A tutorial on Thompson sampling. FNT in Machine Learning 2018;11(1):1-96. [doi: 10.1561/2200000070]

94. Gershman SJ, Daw ND. Reinforcement learning and episodic memory in humans and animals: an integrative framework. Annu Rev Psychol 2017 Jan 03;68(1):101-128 [FREE Full text] [doi: 10.1146/annurev-psych-122414-033625] [Medline: 27618944]
95. Klein MC, Mogles N, Treur J, van Wissen A. A computational model of habit learning to enable ambient support for lifestyle change. In: Modern Approaches in Applied Intelligence. Berlin: Springer; 2011:130-142.
96. Miller KJ, Shenhav A, Ludvig EA. Habits without values. Psychol Rev 2019 Mar;126(2):292-311 [FREE Full text] [doi: 10.1037/rev0000120] [Medline: 30676040]
97. Crutzen R, Peters GY. Evolutionary learning processes as the foundation for behaviour change. Health Psychol Rev 2018 Mar 18;12(1):43-57. [doi: 10.1080/17437199.2017.1362569] [Medline: 28764599]
98. Raynor HA, Epstein LH. Dietary variety, energy regulation, and obesity. Psychol Bull 2001 May;127(3):325-341. [doi: 10.1037/0033-2909.127.3.325] [Medline: 11393299]
99. Trope Y, Liberman N. Construal-level theory of psychological distance. Psychol Rev 2010 Apr;117(2):440-463 [FREE Full text] [doi: 10.1037/a0018963] [Medline: 20438233]
100. Locke EA, Latham GP. Building a practically useful theory of goal setting and task motivation: a 35-year odyssey. Am Psychol 2002 Sep;57(9):705-717. [doi: 10.1037/0003-066X.57.9.705]
101. Abraham C, Michie S. A taxonomy of behavior change techniques used in interventions. Health Psychol 2008 May;27(3):379-387. [doi: 10.1037/0278-6133.27.3.379] [Medline: 18624603]
102. Kok G, Gottlieb NH, Peters GJY, Mullen PD, Parcel GS, Ruiter RAC, et al. A taxonomy of behaviour change methods: an Intervention Mapping approach. Health Psychol Rev 2016 Sep;10(3):297-312 [FREE Full text] [doi: 10.1080/17437199.2015.1077155] [Medline: 26262912]
103. Michie S, Carey RN, Johnston M, Rothman AJ, de Bruin M, Kelly MP, et al. From theory-inspired to theory-based interventions: a protocol for developing and testing a methodology for linking behaviour change techniques to theoretical mechanisms of action. Ann Behav Med 2018 May 18;52(6):501-512 [FREE Full text] [doi: 10.1007/s12160-016-9816-6] [Medline: 27401001]
104. Hekler EB, Michie S, Pavel M, Rivera DE, Collins LM, Jimison HB, et al. Advancing models and theories for digital behavior change interventions. Am J Prev Med 2016 Nov;51(5):825-832 [FREE Full text] [doi: 10.1016/j.amepre.2016.06.013] [Medline: 27745682]
105. Intille SS, Kukla C, Farzanfar R, Bakr W. Just-in-time technology to encourage incremental, dietary behavior change. AMIA Annu Symp Proc 2003:874 [FREE Full text] [Medline: 14728379]
106. Jaimes LG, Calderon J, Lopez J, Raij A. Trends in mobile cyber-physical systems for health just-in time interventions. In: Proceedings of SoutheastCon 2015. 2015 Presented at: SoutheastCon 2015; April 9-12, 2015; Fort Lauderdale, FL, USA p. 1-6. [doi: 10.1109/SECON.2015.7132887]
107. Nahum-Shani I, Hekler EB, Spruijt-Metz D. Building health behavior models to guide the development of just-in-time adaptive interventions: a pragmatic framework. Health Psychol 2015 Dec;34S:1209-1219 [FREE Full text] [doi: 10.1037/hea0000306] [Medline: 26651462]
108. Riley WT, Serrano KJ, Nilsen W, Atienza AA. Mobile and wireless technologies in health behavior and the potential for intensively adaptive interventions. Curr Opin Psychol 2015 Oct 01;5:67-71 [FREE Full text] [doi: 10.1016/j.copsyc.2015.03.024] [Medline: 26086033]
109. Pina L, Ramirez E, Griswold WG. Fitbit+: A behavior-based intervention system to reduce sedentary behavior. In: Proceedings of the 6th International Conference on Pervasive Computing Technologies for Healthcare. 2012 Presented at: 6th International Conference on Pervasive Computing Technologies for Healthcare; May 21-24. 2012; San Diego, CA, USA p. 175-178. [doi: 10.4108/icst.pervasivehealth.2012.248761]
110. Thomas JG, Bond DS. Behavioral response to a just-in-time adaptive intervention (JITAI) to reduce sedentary behavior in obese adults: implications for JITAI optimization. Health Psychol 2015 Dec;34S(Suppl):1261-1267 [FREE Full text] [doi: 10.1037/hea0000304] [Medline: 26651467]
111. Guo L. Quantified-self 2.0: using context-aware services for promoting gradual behaviour change. arXiv:1610.00460. 2016. URL: https://arxiv.org/abs/1610.00460 [accessed 2021-03-17]
112. Adams AT, Costa J, Jung M, Choudhury T. Mindless computing: designing technologies to subtly influence behavior. In: Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing. 2015 Presented at: UbiComp '15: The 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing; September, 2015; Osaka Japan p. 719-730. [doi: 10.1145/2750858.2805843]
113. Schneider C, Weinmann M, vom Brocke J. Digital nudging: guiding online user choices through interface design. Commun ACM 2018 Jun 25;61(7):67-73 [FREE Full text] [doi: 10.1145/3213765]
114. Johnson EJ, Shu SB, Dellaert BGC, Fox C, Goldstein DG, Häubl G, et al. Beyond nudges: tools of a choice architecture. Mark Lett 2012 May 25;23(2):487-504. [doi: 10.1007/s11002-012-9186-1]
115. Thaler RH, Sunstein CR. Nudge: improving decisions about health, wealth, and happiness. New Haven, Connecticut, United States: Yale University Press; 2008:1-293.
116. Johnson EJ, Goldstein D. Medicine. Do defaults save lives? Science 2003 Nov 21;302(5649):1338-1339. [doi: 10.1126/science.1091721] [Medline: 14631022]

117. Lee MK, Kiesler S, Forlizzi J. Mining behavioral economics to design persuasive technology for healthy choices. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2011 Presented at: SIGCHI Conference on Human Factors in Computing Systems; May, 2011; Vancouver, BC, Canada p. 325-334. [doi: 10.1145/1978942.1978989]
118. Zhang C, Starczewski A, Lakens D, IJsselsteijn W. A Decision-Making Perspective on Coaching Behavior Change: A Field Experiment on Promoting Exercise at Work. In: Persuasive Technology. Switzerland: Springer; 2018:87-98.
119. Rahman T, Czerwinski M, Gilad-Bachrach R, Johns P. Predicting about-to-eat moments for just-in-time eating intervention. In: Proceedings of the 6th International Conference on Digital Health Conference. 2016 Presented at: 6th International Conference on Digital Health Conference; April 2016; Montréal Québec Canada p. 141-150. [doi: 10.1145/2896338.2896359]
120. Jaimes LG, Llofriu M, Raij A. A stress-free life: just-in-time interventions for stress via real-time forecasting and intervention adaptation. In: Proceedings of the 9th International Conference on Body Area Networks. 2014 Presented at: 9th International Conference on Body Area Networks; Sep 29th – Oct 1st, 2014; London, Great Britain p. 197-203. [doi: 10.4108/icst.bodynets.2014.258237]
121. Zuboff S. The age of surveillance capitalism: the fight for a human future at the new frontier of power. New York: Profile Books; 2019:-.
122. Kersten-van Dijk ET, Westerink JH, Beute F, IJsselsteijn WA. Personal informatics, self-insight, and behavior change: a critical review of current literature. Hum Comput Interact 2017 Feb 17;32(5-6):268-296. [doi: 10.1080/07370024.2016.1276456]
123. Choe EK, Lee NB, Lee B, Pratt W, Kientz JA. Understanding quantified-selfers' practices in collectingexploring personal data. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2014 Presented at: SIGCHI Conference on Human Factors in Computing Systems; April, 2014; Toronto Ontario Canada p. 1143-1152. [doi: 10.1145/2556288.2557372]
124. Li I, Dey AK, Forlizzi J. Understanding my data, myself: supporting self-reflection with ubicomp technologies. In: Proceedings of the 13th international conference on Ubiquitous computing. 2011 Presented at: UbiComp '11: Proceedings of the 13th international conference on Ubiquitous computing; September 2011; Beijing China p. 405-414. [doi: 10.1145/2030112.2030166]
125. Nacke LE, Deterding S. The maturing of gamification research. Comput Hum Behav 2017 Jun;71:450-454. [doi: 10.1016/j.chb.2016.11.062]
126. Cugelman B. Gamification: what it is and why it matters to digital health behavior change developers. JMIR Serious Games 2013 Dec 12;1(1):e3 [FREE Full text] [doi: 10.2196/games.3139] [Medline: 25658754]
127. Deterding S, Dixon D, Khaled R, Nacke L. From game design elements to gamefulness: defining "gamification". In: Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments. 2011 Presented at: MindTrek '11: 15th International Academic MindTrek Conference: Envisioning Future Media Environments; September, 2011; Tampere Finland p. 9-15. [doi: 10.1145/2181037.2181040]
128. Zuckerman O, Gal-Oz A. Deconstructing gamification: evaluating the effectiveness of continuous measurement, virtual rewards, and social comparison for promoting physical activity. Pers Ubiquit Comput 2014 Jul 5;18(7):1705-1719. [doi: 10.1007/s00779-014-0783-2]
129. Munson SA, Consolvo S. Exploring goal-setting, rewards, self-monitoring,sharing to motivate physical activity. In: Proceedings of the 6th International Conference on Pervasive Computing Technologies for Healthcare. 2012 Presented at: 6th International Conference on Pervasive Computing Technologies for Healthcare; May 21–24, 2012; San Diego, United States p. 25-32. [doi: 10.4108/icst.pervasivehealth.2012.248691]
130. DeSmet A, Shegog R, Van Ryckeghem D, Crombez G, De Bourdeaudhuij I. A systematic review and meta-analysis of interventions for sexual health promotion involving serious digital games. Games Health J 2015 Apr;4(2):78-90. [doi: 10.1089/g4h.2014.0110] [Medline: 26181801]
131. Nicholson S. Two paths to motivation through game design elements: reward-based gamification and meaningful gamification. Proceedings of iConference. 2013. URL: https://core.ac.uk/download/pdf/10202074.pdf [accessed 2021-03-17]
132. Berkovsky S, Coombe M, Freyne J, Bhandari D, Baghaei N. Physical activity motivating games: virtual rewards for real activity. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2010 Presented at: CHI '10: SIGCHI Conference on Human Factors in Computing Systems; April 2010; Atlanta Georgia USA p. 243-252. [doi: 10.1145/1753326.1753362]
133. Novak JI, Loy J. A pilot study for utilizing additive manufacturing and responsive rewards in physical activity gamification. Design for Health 2019 Jan 12;2(2):266-284. [doi: 10.1080/24735132.2018.1541400]
134. Meder M, Plumbaum T, Raczkowski A, Jain B, Albayrak S. Gamification in e-commerce: tangible vs. intangible rewards. In: Proceedings of the 22nd International Academic Mindtrek Conference. 2018 Presented at: Mindtrek '18: 22nd International Academic Mindtrek Conference; October 2018; Tampere Finland p. 11-19. [doi: 10.1145/3275116.3275126]
135. Shahrestani A, Van Gorp P, Le BP, Greidanus F, De GK, Leermakers J. Unified Health Gamification can significantly improve well-being in corporate environments. In: Proceedings of the 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). 2017 Presented at: 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); July 11-15, 2017; Jeju, Korea (South) p. 4507-4511. [doi: 10.1109/embc.2017.8037858]

136. Saksono H, Castaneda-Sceppa C, Hoffman J, Morris V, Seif EM, Parker A. Storywell: designing for family fitness app motivation by using social rewards and reflection. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. 2020 Presented at: 2020 CHI Conference on Human Factors in Computing Systems; April 2020; Honolulu HI USA p. 1-13. [doi: 10.1145/3313831.3376686]
137. Armstrong AW, Watson AJ, Makredes M, Frangos JE, Kimball AB, Kvedar JC. Text-message reminders to improve sunscreen use: a randomized, controlled trial using electronic monitoring. Arch Dermatol 2009;145(11):1230-1236. [doi: 10.1001/archdermatol.2009.269] [Medline: 19917951]
138. Thakkar J, Kurup R, Laba T, Santo K, Thiagalingam A, Rodgers A, et al. Mobile telephone text messaging for medication adherence in chronic disease: a meta-analysis. JAMA Intern Med 2016 Mar;176(3):340-349. [doi: 10.1001/jamainternmed.2015.7667] [Medline: 26831740]
139. Kolb DA, Boyatzis RE. Goal-setting and self-directed behavior change. Hum Relat 2016 Apr 22;23(5):439-457. [doi: 10.1177/001872677002300506]
140. Radha M, Willemsen MC, Boerhof M, IJsselsteijn WA. Lifestyle recommendations for hypertension through Rasch-based feasibility modeling. In: Proceedings of the 2016 Conference on User Modeling Adaptation and Personalization. 2016 Presented at: UMAP '16: Proceedings of the 2016 Conference on User Modeling Adaptation and Personalization; July 2016; Halifax Nova Scotia p. 239-247. [doi: 10.1145/2930238.2930251]
141. Rutjes H, Willemsen MC, IJsselsteijn WA. Beyond behavior: the coach's perspective on technology in health coaching. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. 2019 Presented at: CHI '19: CHI Conference on Human Factors in Computing Systems; May 2019; Glasgow, Scotland, UK p. 1-14. [doi: 10.1145/3290605.3300900]
142. Zhou M, Mintz Y, Fukuoka Y, Goldberg K, Flowers E, Kaminsky P, et al. Personalizing mobile fitness apps using reinforcement learning. CEUR Workshop Proc 2018 Mar 07;2068:- [FREE Full text] [Medline: 32405286]
143. Geurts E, Van Geel , F, Feys P, Coninx K. WalkWithMe: personalized goal setting and coaching for walking in people with multiple sclerosis. In: Proceedings of the 27th ACM Conference on User Modeling, Adaptation and Personalization. 2019 Presented at: 7th ACM Conference on User Modeling, Adaptation and Personalization; June 2019; Larnaca Cyprus p. 51-60. [doi: 10.1145/3320435.3320459]
144. Chokshi NP, Adusumalli S, Small DS, Morris A, Feingold J, Ha YP, et al. Loss‐framed financial incentives and personalized goal‐setting to increase physical activity among ischemic heart disease patients using wearable devices: The ACTIVE REWARD randomized trial. J Am Heart Assoc 2018 Jun 19;7(12):-. [doi: 10.1161/jaha.118.009173]
145. Adriaanse MA, Vinkers CD, De Ridder DT, Hox JJ, De Wit JB. Do implementation intentions help to eat a healthy diet? A systematic review and meta-analysis of the empirical evidence. Appetite 2011 Feb;56(1):183-193 [FREE Full text] [doi: 10.1016/j.appet.2010.10.012] [Medline: 21056605]
146. Stawarz KM. Towards better medication adherence apps: preventing forgetfulness by facilitating the formation of routine-based remembering strategies. University College London. 2017. URL: https://discovery.ucl.ac.uk/id/eprint/1546254/ [accessed 2021-03-17]
147. Luszczynska A, Sobczyk A, Abraham C. Planning to lose weight: randomized controlled trial of an implementation intention prompt to enhance weight reduction among overweight and obese women. Health Psychol 2007 Jul;26(4):507-512. [doi: 10.1037/0278-6133.26.4.507] [Medline: 17605571]
148. Pinder C, Vermeulen J, Wicaksono A, Beale R, Hendley RJ. If this, then habit exploring context-aware implementation intentions on smartphones. In: Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct. 2016 Presented at: MobileHCI '16: 18th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct; 2016; Florence Italy p. 690-697. [doi: 10.1145/2957265.2961837]
149. Dogangün A, Schwarz M, Kloppenborg K, Le R. An approach to improve physical activity by generating individual implementation intentions. In: Proceedings of the 25th Conference on User Modeling, Adaptation and Personalization. 2017 Presented at: UMAP '17: Adjunct Publication of the 25th Conference on User Modeling, Adaptation and Personalization; July, 2017; Bratislava Slovakia p. 370-375. [doi: 10.1145/3099023.3099101]
150. Lehto T, Oinas-Kukkonen H. Persuasive features in web-based alcohol and smoking interventions: a systematic review of the literature. J Med Internet Res 2011 Jul 22;13(3):e46 [FREE Full text] [doi: 10.2196/jmir.1559] [Medline: 21795238]
151. Mercer K, Li M, Giangregorio L, Burns C, Grindrod K. Behavior change techniques present in wearable activity trackers: a critical analysis. JMIR Mhealth Uhealth 2016 Apr 27;4(2):e40 [FREE Full text] [doi: 10.2196/mhealth.4461] [Medline: 27122452]
152. Payne HE, Lister C, West JH, Bernhardt JM. Behavioral functionality of mobile apps in health interventions: a systematic review of the literature. JMIR Mhealth Uhealth 2015 Feb 26;3(1):e20 [FREE Full text] [doi: 10.2196/mhealth.3335] [Medline: 25803705]
153. Zhao J, Freeman B, Li M. Can mobile phone apps influence people's health behavior change? An evidence review. J Med Internet Res 2016 Oct 31;18(11):e287 [FREE Full text] [doi: 10.2196/jmir.5692] [Medline: 27806926]
154. Michie S, Abraham C, Whittington C, McAteer J, Gupta S. Effective techniques in healthy eating and physical activity interventions: a meta-regression. Health Psychol 2009 Nov;28(6):690-701. [doi: 10.1037/a0016136] [Medline: 19916637]

155. Hermsen S, Frost J, Renes RJ, Kerkhof P. Using feedback through digital technology to disrupt and change habitual behavior: a critical review of current literature. Comput Hum Behav 2016 Apr;57:61-74. [doi: 10.1016/j.chb.2015.12.023]
156. van Dijk ET, Westerink JHDM, Beute F, IJsselsteijn WA. In sync: the effect of physiology feedback on the match between heart rate and self-reported stress. Biomed Res Int 2015;2015:- [FREE Full text] [doi: 10.1155/2015/134606] [Medline: 26146611]
157. Klein M, Mogles N, van Wissen A. Why won't you do what's good for you? Using intelligent support for behavior change. In: Human Behavior Understanding. Berlin, Heidelberg: Springer; 2011:104-115.
158. Klein M, Mogles N, van Wissen A. Intelligent mobile support for therapy adherence and behavior change. J Biomed Inform 2014 Oct;51:137-151 [FREE Full text] [doi: 10.1016/j.jbi.2014.05.005] [Medline: 24858491]
159. de Vries H. An integrated approach for understanding health behavior; the I-change model as an example. Psychol Behav Sci Inter J 2017 Mar 9;2(2):-. [doi: 10.19080/pbsij.2017.02.555585]
160. Riley WT, Martin CA, Rivera DE, Hekler EB, Adams MA, Buman MP, et al. Development of a dynamic computational model of social cognitive theory. Transl Behav Med 2016 Dec;6(4):483-495 [FREE Full text] [doi: 10.1007/s13142-015-0356-6] [Medline: 27848208]
161. Zhang C, Vanschoren J, van Wissen A, Lakens D, de RB, IJsselsteijn WA. Theory-based habit modeling for enhancing behavior prediction in behavior change support systems. Manuscript in preparation 2021:- Manuscript in preparation.
162. West R. Theory of addiction. Oxford, UK: Blackwell; 2006.
163. Hall PA, Fong GT. Temporal self-regulation theory: a model for individual health behavior. Health Psychol Rev 2007 Mar;1(1):6-52. [doi: 10.1080/17437190701492437]
164. Vlaev I, Dolan P. Action change theory: a reinforcement learning perspective on behavior change. Rev Gen Psychol 2015 Mar;19(1):69-95. [doi: 10.1037/gpr0000029]
165. Hastie R. Problems for judgment and decision making. Annu Rev Psychol 2001 Feb;52(1):653-683. [doi: 10.1146/annurev.psych.52.1.653] [Medline: 11148321]
166. Dolan RJ, Dayan P. Goals and habits in the brain. Neuron 2013 Oct 16;80(2):312-325 [FREE Full text] [doi: 10.1016/j.neuron.2013.09.007] [Medline: 24139036]
167. Heckhausen J. The motivation-volition divide and its resolution in action-phase models of developmental regulation. Res Hum Dev 2007 Nov 14;4(3-4):163-180. [doi: 10.1080/15427600701662983]
168. Heckhausen H, Gollwitzer PM. Thought contents and cognitive functioning in motivational versus volitional states of mind. Motiv Emot 1987 Jun;11(2):101-120. [doi: 10.1007/bf00992338]
169. Usher M, McClelland JL. The time course of perceptual choice: the leaky, competing accumulator model. Psychol Rev 2001;108(3):550-592. [doi: 10.1037//0033-295x.108.3.550]
170. Daw ND, Niv Y, Dayan P. Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. Nat Neurosci 2005 Dec 6;8(12):1704-1711. [doi: 10.1038/nn1560] [Medline: 16286932]
171. Keramati M, Dezfouli A, Piray P. Speed/accuracy trade-off between the habitual and the goal-directed processes. PLoS Comput Biol 2011 May;7(5):- [FREE Full text] [doi: 10.1371/journal.pcbi.1002055] [Medline: 21637741]
172. Nilsen WJ, Pavel M. Moving behavioral theories into the 21st century: technological advancements for improving quality of life. IEEE Pulse 2013 Sep;4(5):25-28. [doi: 10.1109/MPUL.2013.2271682] [Medline: 24056790]
173. Spruijt-Metz D, Hekler E, Saranummi N, Intille S, Korhonen I, Nilsen W, et al. Building new computational models to support health behavior change and maintenance: new opportunities in behavioral research. Transl Behav Med 2015 Sep;5(3):335-346 [FREE Full text] [doi: 10.1007/s13142-015-0324-1] [Medline: 26327939]
174. Zhang C, van Wissen A, Lakens D, Vanschoren J, De RB, IJsselsteijn W. Anticipating habit formation: a psychological computing approach to behavior change support. In: Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct. 2016 Presented at: UbiComp '16: Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct; September 201; Heidelberg Germany p. 1247-1254. [doi: 10.1145/2968219.2968439]
175. IJsselsteijn WA. Psychology 2.0 : towards a new science of mind and technology. Eindhoven, NL: Technische Universiteit Eindhoven; 2013.
176. Miller G. The smartphone psychology manifesto. Perspect Psychol Sci 2012 May 16;7(3):221-237. [doi: 10.1177/1745691612441215] [Medline: 26168460]

Abbreviations

BCT: behavior change technique
COMBI: computerized behavior intervention
JITAI: just-in-time adaptive intervention
SCT: Social Cognitive Theory
TPB: Theory of Planned Behavior
TTM: Transtheoretical Model

Edited by G Eysenbach; submitted 19.11.19; peer-reviewed by R Crutzen, S Sittig; comments to author 18.05.20; revised version received 26.09.20; accepted 17.02.21; published 09.04.21.

Please cite as:
Zhang C, Lakens D, IJsselsteijn WA
Theory Integration for Lifestyle Behavior Change in the Digital Age: An Adaptive Decision-Making Framework
J Med Internet Res 2021;23(4):e17127
URL: https://www.jmir.org/2021/4/e17127
doi:10.2196/17127
PMID:33835036

Review

Radiomic and Genomic Machine Learning Method Performance for Prostate Cancer Diagnosis: Systematic Literature Review

Rossana Castaldo[1], PhD; Carlo Cavaliere[1], PhD; Andrea Soricelli[1], PhD; Marco Salvatore[1], PhD; Leandro Pecchia[2*], PhD; Monica Franzese[1*], PhD

[1]IRCCS SDN, Naples, Italy

[2]University of Warwick, Coventry, United Kingdom

[*]these authors contributed equally

Corresponding Author:
Carlo Cavaliere, PhD
IRCCS SDN
113 Via E Gianturco
Naples, 80143
Italy
Phone: 39 3470563424
Email: carlo.cavaliere@synlab.it

Abstract

Background: Machine learning algorithms have been drawing attention at the joining of pathology and radiology in prostate cancer research. However, due to their algorithmic learning complexity and the variability of their architecture, there is an ongoing need to analyze their performance.

Objective: This study assesses the source of heterogeneity and the performance of machine learning applied to radiomic, genomic, and clinical biomarkers for the diagnosis of prostate cancer. One research focus of this study was on clearly identifying problems and issues related to the implementation of machine learning in clinical studies.

Methods: Following the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) protocol, 816 titles were identified from the PubMed, Scopus, and OvidSP databases. Studies that used machine learning to detect prostate cancer and provided performance measures were included in our analysis. The quality of the eligible studies was assessed using the QUADAS-2 (quality assessment of diagnostic accuracy studies–version 2) tool. The hierarchical multivariate model was applied to the pooled data in a meta-analysis. To investigate the heterogeneity among studies, I^2 statistics were performed along with visual evaluation of coupled forest plots. Due to the internal heterogeneity among machine learning algorithms, subgroup analysis was carried out to investigate the diagnostic capability of machine learning systems in clinical practice.

Results: In the final analysis, 37 studies were included, of which 29 entered the meta-analysis pooling. The analysis of machine learning methods to detect prostate cancer reveals the limited usage of the methods and the lack of standards that hinder the implementation of machine learning in clinical applications.

Conclusions: The performance of machine learning for diagnosis of prostate cancer was considered satisfactory for several studies investigating the multiparametric magnetic resonance imaging and urine biomarkers; however, given the limitations indicated in our study, further studies are warranted to extend the potential use of machine learning to clinical settings. Recommendations on the use of machine learning techniques were also provided to help researchers to design robust studies to facilitate evidence generation from the use of radiomic and genomic biomarkers.

(J Med Internet Res 2021;23(4):e22394) doi:10.2196/22394

KEYWORDS

prostate cancer; machine learning; systematic review; meta-analysis; diagnosis; imaging; radiomics; genomics; clinical; biomarkers

Introduction

Prostate cancer (PCa) is the second most diagnosed cancer worldwide in men [1,2]. To guarantee cancer-specific survival, early detection of PCa is essential at a treatable stage. The most common method to diagnose PCa is via transrectal ultrasonography (TRUS) [3]. The rapid development of medical imaging techniques and modalities has demonstrated great value

in the screening, diagnosis, treatment response measurement, and prognosis evaluation of PCa. In particular, radiomic investigation, defined as computationally extracting quantitative image features for the characterization of disease patterns [4], has been intensively applied to tumor detection, localization, staging, aggressiveness assessment, treatment decision-making assistance, and patient follow-up in PCa [5] .

More recently, multiparametric magnetic resonance imaging (mpMRI) has been demonstrated to be a better radiomic biomarker than systematic TRUS biopsy, achieving high diagnostic accuracy and becoming a clinical routine investigation for suspected PCa patients [6,7]. The second version of the Prostate Imaging Reporting and Data System (PI-RADS-V2) was updated in regard to minimum technical acquisition parameters and image interpretation [8]. It describes a standard prostate mpMRI protocol that combines anatomical T2-weighted images with functional sequences, that is, diffusion-weighted imaging (DWI) or dynamic contrast-enhanced (DCE) sequences.

Alongside radiomic investigation, there are numerous Food and Drug Administration–approved genomic biomarkers underlying the biomolecular functions most strongly associated with clinical outcomes. In fact, a major focus of personalized medicine has been the biomolecular characterization of tumors by integrating genomics into clinical oncology to identify unique druggable targets and generate higher-order tumor classification methods that can support clinical treatment decisions [9]. They are mainly used to decide whether biopsy screening is necessary and whether patients require primary treatment (such as radical prostatectomy or radiation therapy) [10]. The combination of biopsy screening and evaluation of the Gleason score still remains the most widely accepted grading system in the evaluation of prostatic adenocarcinoma [11]. The Gleason grading system is based on a morphologic continuum of architectural dedifferentiation and is directly correlated with response to therapy and mortality rate. However, novel biomarker tests that can potentially detect PCa from blood, urine, tissue, and semen samples continue to be investigated. Prostate-specific antigen (PSA) is the most commonly used biomarker for the management of PCa [12]. Increased PSA density has been shown to be associated with increased risk of PCa compared to healthy or benign prostatic hyperplasia patients [13]. The Prostate Health Index and 4Kscore utilize isoforms of PSA and its precursors to help risk-stratify patients with an abnormal PSA level. In addition, microRNAs have an important role during tumor progression, and their combination with PSA serum can improve prediction of PCa status [14-16]. Other proposed biomarkers that belong to various classes of biological compounds, including proteins and metabolites, have shown to be noninvasive methods with high diagnostic potential [17].

Over the last decade, the landscape for PCa detection tools has expanded to include novel biomarkers, clinical information, genomic assays, and noninvasive imaging tests. The prospect of detecting PCa using readily available clinical and demographic health information is a potentially innovative part of improving screening practices [18].

In this scenario, machine learning (ML) is helping researchers in identifying and discovering new biomarkers to detect PCa. ML is a branch of artificial intelligence (AI) techniques based on the development and training of algorithms by learning from data and the performance of predictions. ML methods are able to improve and learn over time in a more efficient way than classical statistical approaches [19]. Therefore, ML has been widely used in radiology and recently in the field of bioinformatics [6,20]. A recent field of ML, deep learning (DL), is based on artificial neural networks, which offer superior problem-solving capabilities applied to large heterogenous data sets [20,21]. Specifically, ML allows the integration or combination of different layers of data, such as those from medical images, laboratory results, clinical outcomes, biomarkers, and other biological features, for better prognostication and stratification of patients toward personalized medicine [22,23]. However, the accuracy of such algorithms can be highly impacted by the complex workflows adopted to develop and generalize such ML algorithms [24,25]. High heterogeneity is expected, as ML problems are usually regarded as black boxes, and the consideration of all possible risk factors and transformation is tremendously difficult [26,27]. Moreover, there are no clear guidelines on how to develop ML approaches for medical studies.

Therefore, this study aimed to suggest an integrated estimate of the accuracy for use of ML algorithms in detecting PCa through a systematic review and meta-analysis of the available studies. Due to the internal heterogeneity of ML algorithms, subgroup analyses helped in investigating the diagnostic capability of ML systems and highlighting the sources of bias and common pitfalls to avoid in order to assure reproducibility among studies. Subgroup analyses were mainly based on the model choice, model development, and validation methods to identify potential covariates that could influence the diagnostic performance of ML.

This review helps to support ML studies in rising up the pyramid of evidence. In fact, we identify and discuss recurrent factors that hinder the uptake of these studies in clinical settings.

To the best of the authors' knowledge, there are no systematic review and meta-analysis studies evaluating the performance and estimating the current status of existing approaches on PCa detection. Therefore, this study aims to fill the gap in the existing literature and gather recommendations on ML model development to achieve robust results to automatically detect PCa.

Methods

We conducted and reported this meta-analysis in accordance with PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [28]. Two researchers (RC and MF), who were blinded to the articles' author information, conducted the study inclusion, data extraction, and assessment of the risk of bias independently. A third author (CC) was consulted in case of disagreements.

Search Strategy

The PubMed, Scopus, and OvidSP (ie, Embase) databases were searched to identify studies evaluating the accuracy of radiomic, clinical, and genomic biomarkers in the diagnosis of PCa. The following criteria were used to limit the research: papers published in the last 5 years (from 2015 to 2020) to guarantee homogeneity among radiomic studies, as the new protocol (PI-RADS) for mpMRI was updated in 2015 [8]; study on adult humans (ie, not animals); language (English); and full-text publications. The search took place on February 24, 2020. The reference lists of the included studies were checked, and the authors were contacted if required. The search strategy and queries for each search database are presented in Table S1 in Multimedia Appendix 1.

An author (RC) retrieved the initial search results and removed duplicates via Excel (Microsoft). Subsequently, another author (MF) manually searched for and removed any remaining duplicates. Finally, RC and MF independently screened the studies by title, abstract, and keywords, after which the full texts of the selected studies were assessed by inclusion and exclusion criteria. The main considerations for study inclusion were if machine learning was fully applied in distinguishing individuals or lesions with clinically diagnosed PCa from controls and if the study assessed the accuracy of such applications. Detailed inclusion and exclusion criteria are reported in Table S2 in Multimedia Appendix 1.

Data Extraction and Outcomes of Interest

After the evaluation was completed, two authors extracted the following information from the selected literature: literature data—the first author, publication date, study population, number of patients, study design, and data collection; basic research information—age, Gleason score, and PSA level, where possible; information regarding the reference standard used in individual studies; definitions of positive and negative PCa (PCa positive and control) and methodologies to distinguish individuals or lesions with PCa from the control group; specific methodologies to process and classify data for use in machine learning algorithms; and the sensitivity, specificity, and, if available, true-positive (TP), true-negative (TN), false-positive (FP), and false-negative (FN) rates.

The authors independently graded the quality of the eligible studies using the quality assessment of diagnostic accuracy studies–version 2 (QUADAS-2) tool [29]. The full process is provided in the supplementary materials in Multimedia Appendix 1.

Meta-analysis Paper Inclusion Criteria and Subgroup Analysis

For radiomic analysis, due to the very low number of included studies investigating central gland and transition zone (TZ) prostate tumors, only studies investigating the peripheral zone were included in the meta-analysis. This was also due to the fact that central gland and TZ prostate tumors have significantly different quantitative imaging signatures [30], and they could have highly biased the final results.

Due to the low number of studies employing 3D volumes of interest (VOIs) to extract quantitative features, only studies delineating 2D regions of interest (ROIs) were included in the meta-analysis to reduce the risk of bias. This was mainly due to the fact that significant differences were found between prediction performance when using 3D VOIs and that when using 2D ROIs [31]. If studies investigated several diagnostic imaging techniques via ML, only classification models using mpMRI sequences were included in the meta-analysis.

To reduce heterogeneity among the selected studies, subgroup analyses were carried out for radiomic and genomic studies due to their intrinsic differences in data acquisition, analysis, and feature extraction. Radiomic subgroup analyses helped to investigate the role of the mpMRI biomarker in detecting PCa via ML, whereas genomic subgroup analyses were carried out to understand the role of genomic biomarkers in detecting PCa via ML.

Several covariates suitable for subgroup analysis were identified during the review process where the individual peculiarities of the studies, which may affect the outcome, were investigated.

The included studies were investigated if they explored a patient- or lesion-based model, validation approaches (cross-validation, hold-out approach or external validation, or no validation), ML algorithms (regression-based model, tree-based model, or deep learning algorithms), whether the studies used a DL or ML approach, or whether the employed data set was balanced or unbalanced. For genomic studies, the use of different specimens (ie, urine, serum, semen, and tissue) was also investigated in a subgroup analysis. One study [17] investigated both urine and serum specimens separately; therefore, ML performance was included for both predictors in the meta-analysis.

In case a study investigated multiple ML algorithms, only the method achieving the highest area under the curve (AUC) was included in the meta-analysis, as AUC is a good estimator of ML performance.

Statistical Analysis and Software Tools

This meta-analysis was conducted via the Open Meta-Analyst Software tool, and statistical significance was expressed with 95% CIs. Pooled estimates for sensitivity and specificity with the corresponding 95% CIs were used to determine the accuracy of machine learning for detecting PCa in radiomic and genomic studies. From these data, we generated a hierarchical summary receiver operating characteristic curve (HSROC) and coupled forest plots by random-effects model. Heterogeneity among studies was assessed by calculation of the inconsistency index (I^2) and evaluation of the Cochran χ^2 test (Q test). An I^2 of ≥50% and P<.001 indicated substantial between-study heterogeneity. The TP/FP/TN/FN values were extracted or calculated from each independent study. A correction factor of 0.5 was added if any of the TP/FP/TN/FN rates reported a value of 0, in order to prevent zero cell count problem [32].

In our meta-analysis, a multivariate random-effects model was used to consider both within- and between-subject variability and threshold effects [33]. The HSROC curve was specified by pooled sensitivity and specificity point. Attempts were made

to resolve the heterogeneity by performing a subgroup analysis [34].

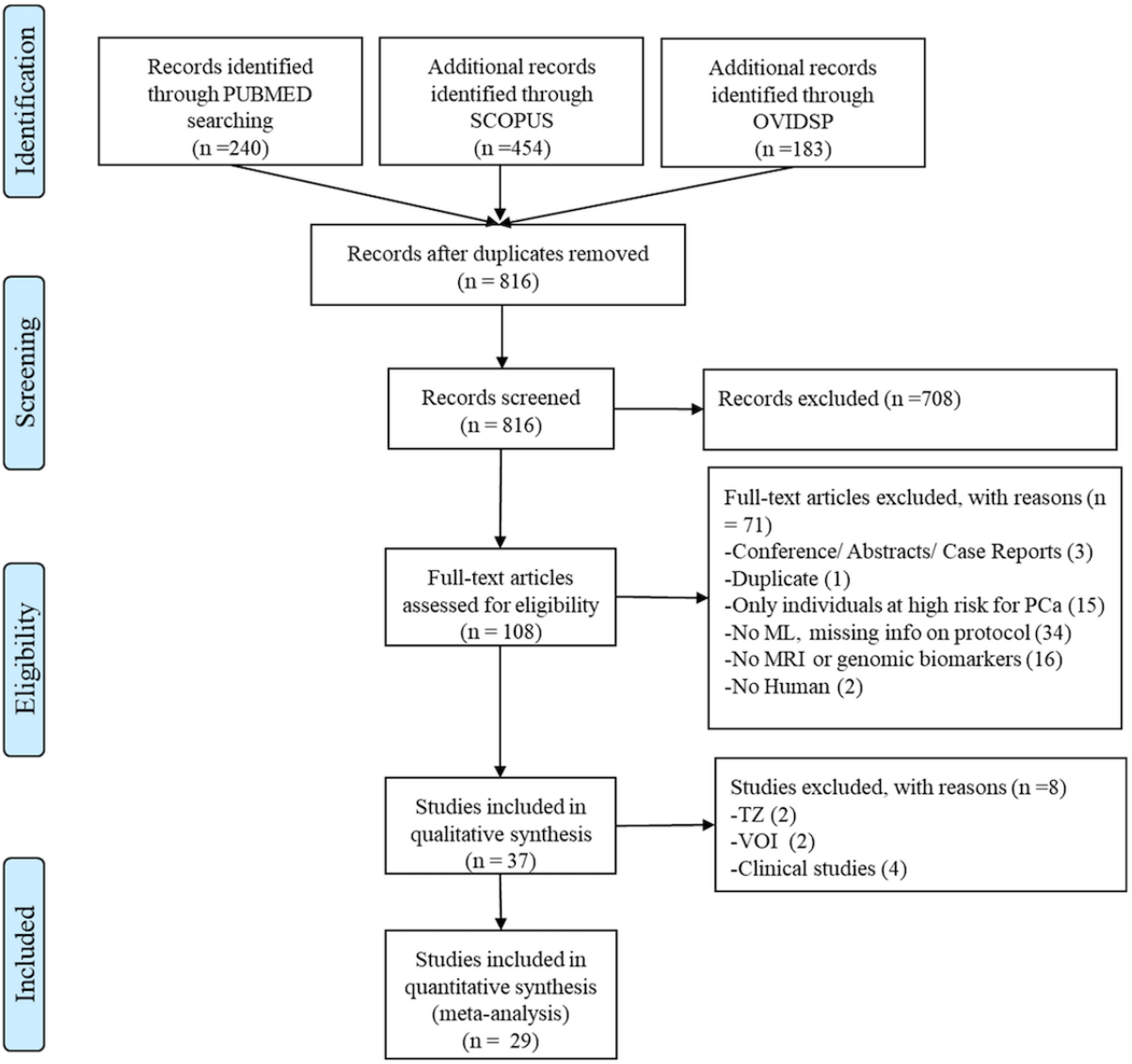
Results

Literature Search

According to the search strategy described above, 877 titles were identified in PubMed, Scopus, and OvidSP. After removing duplicates, 816 titles were considered. Of these, 708 were excluded after reading of the abstracts because they did not meet the inclusion criteria. From the remaining 108 full-text articles, 71 were removed due to the exclusion criteria. Finally, 37 full texts were included in the qualitative analysis, and 29 studies were considered appropriate for inclusion in the meta-analysis. A flowchart of the literature search is shown in Figure 1.

The distribution of the risk of bias evaluated via the QUADAS-2 tool for the included studies is presented in the supplementary materials (Figure S1 in Multimedia Appendix 1).

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart of literature search: included/excluded titles, abstracts, and full papers. ML: machine learning; MRI: magnetic resonance imaging; PCa: prostate cancer; TZ: transition zone; VOI: volume of interest.



Characteristics of the Included Studies

The publication years ranged from 2015 to 2020 to guarantee homogeneity among radiomic studies, as the new PI-RADS was updated in 2015 [8]. All patients were diagnosed with PCa by biopsy. The main characteristics of the studies are reported in Table 1. The extracted raw data are presented in Tables S3 and S4 in Multimedia Appendix 1.

Table 1. Characteristics of 37 studies included in the systematic review.

Characteristics	Studies, n	Patients (average over the number of studies), n
Study type		
Prospective	8	2210 (276.25)
Retrospective	29	6414 (221.17)
Data set type		
Private data set	33	7760 (235.15)
Public database (SPIE-AAPM-NCI[a] PROSTATEx challenge)	2	399 (199.5)
Mixed (private and public) data set	2	465 (232.5)
Classification algorithms		
Random forest	4	1621(405.25)
Regression-based models	20	4678 (233.9)
Partial least squares discriminant analysis (PLS-DA)	2	180 (90)
Linear discriminant analysis (LDA)	1	53
Support vector machine (SVM)	2	65 (32.5)
Classification and regression tree (CART)	1	67
Artificial neural networks (ANNs)	2	1012 (506)
Deep neural networks (DNNs)	1	195
Convolutional neural networks (CNNs)	3	696 (232)
Deep learning: SNCSAE[b]	1	57
Predictor type		
Multiparametric MRI[c]	20	5058 (252.9)
Genetic or molecular biomarker	13	3132 (240.92)
Urine	6	930 (155)
Serum	3	901 (300.3)
Semen	2	108 (54)
Tissue	2	800 (400)
Clinical data	4	2812 (703)
Validation method		
Internal validation	29	6540 (225.52)
External validation	3	1380 (460)
Internal and external validation	1	364
Unknown	5	704 (140.8)

[a]SPIE-AAPM-NCI: International Society for Optics and Photonics–American Association of Physicists in Medicine–National Cancer Institute.

[b]SNCSAE: stacked nonnegativity constraint sparse autoencoders.

[c]MRI: magnetic resonance imaging.

Quantitative Analysis (Meta-analysis)

Of the final 37 papers, 29 were considered for the meta-analysis. Eight studies were excluded to reduce heterogeneity among the studies. Of those, 2 studies were excluded because they extracted radiomic features from VOIs [35,36], and 2 studies [37,38] were excluded because they only focused on detecting TZ tumors. Due to the low number of studies investigating TZ tumors, a comparative assessment of the results for the peripheral zone, central gland, and TZ was not possible.

Studies [18,39-41] employing only clinical information were excluded because a minimum sample of 5 studies is recommended for a meta-analysis [34,42]. In fact, 5 or more studies are needed to reasonably achieve power from random-effects meta-analyses [43].

Radiomic

All the included studies for the radiomic analysis are reported in Table 2. A total of 4438 independent samples were inspected

from 16 studies with sensitivity and specificity ranging from 0.62 to 0.99 and 0.51 to 0.98, respectively.

Multivariate meta-analysis via the HSROC model was assessed for all the studies (Figure S2 in Multimedia Appendix 1). The pooled sensitivity and specificity were 0.815 (95% CI 0.410-0.999) and 0.828 (95% CI 0.424-0.999), respectively.

The calculated heterogeneity values for pooled sensitivity and specificity were 84% and 79% (P<.001), respectively; therefore, a random-effects model was adopted to generate coupled forest plots (Figure S3 in Multimedia Appendix 1).

Subgroup Analysis

To resolve the heterogeneity, subgroup analysis was conducted for different covariates. The subgroup analysis per model-based covariate is shown in Figure 2. Subgroup 1 included the studies that employed a lesion-based ML approach. Those studies [44-54] employed multiple lesions for each patient enrolled in the study. Subgroup 2 gathered those studies [55-59] that enrolled two distinct groups (PCa and controls) and employed a patient-based ML approach. The heterogeneity in the subgroups was greater than 70% (subgroup 1: P<.001, subgroup 2: P=.002).

Table 2. Accuracy measures of radiomic studies for the systematic review.

Study, year	Model basis[a]	Patients, n	Total sample (PCa+, PCa-)[b]	Crossval[c]/ split/none	ML[d] methods[e]	TP,[f] n	FN,[g] n	FP,[h] n	TN,[i] n	Sen[j] (lower-upper)	Spe[k] (lower-upper)
Zhao, 2015 [44]	LB	71	238 (92, 146)	120 (60, 60)	ANN	57	35	16	130	0.620 (0.517-0.713)	0.890 (0.829-0.932)
Valerio, 2016 [45]	LB	53	106 (53, 53)	None	LDA	51	2	1	53	0.962 (0.861-0.991)	0.981 (0.880-0.997)
Lay, 2017 [46]	LB	224	410 (123, 287)	Crossval	RF	109	14	57	230	0.886 (0.817-0.931)	0.801 (0.751-0.844)
Reda, 2017 [47]	LB	18	53 (26, 27)	Crossval	SNCSAE	26	1	1	27	0.963 (0.779-0.995)	0.964 (0.786-0.995)
Starobinets, 2017 [48]	LB	169	509 (291, 218)	Crossval	LR	264	27	24	194	0.907 (0.868-0.936)	0.890 (0.841-0.925)
Wang, 2017 [36]	PB	172	172 (79, 93)	Crossval	DCNN	55	24	15	78	0.696 (0.587-0.787)	0.839 (0.750-0.900)
Le, 2017 [52]	LB	364	913 (463, 450)	275 (139, 135)	multi-modal CNN	125	14	6	129	0.899 (0.837-0.939)	0.956 (0.905-0.980)
Kwon, 2018 [49]	LB	204	191 (36, 155)	Crossval	LASSO LR	35	5	9	90	0.875 (0.733-0.947)	0.909 (0.834-0.952)
Song, 2018 [50]	LB	195	547 (261, 286)	55 (23, 32)	DNN	20	3	3	29	0.870 (0.665-0.957)	0.906 (0.746-0.969)
Chen, 2019 [56]	PB	381	381 (182, 199)	155 (55, 60)	LR	55	1	1	59	0.982 (0.884-0.997)	0.983 (0.891-0.998)
Devine, 2019 [51]	LB	65	97 (81, 16)	Crossval	LR	61	20	2	14	0.753 (0.648-0.835)	0.875 (0.614-0.969)
Gholizadeh, 2019 [54]	LB	11	297 (161, 136)	Crossval	SVM	161	1	9	127	0.994 (0.958-0.999)	0.934 (0.878-0.965)
Ma, 2019 [58]	PB	81	81 (44, 37)	None	LR	42	2	5	32	0.955 (0.836-0.989)	0.865 (0.714-0.943)
Mazaheri, 2019 [53]	LB	67	170 (102, 68)	91 (52, 39)	CART	51	1	19	20	0.981 (0.876-0.997)	0.513 (0.360-0.664)
Qi, 2019 [57]	PB	199	199 (85, 114)	66 (28, 38)	LR	23	5	3	35	0.821 (0.636-0.924)	0.921 (0.782-0.974)
Zhang, 2019 [59]	PB	140	140 (60, 80)	Crossval	RF	14	6	5	22	0.700 (0.473-0.859)	0.815 (0.625-0.921)

[a]LB: lesion-based model; PB: patient-based model.

[b]PCa: prostate cancer.

[c]Crossval: cross-validation techniques.

[d]ML: machine learning.

[e]ANN: artificial neural networks; LDA: linear discriminant analysis; RF: random forest; SNCSAE: stacked nonnegativity constraint sparse autoencoders; LR: logistic regression; DCNN: deep convolutional neural networks; LASSO: least absolute shrinkage and selection operator; DNN: deep neural networks; SVM: support vector machine; CART: classification and regression tree.

[f]TP: true-positive.

[g]FN: false-negative.

[h]FP: false-positive.

[i]TN: true-negative.

[j]Sen: sensitivity.

[k]Spe: specificity.

Figure 3 shows the subgroup analysis among studies that employed internal cross-validation techniques (subgroup 1) [46-49,51,54,55,59], split validation approaches (subgroup 2) [44,50,52,53,56,57], and no validation (subgroup 3) [45,58]. The heterogeneity for subgroups 1 and 2 was around 80% (P<.001).

Figure 2. Subgroup analysis for the model-based covariate in radiomic studies. Subgroup 1: lesion-based models; subgroup 2: patient-based models. FN: false-negative; FP: false-positive; TN: true-negative; TP: true-positive.

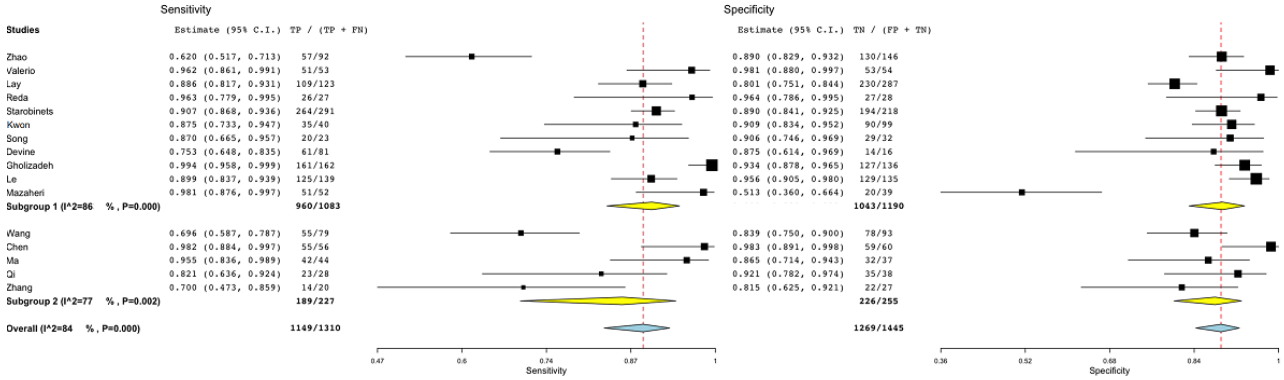


Figure 3. Subgroup analysis for the validation covariate in radiomic studies. Subgroup 1: internal cross-validation; subgroup 2: hold-out approach or external validation; subgroup 3: no validation. FN: false-negative; FP: false-positive; TN: true-negative; TP: true-positive.

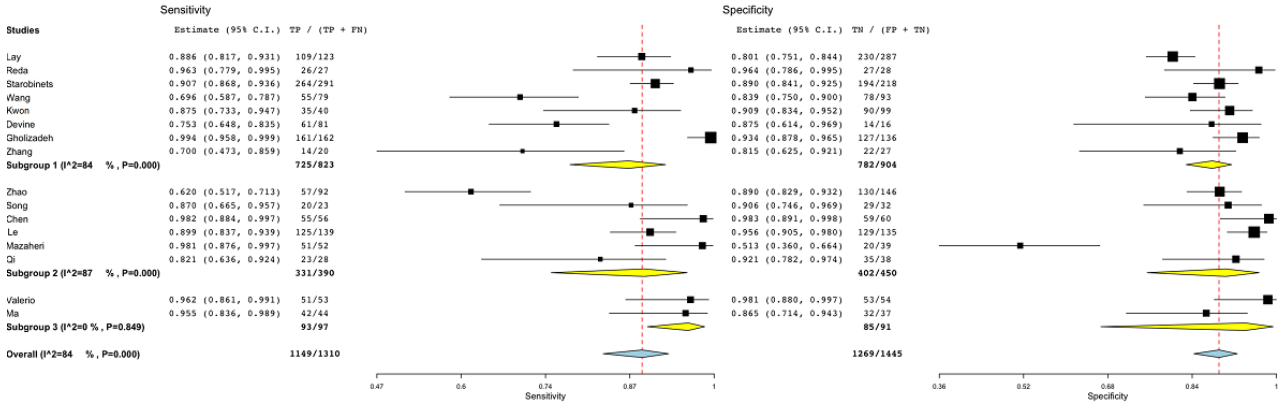


Figure 4 shows the subgroup analysis for regression-based models (subgroup 1) [45,48,49,51,56-58], tree-based models (subgroup 2) [46,53,59], and DL methods (subgroup 3) [44,47,50,52,55]. One study was not included [54], as it was the only study employing a support vector machine model. The heterogeneity among groups oscillated between 74% and 86% (subgroup 1: P=.001, subgroup 2: P=.01, subgroup 3: P<.001).

The results of the subgroup analysis to discriminate among machine and deep learning methods are reported in Figure 5. Subgroup 1 included the studies [45,46,48,49,51,53,56-59] employing ML methods, whereas subgroup 2 comprised the studies [44,47,50,52,55] employing DL methods (based on artificial neural networks) such as convolutional neural networks and deep neural networks. The I^2 statistics for subgroups 1 and 2 were 76% and 86% (P<.001), respectively.

Figure 6 shows the subgroup analysis based on whether the studies employed a balanced or unbalanced data set. A data set was defined as unbalanced if it had more than 30% of the total observations in one specific class rather than the other (PCa and controls) and did not apply any correction on performance (eg, synthetic minority oversampling technique [SMOTE] or voting techniques). The heterogeneity of subgroup 1 [36,44,51,53] was around 58% (P=.005). As a result, among the several covariates, the imbalance covariate was the only one by which the heterogeneity could be partially resolved.

Therefore, Devine et al [51], Wang et al [36], Mazaheri et al [53], and Zhao et al [44] were excluded from the coupled forest plot (Figure 7).

Figure 8 shows the HSROC curve for the studies employing balanced data sets to automatically detect PCa. The pooled sensitivity and specificity were 0.808 (95% CI 0.38-0.999) and 0.831 (95% CI 0.41-0.999), respectively.

Figure 4. Subgroup analysis for the machine learning algorithm covariate in radiomic studies. Subgroup 1: regression-based models; subgroup 2: tree-based models; subgroup 3: deep learning methods. FN: false-negative; FP: false-positive; TN: true-negative; TP: true-positive.

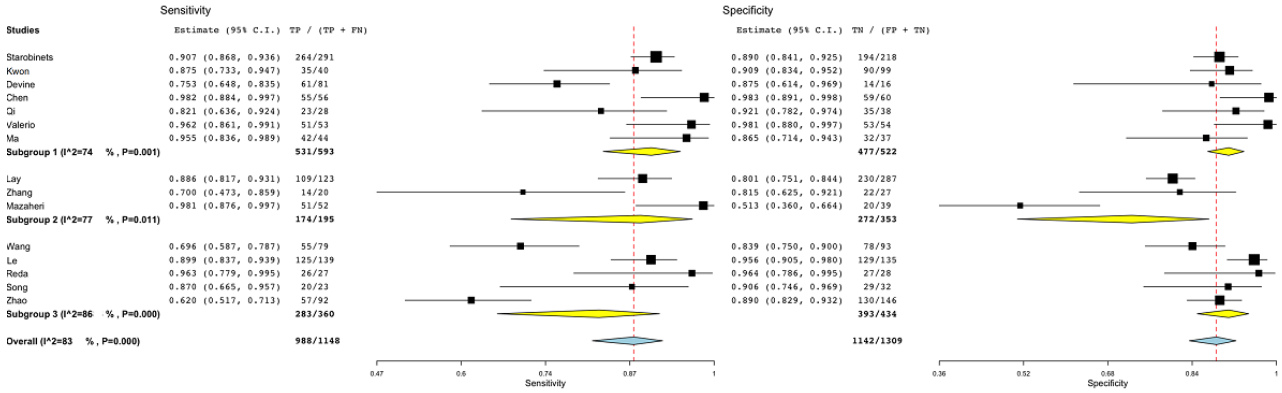


Figure 5. Subgroup analysis for the machine learning or deep learning covariate in radiomic studies. Subgroup 1: machine learning–based models; subgroup 2: deep learning methods. FN: false-negative; FP: false-positive; TN: true-negative; TP: true-positive.

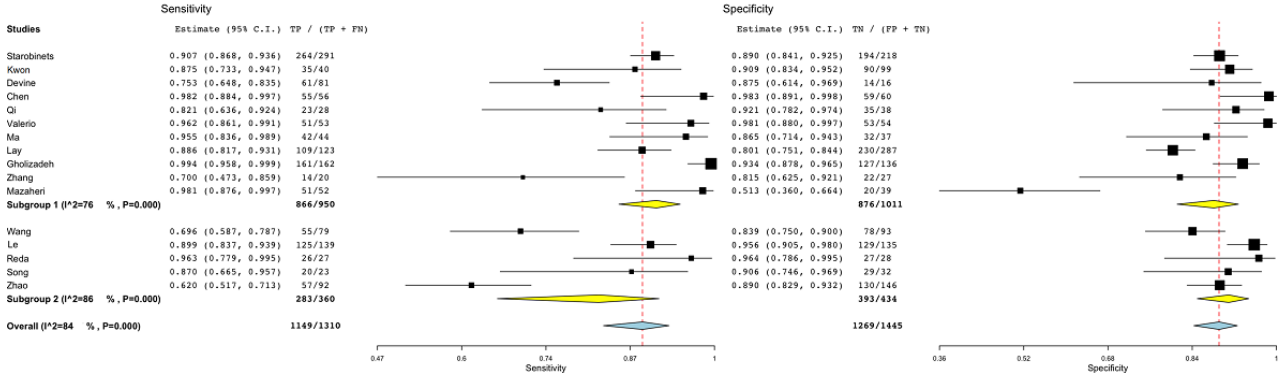


Figure 6. Subgroup analysis for the imbalance covariate in radiomic studies. Subgroup 1: balanced data sets; subgroup 2: unbalanced data sets. FN: false-negative; FP: false-positive; TN: true-negative; TP: true-positive.

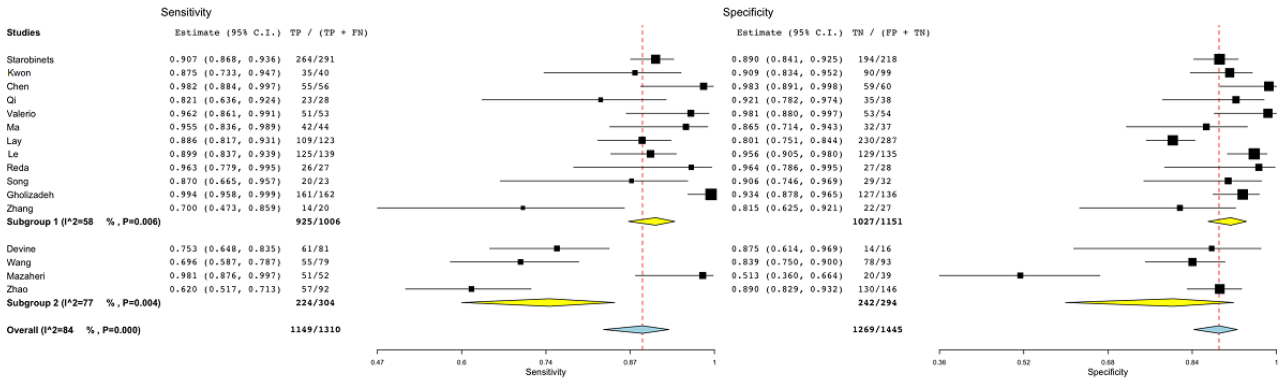


Figure 7. Subgroup analysis for the model-based covariate in a subset of radiomic studies. Subgroup 1: lesion-based models; subgroup 2: patient-based models. FN: false-negative; FP: false-positive; TN: true-negative; TP: true-positive.

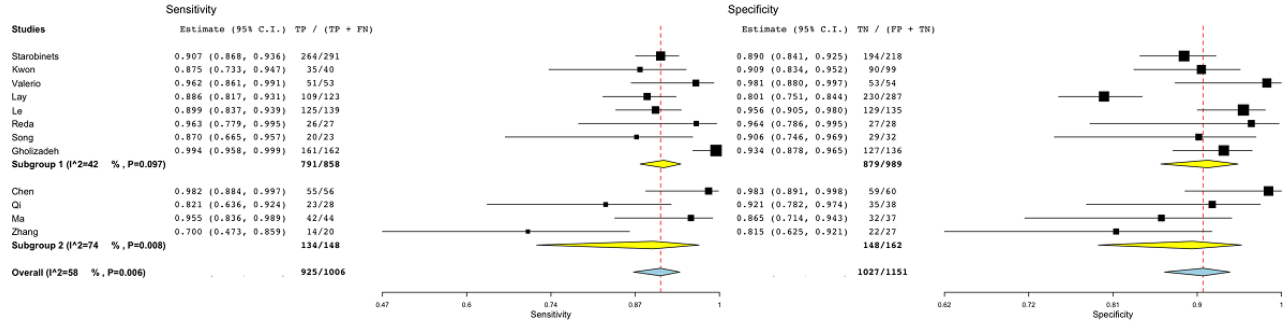
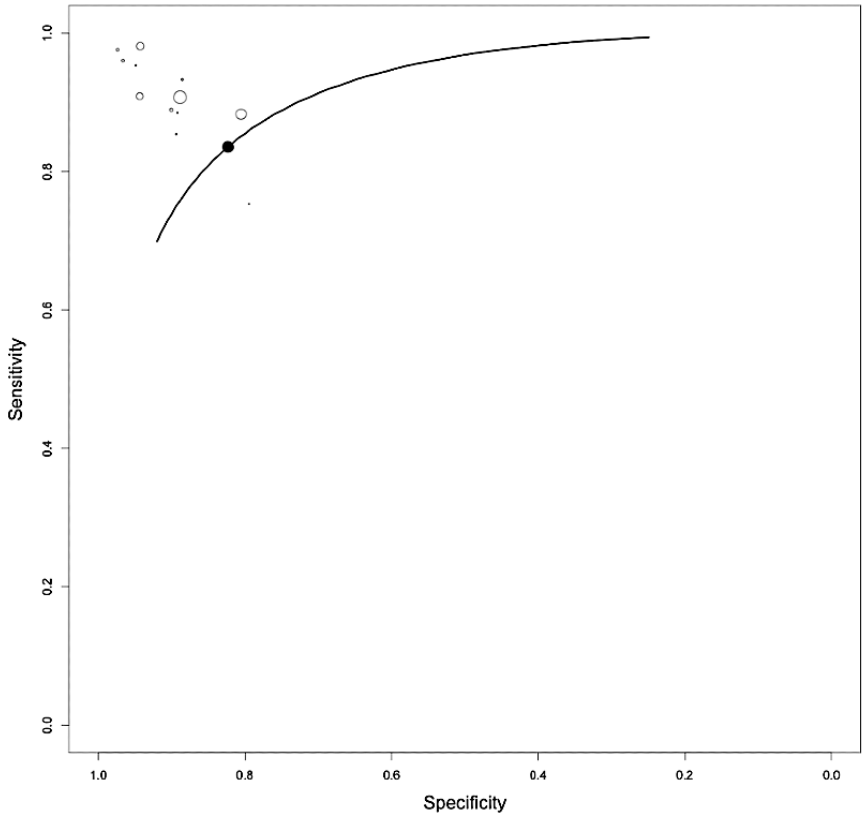


Figure 8. Overall hierarchical summary receiver operating characteristic curve (HSROC) for a subset of radiomic studies. HSROC was calculated for radiomic studies with low heterogeneity, excluding 4 studies [36,44,51,53].



Genomic

All the included studies for the genomic analysis are reported in Table 3. A total of 3221 independent samples were inspected from 14 studies and included in the meta-analysis, with sensitivity and specificity ranging from 0.67 to 0.95 and 0.15 to 0.97, respectively.

An HSROC model was assessed for all genomic studies (Figure S4 in Multimedia Appendix 1). The pooled sensitivity and specificity were 0.883 (95% CI 0.541-0.999) and 0.734 (95% CI 0.330-0.999), respectively.

The calculated heterogeneity values for the pooled sensitivity and specificity were 73% and 92% (P<.001), respectively; therefore, a random-effects model was adopted to generate the coupled forest plots (Figure S5 in Multimedia Appendix 1).

Subgroup Analysis

To resolve this heterogeneity, subgroup analyses were conducted for several covariates. The subgroup analysis for model-based covariates is shown in Figure 9. Subgroup 1 included the studies [60,61] that used malignant lesions and benign-adjacent tissue from PCa patients. Subgroup 2 gathered those studies [15-17,62-69] that enrolled two distinct groups (PCa and controls) and employed a patient-based ML approach. The heterogeneity for subgroup 1 was greater than 80%, whereas for subgroup 2 it was around 60%. However, subgroup 2 only included 2 studies.

Table 3. Accuracy measures of genomic studies for the systematic review.[a]

Study, year	Model basis[b]	Predictor	Patients, n	Total sample (PCa+, PCa-)[c]	Crossval[d]/ split/none	TP,[e] n	FN,[f] n	FP,[g] n	TN,[h] n	Sen[i] (lower-upper)	Spe[j] (lower-upper)
Donovan, 2015 [62]	PB	Urine	195	195 (89, 106)	None	80	9	84	22	0.899 (0.817-0.947)	0.208 (0.141-0.295)
Roberts, 2015 [16]	PB	Semen	66	66 (12, 54)	Crossval	11	1	32	20	0.917 (0.587-0.988)	0.385 (0.263-0.522)
Zhang, 2015 [63]	PB	Serum	580	580 (180, 400)	320 (120, 200)	84	36	5	195	0.7 (0.612-0.775)	0.975 (0.941-0.99)
Mengual, 2016 [64]	PB	Urine	224	224 (15, 73)	Crossval	116	35	12	61	0.768 (0.694-0.829)	0.836 (0.732-0.904)
Salido-Guadarrama, 2016 [15]	PB	Urine	143	143 (73, 70)	None	60	13	13	57	0.822 (0.717-0.894)	0.814 (0.706-0.889)
Dereziński, 2017 [17]	PB	Serum	89	89 (49, 40)	34 (19, 15)	13	6	0	15	0.675 (0.449-0.841)	0.969 (0.65-0.998)
Dereziński, 2017a [17]	PB	Urine	89	89 (49, 40)	34 (19,15)	17	2	4	11	0.895 (0.663-0.974)	0.733 (0.467-0.896)
Kirby, 2017 [60]	LB	Tissue	101	398 (286, 112)	262 (213, 49)	180	33	4	45	0.845 (0.79-0.888)	0.918 (0.802-0.969)
Barceló, 2018 [65]	PB	Semen	42	42 (34, 18)	None	22	2	5	13	0.917 (0.721-0.979)	0.722 (0.481-0.879)
Amante, 2019 [66]	PB	Urine	91	91 (43, 48)	Crossval	40	3	5	43	0.93 (0.805-0.977)	0.896 (0.773-0.956)
Brikun, 2019 [67]	PB	Urine	94	94 (42, 52)	29 (13, 16)	12	1	5	11	0.923 (0.609-0.989)	0.687 (0.433-0.864)
Gao, 2019 [69]	PB	Urine	183	183 (108, 75)	77 (55, 22)	48	7	5	17	0.873 (0.756-0.938)	0.773 (0.556-0.902)
Patel, 2019 [61]	LB	Tissue	699	795 (699, 96)	242 (212, 30)	199	13	2	28	0.939 (0.897-0.964)	0.933 (0.769-0.983)
Santotoribio, 2019 [68]	PB	Serum	232	232 (32, 200)	None	30	2	58	142	0.937 (0.782-0.984)	0.71 (0.643-0.769)

[a]All studies employed regression-based models.

[b]LB: lesion-based model; PB: patient-based model.

[c]PCa: prostate cancer.

[d]Crossval: cross-validation techniques.

[e]TP: true-positive.

[f]FN: false-negative.

[g]FP: false-positive.

[h]TN: true-negative.

[i]Sen: sensitivity.

[j]Spe: specificity.

The subgroup analysis among studies that employed internal cross-validation techniques (subgroup 1) [16,64,66], split validation approaches (subgroup 2) [17,60,61,63,67,69], and no validation (subgroup 3) [15,62,65,68] is shown in Figure 10. For subgroups 1 and 2, the heterogeneity was greater than 50%. In subgroup 3, the heterogeneity was around 20%.

Figure 9. Subgroup analysis for the model-based covariate in genomic studies. Subgroup 1: lesion-based models; subgroup 2: patient-based models. FN: false-negative; FP: false-positive; TN: true-negative; TP: true-positive.

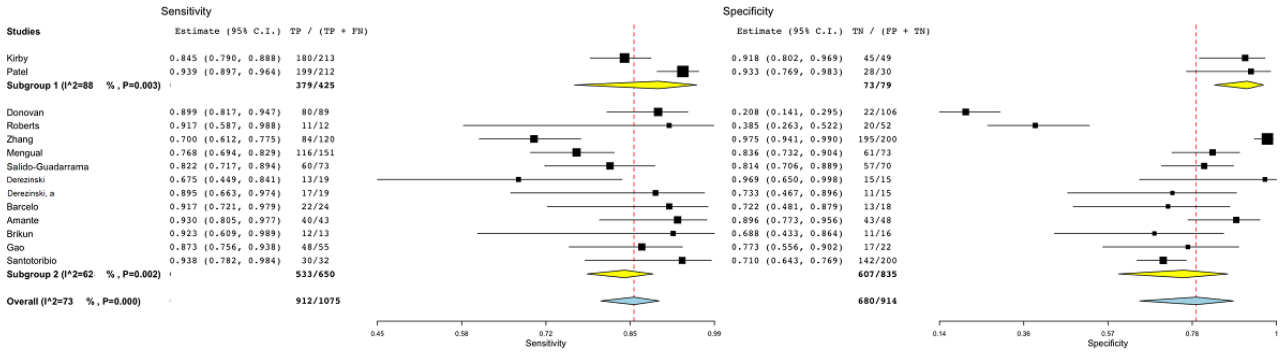
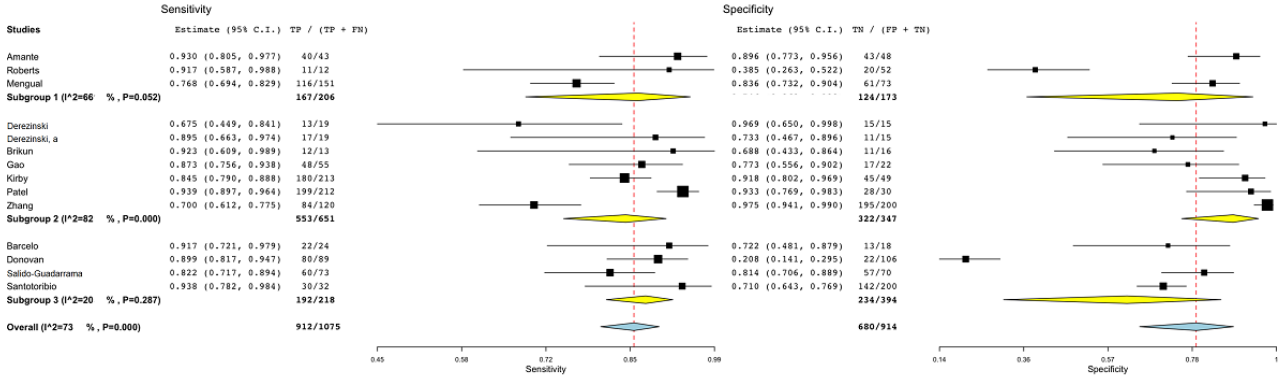


Figure 10. Subgroup analysis for the validation covariate in genomic studies. Subgroup 1: internal cross-validation; subgroup 2: hold-out approach or external validation; subgroup 3: no validation. FN: false-negative; FP: false-positive; TN: true-negative; TP: true-positive.



A subgroup analysis was also carried out based on the specimen used by the genomic studies (ie, urine [15,17,62,64,66,67,69], semen [16,65], serum [17,63,68], and tissue [60,61] biomarkers). The subgroup of studies investigating urine biomarkers to automatically detect PCa presented a lower heterogeneity than studies employing tissue and serum biomarkers and included more than 5 studies (Figure 11).

An inspection of ML algorithms among genomic studies was not possible because all the included studies employed a regression-based model (Table S4 in Multimedia Appendix 1).

Finally, the effect of using balanced or highly unbalanced data sets in ML approaches was investigated (Figure 12). Seven studies were included in subgroup 2, as they employed highly unbalanced data sets. The heterogeneity of subgroup 1 was around 36%, whereas subgroup 2 showed a high heterogeneity (I^2=84%, P<.001).

As a result, among several covariates, the imbalance covariate was the only one by which the heterogeneity could be partially resolved for more than 5 studies.

By inspecting Figure 12, Donovan et al [62] presented a very low value for specificity; this was due to the fact that they fixed the sensitivity threshold value at 90%.

Five studies employing urine specimens and balanced data sets showed a very low heterogeneity (Figure 13) [15,17,66,67,69].

The HSROC curve for the studies employing balanced data sets to automatically detect PCa via urine biomarkers is shown in Figure 14. The pooled sensitivity and specificity were 0.812 (95% CI 0.577-0.999) and 0.8101 (95% CI 0.544-0.999), respectively.

Figure 11. Subgroup analysis for the predictor covariate in genomic studies. FN: false-negative; FP: false-positive; TN: true-negative; TP: true-positive.

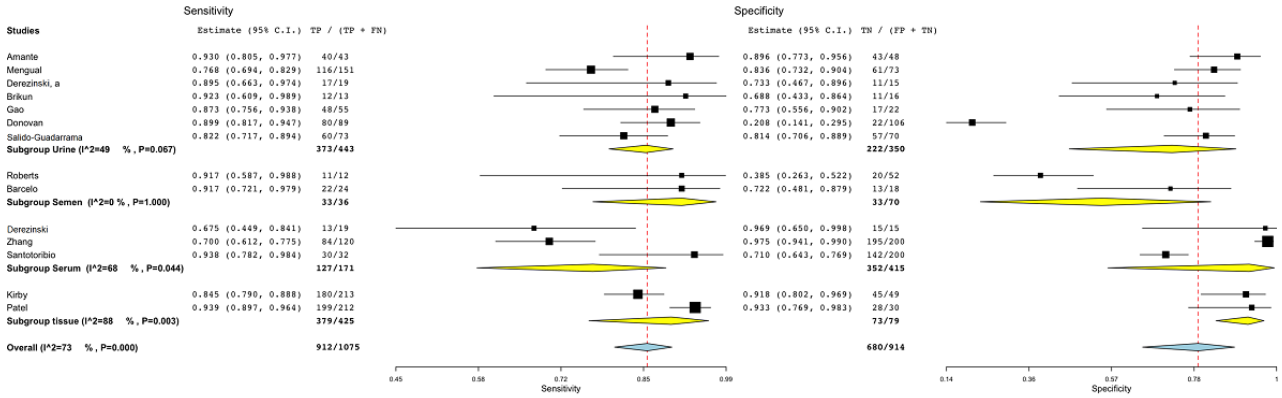


Figure 12. Subgroup analysis for the imbalance covariate in genomic studies. Subgroup 1: balanced data sets; subgroup 2: unbalanced data sets. FN: false-negative; FP: false-positive; TN: true-negative; TP: true-positive.

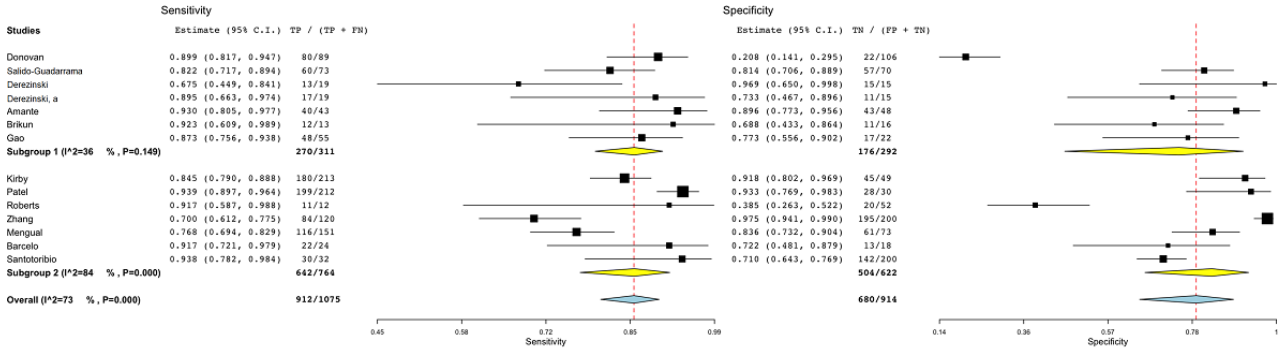


Figure 13. Coupled forest plots for balanced studies. The included studies investigated urine specimens. FN: false-negative; FP: false-positive; TN: true-negative; TP: true-positive.

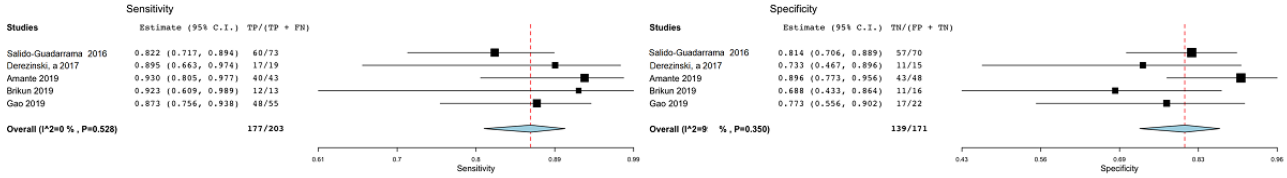
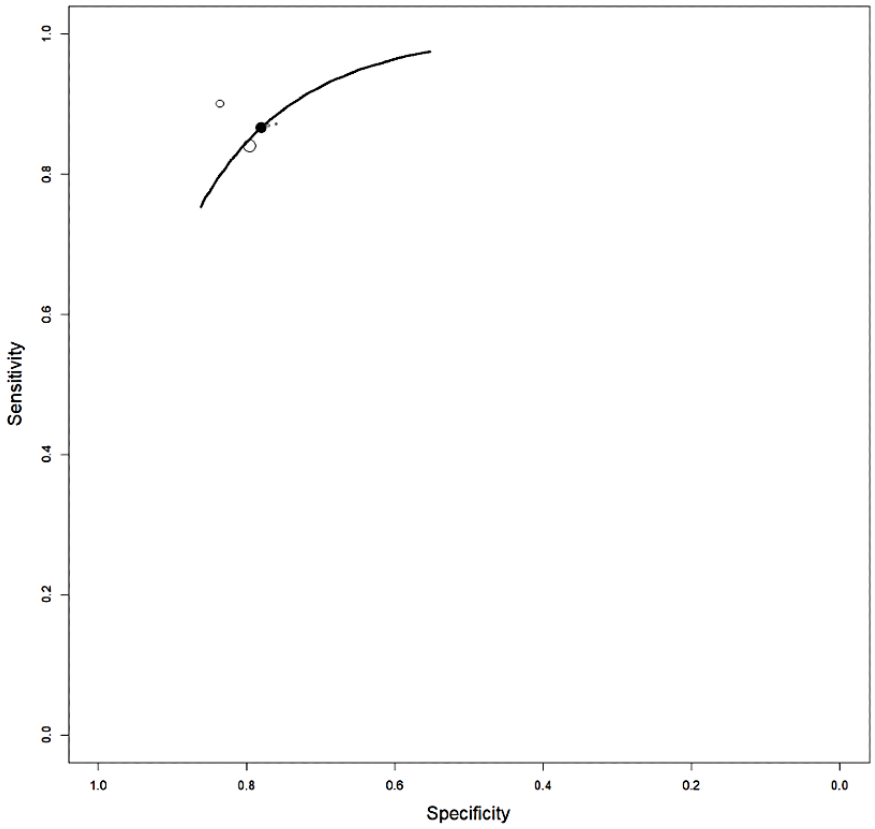


Figure 14. Hierarchical summary receiver operating characteristic curve (HSROC) for a subset of genomic studies. HSROC was calculated for genomic studies with low heterogeneity [15,17,66,67,69].



Discussion

Principal Findings

This paper presents the results of a systematic literature review with meta-analysis of articles investigating machine learning algorithms to detect PCa via radiomic or genomic analysis. One research focus of this study was on clearly evaluating how the implementation of different ML approaches impacts the clinical results. At this stage, due to the high heterogeneity of methods and tools employed in the existing literature, no clear clinical relevance on the use of ML for PCa can be drawn from this study. This review shows that ML has helped to improve the diagnostic performance of the detection of PCa, but challenges still remain for clinical applicability of such methods, and more research is needed. The presented literature aims to help in building an ML system that is robust and computationally efficient to assist clinicians in the diagnosis of PCa via radiomic and genomic biomarkers.

In this review, 37 studies were shortlisted, and 29 studies were included in a meta-analysis. All patients were diagnosed with PCa by biopsy. However, not all the included studies reported full information on the methods used to carry out biopsy (eg, direct MRI-guided, cognitive fusion, or MRI-TRUS fusion biopsy).

In the radiomic and genomic meta-analysis, 16 and 14 studies were included, respectively. Heterogeneity among radiomic and genomic studies was 84% and 73%, respectively. This was expected, as ML methods are usually regarded as black boxes, and the consideration of all possible transformations is onerous. Moreover, there are no clear guidelines on how to develop AI approaches for medical studies, even though a few recommendations have been summarized by Foster et al [24] and Chen et al [25]. Another font of heterogeneity in radiomic studies may be due to the inclusion of PI-RADS score 3 and Gleason score 3+3 lesions, which are equivocal and should be disregarded in classification processes.

To partially solve the heterogeneity for the included studies, subgroup analyses were conducted based on several covariates. In the field of ML, applications where repeated measures or records have been captured on each subject can affect the overall performance. In most studies, the main aim is to predict if a given subject is “sick” or a “control” subject. In these applications, each subject has a single label type (eg, “sick” or control case). Nonetheless, there are other classification problems where each subject can have multiple labels. For instance, multiple lesions can be extracted from the same subject, and the control part can be represented by the benign-adjacent prostate lesion. It has been demonstrated that this phenomenon, known as identity confounding, can cause discrepancy in classification performance [70,71]. Therefore, the studies included in the meta-analysis were investigated to determine whether they explored patient- or lesion-based models. A patient-based model could be defined as a model that is developed and assessed in a “subject-wise” fashion, where all the records of each subject are considered as a group in the training and testing set and when assessing the model performance; conversely, a lesion-based model could be defined as a model that is developed and assessed in a “record-wise” fashion, where each measurement or record contributes to both

the training and test sets and when assessing the model performance [70].

In both radiomic and genomic studies, patient-based models presented lower heterogeneity and performance than lesion-based models; this could be due to the fact that lesion-based models employed a bigger size sample, but the models may be overfit due to repeated measures.

A second important covariate to examine in ML problems is the data set construction. In particular, the data set is usually divided into training and testing sets in order to reduce overfitting problems [70,71]. The training set is often further split into a training set and a validation set, which is used to update model parameters. At least one procedure of internal or external validation is required in ML approaches. Cross-validation techniques are preferred if availability of data is not a problem. It is also strongly suggested to retrain on a subset of data or use an independent data set for external testing. Therefore, "validation approach" was used as a covariate in subgroup analysis. Validation approaches were divided into cross-validation, hold-out approach (split) or external validation, and no validation. In both radiomic and genomic analysis, studies employing cross-validation techniques and hold-out approaches had very high heterogeneity and similar performances among them. High heterogeneity may be due to the different cross-validation techniques used (eg, bootstrapping [16,40,52], Monte Carlo cross-validation [17]) or the choice of number of folders used in cross-validation methods; if an external data set was used [52,60,61,63], differences in the study protocols may have increased the bias among studies. Moreover, few studies in radiomic [50,53,57,59] and genomic [17,67] analysis employed both cross-validation and external testing. Studies employing no validation showed very low heterogeneity (only 2 studies in radiomic analysis), which may be due to the absence of other confounding variables, and high performances may be due to overfitting problems. A lower specificity was only noted in genomic analysis; this was due to Donovan et al [62], which used a fixed threshold for sensitivity at 90%.

Different ML approaches were also investigated among radiomic studies as a possible covariate factor. There were no relevant differences in heterogeneity or performance among subgroups (Figure 4). All genomic studies employed regression-based models. In fact, one limitation of the genomic studies was that none of the selected studies explored the potential of ML techniques at full capacity. Subgroup analysis was also conducted among radiomic studies employing ML or DL (ie, based on artificial neural networks) approaches. As expected, heterogeneity among DL studies was higher than among the studies employing other ML approaches to detect PCa. This could be mainly due to the high complexity of DL methods and hyperparameters. Moreover, DL approaches showed lower performance due to the small sample sizes used; they need large volumes of data to automatically identify patterns and achieve high performance.

The imbalance covariate was crucial in this study. Unbalanced and small data sets are very common in the medical field, and ML algorithms tend to produce unsatisfactory classifiers when handled with imbalanced data sets. Therefore, several techniques to overcome this problem have been proposed over time [72]. In this review, none of the studies included in the subgroup of unbalanced data sets had used any techniques to overcome the problem. Only one study [56] used SMOTE, but it did not employ a highly unbalanced data set.

For radiomic studies, after excluding studies that employed highly unbalanced data sets, the heterogeneity was less than 50%. The final pooled sensitivity and specificity for the use of mpMRI were 0.808 (95% CI 0.38-0.999) and 0.831 (95% CI 0.41-0.999), respectively.

For genomic studies, the heterogeneity dropped to 36% and reached a value close to zero when Donovan et al [62] was excluded because they fixed a threshold of 90% for sensitivity. The final pooled sensitivity and specificity were 0.812 (95% CI 0.577-0.999) and 0.8101 (95% CI 0.544-0.999), respectively. The predictor used to estimate the final pooled sensitivity and specificity was urine specimen.

Only 4 studies [18,39-41] investigating clinically based models were identified through the search. All the included studies adopted internal validation techniques (3 cross-validation [39-41] and 1 internal split validation [18]). Two studies [40,41] employed regression-based models, one [39] employed a tree-based model, and lastly, one employed a DL approach [18]. Heterogeneity was very high among them (I^2=96%, P=.01) due to different sample sizes and diversity of predictors. However, contributions from genomic and imaging biomarkers should be considered to improve the overall performance of the clinically based diagnostic models.

Comparison among genomic and radiomic studies was not possible because they describe two different but complementary prospective approaches to the disease. However, the pooled sensitivity and specificity for both mpMRI and urine biomarkers were around 80%, showing them to be promising biomarkers in the detection of PCa via ML in clinical practice. The use of mpMRI has shown great diagnostic potential [73]; however, its analysis and interpretation are quite challenging, and there is not a consensus on how to optimally extract significant information. On the other side, genomic analyses have significantly increased our understanding of PCa and greatly improved patient risk classification, thus impacting treatment decision making. Therefore, a new prospective approach is the integration of radiomic and genomic signatures, commonly known as radiogenomics [74-76], in order to improve the overall performance of diagnostic tools to automatically detect PCa. In the existing literature, only a few studies have investigated "radiophenotypes" to complement existing validated clinical and genomic risk stratification biomarkers [77-79].

In this scenario, a typical ML postprocessing pipeline for radiomic and genomic analysis to automatically detect PCa may be constituted of a few crucial steps. In the case of radiomic studies, a common pipeline may be constituted of (1) examination of mpMRI; (2) image segmentation through the delineation of ROIs or VOIs, which can include whole gland volume, a specific zone, and one or multiple lesions, which should be explicitly specified in the manuscript; (3) image preprocessing; (4) filtering; (5) feature extraction; (6) integration

of radiomic data with clinical data, genomic data, or both; (7) feature selection in relation to the target class; and (8) algorithm training, validation, and testing. Alternatively, a DL approach would only require the examination of the images and annotation of the ROIs or VOIs of the whole image, according to the desired classification output.

The image processing pipeline should be carefully described in the manuscripts, and the spatial coregistration of DWIs is a critical factor in the correct analysis of diffusion tensor imaging data, which has often been used as a predictor of PCa diagnosis. Moreover, the use of endorectal coil can cause high deformation of the prostate compared with other coils and may not provide adequate MR image quality [80]. Therefore, further processing of the images should also be considered, especially when the study is multicenter and different protocols have been adopted.

Due to the high heterogeneity of genomic studies, a standard pipeline configuration could be structured into (1) missing value management; (2) filtering to remove low-variance features; (3) data normalization due to data coming from heterogeneous formats; (4) a feature selection step to remove irrelevant features due to the high dimension of data; (5) dealing with class imbalance distribution present in this type of large-scale data set; and (6) algorithm training, validation, and testing. Alternatively, a DL approach would handle filtering and feature selection to generate handcrafted features. Deep learning is a powerful tool to integrate different "omics" and increase the computational power of diagnostic tools.

Further general recommendations on how to avoid bias and pitfalls in applying ML to medical problems are as follows: (1) in the case of multicenter studies, it is recommended to use batch effect approaches to prevent any bias due to different study protocols and feature normalization procedures to reduce within-subject bias [81]; and (2) for classifier performance, it is necessary to report if any threshold has been used to identify sensitivity and specificity and whether the performance was reported on patient-based or lesion-based data sets.

Limitations

Our study presents several limitations. Some variability still remains due to the actual thresholds between studies. However, the multiple hierarchical model accounts for between- and within-subject variability among studies, including threshold effects. Another factor that could have affected the heterogeneity among studies is the use of different predictors among radiomic and genomic studies. Moreover, several studies reported little or incomplete information on the parameters used to develop ML models. Therefore, the number of parameters that are estimated by each technique was not investigated as a possible source of heterogeneity among studies. Additional heterogeneity in the observed results is due to the variability of calibration differences between equipment and differences between readers or observers, as well as variation in the implementation of tests. Another possible bias may be due to the preprocessing techniques on the extracted data and feature selection and feature normalization methods.

We limited the search to English-only studies; although this is common in systematic reviews, this exclusion criterion could have reduced the generalizability of the findings. However, the extent and effects of language bias have recently diminished because of a shift toward publication of studies in English [82]. At this stage, we also excluded PCa risk stratification studies to reduce bias and heterogeneity among studies, but further investigation on the use of ML methods to assess risk stratification biomarkers could give a comparative perspective on the treatment selection.

Finally, publication bias was not assessed in our analysis, as there are currently no statistically adequate models in the field of meta-analysis of diagnostic test accuracy [29].

Conclusion

ML has shown its potential to empower clinicians in the detection of prostate cancer. The accuracy of ML algorithms for diagnosis of PCa was considered acceptable, in terms of heterogeneity, for 12 radiomic studies investigating mpMRI and 5 genomic studies using urine biomarkers.

However, given the limitations indicated in our study, further well-designed studies are warranted to extend the potential use of ML algorithms to clinical settings. Recommendations on the use of these techniques were also provided to help researchers to design robust studies aiming to identify radiomic and genomic biomarkers to detect cancer.

Acknowledgments

This work was supported by "Progetti di Ricerca Corrente" funded by the Italian Ministry of Health.

Authors' Contributions

RC, MF, and CC collected the data. All authors contributed to project development, data analysis, and the writing and editing of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary material.
[DOCX File , 984 KB - jmir_v23i4e22394_app1.docx]

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2018 Nov;68(6):394-424 [FREE Full text] [doi: 10.3322/caac.21492] [Medline: 30207593]
2. Grönberg H. Prostate cancer epidemiology. Lancet 2003 Mar 08;361(9360):859-864. [doi: 10.1016/S0140-6736(03)12713-4] [Medline: 12642065]
3. Mottet N, Bellmunt J, Bolla M, Briers E, Cumberbatch MG, De Santis M, et al. EAU-ESTRO-SIOG Guidelines on Prostate Cancer. Part 1: Screening, Diagnosis, and Local Treatment with Curative Intent. Eur Urol 2017 Apr;71(4):618-629. [doi: 10.1016/j.eururo.2016.08.003] [Medline: 27568654]
4. Rizzo S, Botta F, Raimondi S, Origgi D, Fanciullo C, Morganti AG, et al. Radiomics: the facts and the challenges of image analysis. Eur Radiol Exp 2018 Nov 14;2(1):36 [FREE Full text] [doi: 10.1186/s41747-018-0068-z] [Medline: 30426318]
5. Yao S, Jiang H, Song B. Radiomics in prostate cancer: basic concepts and current state-of-the-art. Chin J Acad Radiol 2019 Nov 12;2(3-4):47-55. [doi: 10.1007/s42058-019-00020-3]
6. Cuocolo R, Cipullo MB, Stanzione A, Ugga L, Romeo V, Radice L, et al. Machine learning applications in prostate cancer magnetic resonance imaging. Eur Radiol Exp 2019 Aug 07;3(1):35 [FREE Full text] [doi: 10.1186/s41747-019-0109-2] [Medline: 31392526]
7. Cantiello F, Russo GI, Kaufmann S, Cacciamani G, Crocerossa F, Ferro M, et al. Role of multiparametric magnetic resonance imaging for patients under active surveillance for prostate cancer: a systematic review with diagnostic meta-analysis. Prostate Cancer Prostatic Dis 2019 May;22(2):206-220. [doi: 10.1038/s41391-018-0113-2] [Medline: 30487646]
8. Weinreb J, Barentsz JO, Choyke PL, Cornud F, Haider MA, Macura KJ, et al. PI-RADS Prostate Imaging - Reporting and Data System: 2015, Version 2. Eur Urol 2016 Jan;69(1):16-40 [FREE Full text] [doi: 10.1016/j.eururo.2015.08.052] [Medline: 26427566]
9. Chin L, Andersen JN, Futreal PA. Cancer genomics: from discovery science to personalized medicine. Nat Med 2011 Mar;17(3):297-303. [doi: 10.1038/nm.2323] [Medline: 21383744]
10. Kornberg Z, Cooperberg MR, Spratt DE, Feng FY. Genomic biomarkers in prostate cancer. Transl Androl Urol 2018 Jun;7(3):459-471 [FREE Full text] [doi: 10.21037/tau.2018.06.02] [Medline: 30050804]
11. Gleason DF. Classification of prostatic carcinomas. Cancer Chemother Rep 1966 Mar;50(3):125-128. [Medline: 5948714]
12. Jiandani D, Randhawa A, Brown RE, Hamilton R, Matthew AG, Kuk JL, et al. The effect of bicycling on PSA levels: a systematic review and meta-analysis. Prostate Cancer Prostatic Dis 2015 Sep;18(3):208-212. [doi: 10.1038/pcan.2015.16] [Medline: 25939515]
13. Kundu SD, Roehl KA, Yu X, Antenor JAV, Suarez BK, Catalona WJ. Prostate specific antigen density correlates with features of prostate cancer aggressiveness. J Urol 2007 Feb;177(2):505-509. [doi: 10.1016/j.juro.2006.09.039] [Medline: 17222621]
14. Fang Y, Gao W. Roles of microRNAs during prostatic tumorigenesis and tumor progression. Oncogene 2014 Jan 09;33(2):135-147. [doi: 10.1038/onc.2013.54] [Medline: 23455326]
15. Salido-Guadarrama AI, Morales-Montor JG, Rangel-Escareño C, Langley E, Peralta-Zaragoza O, Cruz Colin JL, et al. Urinary microRNA-based signature improves accuracy of detection of clinically relevant prostate cancer within the prostate-specific antigen grey zone. Mol Med Rep 2016 Jun;13(6):4549-4560 [FREE Full text] [doi: 10.3892/mmr.2016.5095] [Medline: 27081843]
16. Roberts MJ, Chow CWK, Schirra HJ, Richards R, Buck M, Selth LA, et al. Diagnostic performance of expression of PCA3, Hepsin and miR biomarkers inejaculate in combination with serum PSA for the detection of prostate cancer. Prostate 2015 Apr 01;75(5):539-549. [doi: 10.1002/pros.22942] [Medline: 25597828]
17. Dereziński P, Klupczynska A, Sawicki W, Pałka JA, Kokot ZJ. Amino Acid Profiles of Serum and Urine in Search for Prostate Cancer Biomarkers: a Pilot Study. Int J Med Sci 2017;14(1):1-12 [FREE Full text] [doi: 10.7150/ijms.15783] [Medline: 28138303]
18. Roffman DA, Hart GR, Leapman MS, Yu JB, Guo FL, Ali I, et al. Development and Validation of a Multiparameterized Artificial Neural Network for Prostate Cancer Risk Prediction and Stratification. JCO Clin Cancer Inform 2018 Dec;2:1-10 [FREE Full text] [doi: 10.1200/CCI.17.00119] [Medline: 30652591]
19. Bini SA. Artificial Intelligence, Machine Learning, Deep Learning, and Cognitive Computing: What Do These Terms Mean and How Will They Impact Health Care? J Arthroplasty 2018 Aug;33(8):2358-2361. [doi: 10.1016/j.arth.2018.02.067] [Medline: 29656964]
20. Goldenberg SL, Nir G, Salcudean SE. A new era: artificial intelligence and machine learning in prostate cancer. Nat Rev Urol 2019 Jul;16(7):391-403. [doi: 10.1038/s41585-019-0193-3] [Medline: 31092914]
21. Meskó B, Görög M. A short guide for medical professionals in the era of artificial intelligence. NPJ Digit Med 2020;3:126 [FREE Full text] [doi: 10.1038/s41746-020-00333-z] [Medline: 33043150]
22. Trivizakis E, Papadakis G, Souglakos I, Papanikolaou N, Koumakis L, Spandidos D, et al. Artificial intelligence radiogenomics for advancing precision and effectiveness in oncologic care (Review). Int J Oncol 2020 Jul;57(1):43-53 [FREE Full text] [doi: 10.3892/ijo.2020.5063] [Medline: 32467997]

23. Zitnik M, Nguyen F, Wang B, Leskovec J, Goldenberg A, Hoffman MM. Machine Learning for Integrating Data in Biology and Medicine: Principles, Practice, and Opportunities. Inf Fusion 2019 Oct;50:71-91 [FREE Full text] [doi: 10.1016/j.inffus.2018.09.012] [Medline: 30467459]
24. Foster KR, Koprowski R, Skufca JD. Machine learning, medical diagnosis, and biomedical engineering research - commentary. Biomed Eng Online 2014 Jul 05;13:94 [FREE Full text] [doi: 10.1186/1475-925X-13-94] [Medline: 24998888]
25. Chen PC, Liu Y, Peng L. How to develop machine learning models for healthcare. Nat Mater 2019 May;18(5):410-414. [doi: 10.1038/s41563-019-0345-0] [Medline: 31000806]
26. Luo W, Phung D, Tran T, Gupta S, Rana S, Karmakar C, et al. Guidelines for Developing and Reporting Machine Learning Predictive Models in Biomedical Research: A Multidisciplinary View. J Med Internet Res 2016 Dec 16;18(12):e323 [FREE Full text] [doi: 10.2196/jmir.5870] [Medline: 27986644]
27. Sandhu S, Lin AL, Brajer N, Sperling J, Ratliff W, Bedoya AD, et al. Integrating a Machine Learning System Into Clinical Workflows: Qualitative Study. J Med Internet Res 2020 Nov 19;22(11):e22421 [FREE Full text] [doi: 10.2196/22421] [Medline: 33211015]
28. Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, PRISMA-P Group. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Syst Rev 2015 Jan 01;4:1 [FREE Full text] [doi: 10.1186/2046-4053-4-1] [Medline: 25554246]
29. Whiting PF, Rutjes AWS, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, QUADAS-2 Group. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med 2011 Oct 18;155(8):529-536 [FREE Full text] [doi: 10.7326/0003-4819-155-8-201110180-00009] [Medline: 22007046]
30. Viswanath SE, Bloch NB, Chappelow JC, Toth R, Rofsky NM, Genega EM, et al. Central gland and peripheral zone prostate tumors have significantly different quantitative imaging signatures on 3 Tesla endorectal, in vivo T2-weighted MR imagery. J Magn Reson Imaging 2012 Jul;36(1):213-224 [FREE Full text] [doi: 10.1002/jmri.23618] [Medline: 22337003]
31. Monti S, Brancato V, Di Costanzo G, Basso L, Puglia M, Ragozzino A, et al. Multiparametric MRI for Prostate Cancer Detection: New Insights into the Combined Use of a Radiomic Approach with Advanced Acquisition Protocol. Cancers 2020 Feb 07;12(2):390. [doi: 10.3390/cancers12020390]
32. Sweeting MJ, Sutton AJ, Lambert PC. What to add to nothing? Use and avoidance of continuity corrections in meta-analysis of sparse data. Stat Med 2004 May 15;23(9):1351-1375. [doi: 10.1002/sim.1761] [Medline: 15116347]
33. Lee J, Kim KW, Choi SH, Huh J, Park SH. Systematic Review and Meta-Analysis of Studies Evaluating Diagnostic Test Accuracy: A Practical Review for Clinical Researchers-Part II. Statistical Methods of Meta-Analysis. Korean J Radiol 2015;16(6):1188-1196 [FREE Full text] [doi: 10.3348/kjr.2015.16.6.1188] [Medline: 26576107]
34. Richardson M, Garner P, Donegan S. Interpretation of subgroup analyses in systematic reviews: A tutorial. Clinical Epidemiology and Global Health 2019 Jun;7(2):192-198. [doi: 10.1016/j.cegh.2018.05.005]
35. Bonekamp D, Kohl S, Wiesenfarth M, Schelb P, Radtke JP, Götz M, et al. Radiomic Machine Learning for Characterization of Prostate Lesions with MRI: Comparison to ADC Values. Radiology 2018 Oct;289(1):128-137. [doi: 10.1148/radiol.2018173064] [Medline: 30063191]
36. Wang J, Wu C, Bao M, Zhang J, Wang X, Zhang Y. Machine learning-based analysis of MR radiomics can help to improve the diagnostic performance of PI-RADS v2 in clinically relevant prostate cancer. Eur Radiol 2017 Oct;27(10):4082-4090. [doi: 10.1007/s00330-017-4800-5] [Medline: 28374077]
37. Wu M, Krishna S, Thornhill RE, Flood TA, McInnes MD, Schieda N. Transition zone prostate cancer: Logistic regression and machine-learning models of quantitative ADC, shape and texture features are highly accurate for diagnosis. J Magn Reson Imaging 2019 Sep;50(3):940-950. [doi: 10.1002/jmri.26674] [Medline: 30701625]
38. Yang X, Liu C, Wang Z, Yang J, Min HL, Wang L, et al. Co-trained convolutional neural networks for automated detection of prostate cancer in multi-parametric MRI. Med Image Anal 2017 Dec;42:212-227. [doi: 10.1016/j.media.2017.08.006] [Medline: 28850876]
39. Xiao L, Chen P, Gou Z, Li Y, Li M, Xiang L, et al. Prostate cancer prediction using the random forest algorithm that takes into account transrectal ultrasound findings, age, and serum levels of prostate-specific antigen. Asian J Androl 2017;19(5):586-590 [FREE Full text] [doi: 10.4103/1008-682X.186884] [Medline: 27586028]
40. Nowroozi MR, Amini S, Kasaeian A, Zavarehei MJ, Eshraghian MR, Ayati M. Development, validation and comparison of two nomograms predicting prostate cancer at initial 12-core biopsy. Asia Pac J Clin Oncol 2016 Jun;12(2):e289-e297. [doi: 10.1111/ajco.12186] [Medline: 24684767]
41. Huang C, Song G, Wang H, Ji G, Li J, Chen Y, et al. MultiParametric Magnetic Resonance Imaging-Based Nomogram for Predicting Prostate Cancer and Clinically Significant Prostate Cancer in Men Undergoing Repeat Prostate Biopsy. Biomed Res Int 2018;2018:6368309 [FREE Full text] [doi: 10.1155/2018/6368309] [Medline: 30276213]
42. Macaskill P. Cochrane handbook for systematic reviews of diagnostic test accuracy, Version 0.9.0. London: The Cochrane Collaboration; 2010. URL: https://methods.cochrane.org/sdt/ [accessed 2019-02-20]
43. Jackson D, Turner R. Power analysis for random-effects meta-analysis. Res Synth Methods 2017 Sep;8(3):290-302 [FREE Full text] [doi: 10.1002/jrsm.1240] [Medline: 28378395]

44. Zhao K, Wang C, Hu J, Yang X, Wang H, Li F, et al. Prostate cancer identification: quantitative analysis of T2-weighted MR images based on a back propagation artificial neural network model. Sci China Life Sci 2015 Jul;58(7):666-673. [doi: 10.1007/s11427-015-4876-6] [Medline: 26025283]
45. Valerio M, Zini C, Fierro D, Giura F, Colarieti A, Giuliani A, et al. 3T multiparametric MRI of the prostate: Does intravoxel incoherent motion diffusion imaging have a role in the detection and stratification of prostate cancer in the peripheral zone? Eur J Radiol 2016 Apr;85(4):790-794. [doi: 10.1016/j.ejrad.2016.01.006] [Medline: 26971425]
46. Lay N, Tsehay Y, Greer MD, Turkbey B, Kwak JT, Choyke PL, et al. Detection of prostate cancer in multiparametric MRI using random forest with instance weighting. J Med Imaging (Bellingham) 2017 Apr;4(2):024506 [FREE Full text] [doi: 10.1117/1.JMI.4.2.024506] [Medline: 28630883]
47. Reda I, Shalaby A, Elmogy M, Elfotouh AA, Khalifa F, El-Ghar MA, et al. A comprehensive non-invasive framework for diagnosing prostate cancer. Comput Biol Med 2017 Feb 01;81:148-158. [doi: 10.1016/j.compbiomed.2016.12.010] [Medline: 28063376]
48. Starobinets O, Simko JP, Kuchinsky K, Kornak J, Carroll PR, Greene KL, et al. Characterization and stratification of prostate lesions based on comprehensive multiparametric MRI using detailed whole-mount histopathology as a reference standard. NMR Biomed 2017 Dec;30(12):e3796. [doi: 10.1002/nbm.3796] [Medline: 28961382]
49. Kwon D, Reis IM, Breto AL, Tschudi Y, Gautney N, Zavala-Romero O, et al. Classification of suspicious lesions on prostate multiparametric MRI using machine learning. J Med Imaging (Bellingham) 2018 Jul;5(3):034502 [FREE Full text] [doi: 10.1117/1.JMI.5.3.034502] [Medline: 30840719]
50. Song Y, Zhang Y, Yan X, Liu H, Zhou M, Hu B, et al. Computer-aided diagnosis of prostate cancer using a deep convolutional neural network from multiparametric MRI. J Magn Reson Imaging 2018 Dec;48(6):1570-1577. [doi: 10.1002/jmri.26047] [Medline: 29659067]
51. Devine W, Giganti F, Johnston EW, Sidhu HS, Panagiotaki E, Punwani S, et al. Simplified Luminal Water Imaging for the Detection of Prostate Cancer From Multiecho T MR Images. J Magn Reson Imaging 2019 Sep;50(3):910-917 [FREE Full text] [doi: 10.1002/jmri.26608] [Medline: 30566264]
52. Le MH, Chen J, Wang L, Wang Z, Liu W, Cheng KT, et al. Automated diagnosis of prostate cancer in multi-parametric MRI based on multimodal convolutional neural networks. Phys Med Biol 2017 Jul 24;62(16):6497-6514. [doi: 10.1088/1361-6560/aa7731] [Medline: 28582269]
53. Mazaheri Y, Shukla-Dave A, Goldman DA, Moskowitz CS, Takeda T, Reuter VE, et al. Characterization of prostate cancer with MR spectroscopic imaging and diffusion-weighted imaging at 3 Tesla. Magn Reson Imaging 2019 Jan;55:93-102 [FREE Full text] [doi: 10.1016/j.mri.2018.08.025] [Medline: 30176373]
54. Gholizadeh N, Greer PB, Simpson J, Denham J, Lau P, Dowling J, et al. Characterization of prostate cancer using diffusion tensor imaging: A new perspective. Eur J Radiol 2019 Jan;110:112-120. [doi: 10.1016/j.ejrad.2018.11.026] [Medline: 30599846]
55. Wang X, Yang W, Weinreb J, Han J, Li Q, Kong X, et al. Searching for prostate cancer by fully automated magnetic resonance imaging classification: deep learning versus non-deep learning. Sci Rep 2017 Nov 13;7(1):15415 [FREE Full text] [doi: 10.1038/s41598-017-15720-y] [Medline: 29133818]
56. Chen T, Li M, Gu Y, Zhang Y, Yang S, Wei C, et al. Prostate Cancer Differentiation and Aggressiveness: Assessment With a Radiomic-Based Model vs. PI-RADS v2. J Magn Reson Imaging 2019 Mar;49(3):875-884 [FREE Full text] [doi: 10.1002/jmri.26243] [Medline: 30230108]
57. Qi Y, Zhang S, Wei J, Zhang G, Lei J, Yan W, et al. Multiparametric MRI-Based Radiomics for Prostate Cancer Screening With PSA in 4-10 ng/mL to Reduce Unnecessary Biopsies. J Magn Reson Imaging 2020 Jun;51(6):1890-1899. [doi: 10.1002/jmri.27008] [Medline: 31808980]
58. Ma X, Lv K, Sheng J, Yu Y, Pang P, Xu M, et al. Application evaluation of DCE-MRI combined with quantitative analysis of DWI for the diagnosis of prostate cancer. Oncol Lett 2019 Mar;17(3):3077-3084 [FREE Full text] [doi: 10.3892/ol.2019.9988] [Medline: 30867737]
59. Zhang S. Radiomics analysis potentially reduces over-diagnosis of prostate cancer with PSA levels of 4-10 ng/ml based on DWI data. 2019 Presented at: Medical Imaging 2019: Computer-Aided Diagnosis; 2019; San Diego, California, United States. [doi: 10.1117/12.2511497]
60. Kirby MK, Ramaker RC, Roberts BS, Lasseigne BN, Gunther DS, Burwell TC, et al. Genome-wide DNA methylation measurements in prostate tissues uncovers novel prostate cancer diagnostic biomarkers and transcription factor binding patterns. BMC Cancer 2017 Apr 17;17(1):273 [FREE Full text] [doi: 10.1186/s12885-017-3252-2] [Medline: 28412973]
61. Patel PG, Wessel T, Kawashima A, Okello JBA, Jamaspishvili T, Guérard KP, et al. A three-gene DNA methylation biomarker accurately classifies early stage prostate cancer. Prostate 2019 Oct;79(14):1705-1714. [doi: 10.1002/pros.23895] [Medline: 31433512]
62. Donovan MJ, Noerholm M, Bentink S, Belzer S, Skog J, O'Neill V, et al. A molecular signature of PCA3 and ERG exosomal RNA from non-DRE urine is predictive of initial prostate biopsy result. Prostate Cancer Prostatic Dis 2015 Dec;18(4):370-375. [doi: 10.1038/pcan.2015.40] [Medline: 26345389]
63. Zhang W, Zheng X, Wang X. Oxidative stress measured by thioredoxin reductase level as potential biomarker for prostate cancer. Am J Cancer Res 2015;5(9):2788-2798 [FREE Full text] [Medline: 26609484]

64. Mengual L, Lozano JJ, Ingelmo-Torres M, Izquierdo L, Musquera M, Ribal MJ, et al. Using gene expression from urine sediment to diagnose prostate cancer: development of a new multiplex mRNA urine test and validation of current biomarkers. BMC Cancer 2016 Feb 09;16:76 [FREE Full text] [doi: 10.1186/s12885-016-2127-2] [Medline: 26856686]
65. Barceló M, Castells M, Bassas L, Vigués F, Larriba S. Semen miRNAs Contained in Exosomes as Non-Invasive Biomarkers for Prostate Cancer Diagnosis. Sci Rep 2019 Sep 24;9(1):13772 [FREE Full text] [doi: 10.1038/s41598-019-50172-6] [Medline: 31551516]
66. Amante, Salomone, Alladio, Vincenti, Porpiglia, Bro. Untargeted Metabolomic Profile for the Detection of Prostate Carcinoma-Preliminary Results from PARAFAC2 and PLS-DA Models. Molecules 2019 Aug 22;24(17):3063 [FREE Full text] [doi: 10.3390/molecules24173063] [Medline: 31443574]
67. Brikun I, Nusskern D, Decatus A, Harvey E, Li L, Freije D. A panel of DNA methylation markers for the detection of prostate cancer from FV and DRE urine DNA. Clin Epigenetics 2018;10:91 [FREE Full text] [doi: 10.1186/s13148-018-0524-x] [Medline: 29988684]
68. Santotoribio JD, Jimenez-Romero ME. Serum biomarkers of inflammation for diagnosis of prostate cancer in patients with nonspecific elevations of serum prostate specific antigen levels. Transl. Cancer Res 2019 Feb;8(1):273-278. [doi: 10.21037/tcr.2019.01.31]
69. Gao Q, Su X, Annabi MH, Schreiter BR, Prince T, Ackerman A, et al. Application of Urinary Volatile Organic Compounds (VOCs) for the Diagnosis of Prostate Cancer. Clin Genitourin Cancer 2019 Jun;17(3):183-190. [doi: 10.1016/j.clgc.2019.02.003] [Medline: 30853355]
70. Chaibub Neto E, Pratap A, Perumal TM, Tummalacherla M, Snyder P, Bot BM, et al. Detecting the impact of subject characteristics on machine learning-based diagnostic applications. NPJ Digit Med 2019;2:99 [FREE Full text] [doi: 10.1038/s41746-019-0178-x] [Medline: 31633058]
71. Little MA, Varoquaux G, Saeb S, Lonini L, Jayaraman A, Mohr DC, et al. Using and understanding cross-validation strategies. Perspectives on Saeb et al. Gigascience 2017 May 01;6(5):1-6 [FREE Full text] [doi: 10.1093/gigascience/gix020] [Medline: 28327989]
72. Ganganwar V. An overview of classification algorithms for imbalanced datasets. International Journal of Emerging Technology and Advanced Engineering 2012;2(4):42-47.
73. Sathianathen NJ, Butaney M, Bongiorno C, Konety BR, Bolton DM, Lawrentschuk N. Accuracy of the magnetic resonance imaging pathway in the detection of prostate cancer: a systematic review and meta-analysis. Prostate Cancer Prostatic Dis 2019 Mar;22(1):39-48. [doi: 10.1038/s41391-018-0075-4] [Medline: 30108376]
74. Incoronato M, Aiello M, Infante T, Cavaliere C, Grimaldi A, Mirabelli P, et al. Radiogenomic Analysis of Oncological Data: A Technical Survey. Int J Mol Sci 2017 Apr 12;18(4):805 [FREE Full text] [doi: 10.3390/ijms18040805] [Medline: 28417933]
75. Zanfardino M, Franzese M, Pane K, Cavaliere C, Monti S, Esposito G, et al. Bringing radiomics into a multi-omics framework for a comprehensive genotype-phenotype characterization of oncological diseases. J Transl Med 2019 Oct 07;17(1):337 [FREE Full text] [doi: 10.1186/s12967-019-2073-2] [Medline: 31590671]
76. Zanfardino M, Pane K, Mirabelli P, Salvatore M, Franzese M. TCGA-TCIA Impact on Radiogenomics Cancer Research: A Systematic Review. Int J Mol Sci 2019 Nov 29;20(23):6033 [FREE Full text] [doi: 10.3390/ijms20236033] [Medline: 31795520]
77. Stoyanova R, Pollack A, Lynne C, Jorda M, Erho N, Lam LL, et al. Using radiogenomics to characterize MRI-guided prostate cancer biopsy heterogenity. JCO 2015 Mar 01;33(7_suppl):25-25. [doi: 10.1200/jco.2015.33.7_suppl.25]
78. Jamaspishvili T, Berman DM, Ross AE, Scher HI, De Marzo AM, Squire JA, et al. Clinical implications of PTEN loss in prostate cancer. Nat Rev Urol 2018 Apr;15(4):222-234 [FREE Full text] [doi: 10.1038/nrurol.2018.9] [Medline: 29460925]
79. VanderWeele DJ, McCann S, Fan X, Antic T, Jiang Y, Oto A. Radiogenomics of prostate cancer: Association between qunatitative multiparametric MRI features and PTEN. JCO 2015 Mar 01;33(7_suppl):126-126. [doi: 10.1200/jco.2015.33.7_suppl.126]
80. Martin GV, Kudchadker RJ, Bruno TL, Frank SJ, Wang J. Comparison of prostate distortion by inflatable and rigid endorectal MRI coils in permanent prostate brachytherapy imaging. Brachytherapy 2018;17(2):298-305. [doi: 10.1016/j.brachy.2017.09.014] [Medline: 29169971]
81. Castaldo R, Pane K, Nicolai E, Salvatore M, Franzese M. The Impact of Normalization Approaches to Automatically Detect Radiogenomic Phenotypes Characterizing Breast Cancer Receptors Status. Cancers (Basel) 2020 Feb 24;12(2):518 [FREE Full text] [doi: 10.3390/cancers12020518] [Medline: 32102334]
82. Morrison A, Polisena J, Husereau D, Moulton K, Clark M, Fiander M, et al. The effect of English-language restriction on systematic review-based meta-analyses: a systematic review of empirical studies. Int J Technol Assess Health Care 2012 Apr;28(2):138-144. [doi: 10.1017/S0266462312000086] [Medline: 22559755]

Abbreviations

AI: artificial intelligence
AUC: area under the curve

DCE: dynamic contrast-enhanced
DL: deep learning
DWI: diffusion-weighted imaging
FN: false-negative
FP: false-positive
HSROC: hierarchical summary receiver operating characteristic curve
ML: machine learning
mpMRI: multiparametric magnetic resonance imaging
PCa: prostate cancer
PI-RADS-V2: Prostate Imaging Reporting and Data System
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PSA: prostate-specific antigen
QUADAS-2: quality assessment of diagnostic accuracy studies–version 2
ROI: region of interest
SMOTE: synthetic minority oversampling technique
TN: true-negative
TP: true-positive
TRUS: transrectal ultrasonography
TZ: transition zone
VOI: volume of interest

Edited by G Eysenbach; submitted 10.07.20; peer-reviewed by G Fico, K Kowalewski, A Drory, DA Banerjee; comments to author 17.11.20; revised version received 26.11.20; accepted 17.01.21; published 01.04.21.

Please cite as:
Castaldo R, Cavaliere C, Soricelli A, Salvatore M, Pecchia L, Franzese M
Radiomic and Genomic Machine Learning Method Performance for Prostate Cancer Diagnosis: Systematic Literature Review
J Med Internet Res 2021;23(4):e22394
URL: https://www.jmir.org/2021/4/e22394
doi:10.2196/22394
PMID:33792552

Original Paper

Use of Endoscopic Images in the Prediction of Submucosal Invasion of Gastric Neoplasms: Automated Deep Learning Model Development and Usability Study

Chang Seok Bang[1,2,3,4], MD, PhD; Hyun Lim[1], MD, PhD; Hae Min Jeong[1], MD; Sung Hyeon Hwang[1], MD

[1]Department of Internal Medicine, Hallym University College of Medicine, Chuncheon, Republic of Korea

[2]Institute for Liver and Digestive Diseases, Hallym University, Chuncheon, Republic of Korea

[3]Institute of New Frontier Research, Hallym University College of Medicine, Chuncheon, Republic of Korea

[4]Division of Big Data and Artificial Intelligence, Chuncheon Sacred Heart Hospital, Chuncheon, Republic of Korea

Corresponding Author:
Chang Seok Bang, MD, PhD
Department of Internal Medicine
Hallym University College of Medicine
Sakju-ro 77
Chuncheon, 24253
Republic of Korea
Phone: 82 332405821
Fax: 82 332418064
Email: csbang@hallym.ac.kr

Abstract

Background: In a previous study, we examined the use of deep learning models to classify the invasion depth (mucosa-confined versus submucosa-invaded) of gastric neoplasms using endoscopic images. The external test accuracy reached 77.3%. However, model establishment is labor intense, requiring high performance. Automated deep learning (AutoDL) models, which enable fast searching of optimal neural architectures and hyperparameters without complex coding, have been developed.

Objective: The objective of this study was to establish AutoDL models to classify the invasion depth of gastric neoplasms. Additionally, endoscopist–artificial intelligence interactions were explored.

Methods: The same 2899 endoscopic images that were employed to establish the previous model were used. A prospective multicenter validation using 206 and 1597 novel images was conducted. The primary outcome was external test accuracy. Neuro-T, Create ML Image Classifier, and AutoML Vision were used in establishing the models. Three doctors with different levels of endoscopy expertise were asked to classify the invasion depth of gastric neoplasms for each image without AutoDL support, with faulty AutoDL support, and with best performance AutoDL support in sequence.

Results: The Neuro-T–based model reached 89.3% (95% CI 85.1%-93.5%) external test accuracy. For the model establishment time, Create ML Image Classifier showed the fastest time of 13 minutes while reaching 82.0% (95% CI 76.8%-87.2%) external test accuracy. While the expert endoscopist's decisions were not influenced by AutoDL, the faulty AutoDL misled the endoscopy trainee and the general physician. However, this was corrected by the support of the best performance AutoDL model. The trainee gained the most benefit from the AutoDL support.

Conclusions: AutoDL is deemed useful for the on-site establishment of customized deep learning models. An inexperienced endoscopist with at least a certain level of expertise can benefit from AutoDL support.

(J Med Internet Res 2021;23(4):e25167) doi:10.2196/25167

KEYWORDS

convolutional neural network; deep learning; automated deep learning; endoscopy; gastric neoplasms; neural network; deep learning model; artificial intelligence

Introduction

Artificial intelligence (AI) using deep learning (DL), which mimics the intellectual function of humans, has been increasingly adopted in clinical medicine, especially for cognitive function in computer vision [1-3], including automated image recognition, classification, and segmentation tasks [4-6]. Application of AI to endoscopic examination is noninvasive and can further help in detecting hidden or hard-to-detect lesions in real time. Moreover, automated determination of the optimum classification—providing delineation of the lesions—may be helpful to endoscopists, especially for inexperienced physicians. Optimizing classification facilitates the appropriate selection of high-risk patients who need additional workup or treatment [4,7]. Current established AI models are in the research-based format, which tend to have limited value in real-world clinical practice. However, AI models can potentially be used as add-on testing as a secondary assistant observer for endoscopists.

The accurate prediction of invasion depth for gastric neoplasms is an essential skill of endoscopists [8,9]. Gastric neoplasms confined to the mucosa or superficial submucosa are potential candidates for endoscopic resection [9]. Thus, precisely predicting the invasion depth is essential for determining the therapeutic strategy. Prediction of the invasion depth is based on the gross morphology of the lesions, and there are no standard criteria for classifying invasion depth. Therefore, current practice is limited by the inevitable interobserver variability and inaccurate determination of the invasion depth in gastric neoplasms [9].

The authors previously established DL models for classifying the invasion depth (mucosa-confined versus submucosa-invaded) of gastric neoplasms from endoscopic images using transfer learning of pretrained convolutional neural networks (CNNs) based on the PyTorch platform [10]. The external test accuracy was able to reach 77.3% [9]. However, the establishment of a DL model requires substantial time, and high performance is needed before applying these models to real-world clinical practice.

Automated deep learning (AutoDL) techniques, which enable fast searching of optimal neural architectures and hyperparameters without complex coding, have been widely developed. This “off-the-shelf” software or platform can be used without professional AI expertise and can easily be applied to clinical practice with simple inference structures [11]. However, the performance of AI models established by data scientists in the traditional manner and AutoDL models established by health care researchers for the gastrointestinal endoscopy field have not been directly compared. Moreover, there are scarce data in terms of human-AI interactions. For example, the reaction of endoscopists (ie, approval, indolence, or disregard) to diagnoses made using an AI model remains unknown [12]. This study aimed to establish AutoDL models classifying invasion depth of gastric neoplasms using endoscopic images and compare the diagnostic performance of the AutoDL models with previous CNN models established in the traditional way. Additionally, endoscopist-AI interactions using the newly established model were further examined.

Methods

Construction of the Data Set

This study extends the previous research on this topic [9] by constructing (Figure 1) and evaluating (Figure 2) experimental DL models with AutoDL tools. In order to compare the diagnostic performance of AutoDL–based models to the previous CNN models, the same input images (2899 white-light imaging endoscopic images) that were used to establish the previous model were used again. The detailed data collection process was described previously [9]. Briefly, between 2010 and 2017 in the Chuncheon Sacred Heart Hospital (Republic of Korea), we enrolled consecutive patients with any type of gastric neoplasms discovered during upper gastrointestinal endoscopy and histologically confirmed. Endoscopic images were collected from the in-hospital database in JPEG format, with a minimum resolution of 640×480 pixels [9]. The same previously used external test data set (206 white-light imaging endoscopic images) was also used in classifying the performance of the AutoDL models. This external test data set was constructed by collecting images from consecutive patients who underwent upper gastrointestinal endoscopy between 2019 and 2020, and all the images were mutually exclusive from those of the training and internal validation data set (Table 1) [9].

To guarantee the generalizability of the performance of the newly developed AutoDL model, an additional performance verification (prospective validation) test with another external test data set was conducted. This second external test data set, including 1597 images, was collected from consecutive patients who underwent upper gastrointestinal endoscopy at the Hallym University Sacred Heart Hospital from 2018 to 2020 (Table 1).

Figure 1. Schematic flow for the establishment of automated deep learning models in data construction.

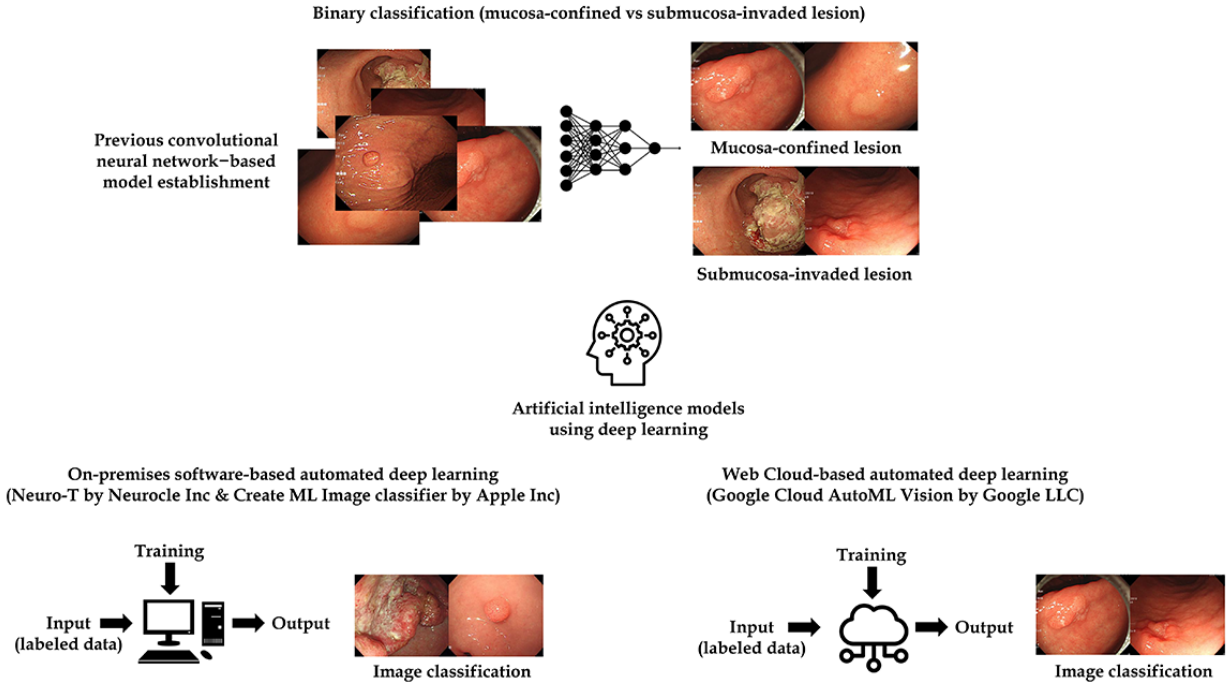


Figure 2. Schematic flow for the establishment of automated deep learning models in performance evaluation.

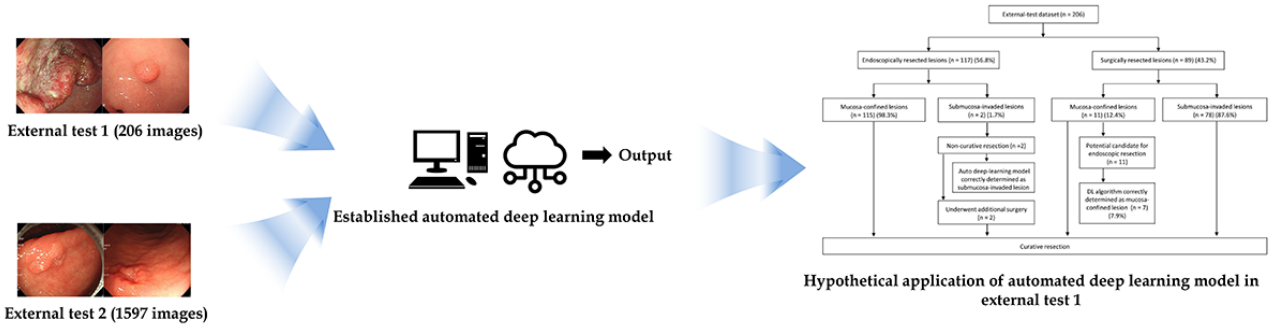


Table 1. Summary of images in each data set.

	Number of images		
Invasion depth of gastric neoplasms	Training and internal validation data set	External test data set 1	External test data set 2
Overall	2899	206	1597
Mucosa-confined lesions	1900	126	1344
Low-grade dysplasia	727	68	734
High-grade dysplasia	421	21	110
Early gastric cancer	752	37	500
Submucosa-invaded lesions	999	80	253
Early gastric cancer	282	23	155
Advanced gastric cancer	717	57	98

AutoDL Tools Used in the Study

AutoDL tools including Neuro-T (version 2.0.2; Neurocle Inc), Create ML Image Classifier (Apple Inc), and Google Cloud AutoML Vision (Google LLC) were used in this study.

Neuro-T has been defined as an AutoDL software that can establish DL algorithms on its own for image recognition and classification based on a graphical user interface (GUI). The software's algorithm analyzes the features of the data set and self-discovers optimal hyperparameters, thus making it easy for

non-AI experts to build the best models. Neuro-T also offers a platform to establish anomaly detection models (supervised anomaly detection based on the clustering algorithm with deep neural networks). Anomaly detection is the identification of observations that raise suspicions by differing significantly from the majority of the training data. The neural network clustering algorithm clusters each training sample and makes its own cluster decision boundary for classifying the test sample into normal class or abnormal class. Meanwhile, Create ML is defined as a framework used to establish customized DL models on the Mac operating system (Apple Inc); Image Classifier can be accessed by GUI or Swift language code. DL models can be established using image data sets through the self-learning process of specific features. Google Cloud AutoML Vision is a web-based service to build customized DL models with automatic neural architecture searching and feature extraction. These 3 AutoDL tools were used based on the manner of the GUI (ie, no coding tool required) to build the DL models.

Preprocessing of Images

The authors used data augmentation methods—such as rotation, and horizontal or vertical flipping of included images—and image normalization with linear transformation in terms of 3 RGB channels in order to build the previous DL model [9]. However, AutoDL tools are determined to have data preprocessing functions. Neuro-T has an automated image normalization process with a resizing function for input images. All of the included images were resized with a resolution of 512×480 pixels while building the Neuro-T–based models. Create ML Image Classifier offers GUI–based data augmentation options. These include 6 image data augmentation methods, such as “add noise,” “blur,” “crop,” “expose,” “flip,” or “rotate” functions. In order to identify the best models, we conducted multiple experiments (with or without data augmentation and single or combination data augmentation options) in Create ML. In terms of AutoML Vision, no GUI-based data augmentation option was determined. Developers can add image augmentation codes using a Python application programming interface. However, considering that the aim of this study was to develop AutoDL models without complex coding or AI expertise, we only selected the GUI-based function without data augmentation while building the AutoML Vision–based models.

Training of AutoDL Models

The 2899 input images were uploaded to each AutoDL tool. Neuro-T and Create ML were considered on-premise software; however, AutoML Vision is a cloud-based service. The input images were uploaded to Neuro-T and Create ML, and a bucket in Google Cloud Storage system was used for data uploading in the AutoML Vision. After selecting data preprocessing options (as described above, including resizing/normalization in Neuro-T and image augmentation in Create ML), AutoDL models were trained in each specified way of self-learning. Images were then randomly split into training and internal validation sets. The Neuro-T variable options—such as 9:1, 8:2, or 7:3—were set as per the user’s preference. Multiple experiments were further conducted to determine the model with the best performance with variable splitting ratios. However, Create ML Image Classifier automatically sets an internal validation set using approximately 5.1% of the images; thus, 149 images were allocated in the internal validation data set. In AutoML Vision, the ratio of training, internal validation, and internal test sets was 8:1:1. For the training of the anomaly detection model in Neuro-T, only images with mucosa-confined lesions could be used. Therefore, 1714 mucosa-confined images were used for training, and 186 mucosa-confined images and 999 submucosa-invaded images were used for the internal validation set. The number of iterations in training can be set for Create ML. Experiments for the different iteration numbers were conducted to prevent overfitting (ie, model learns too much about training images, and predictions are not well generalized to new images) [3].

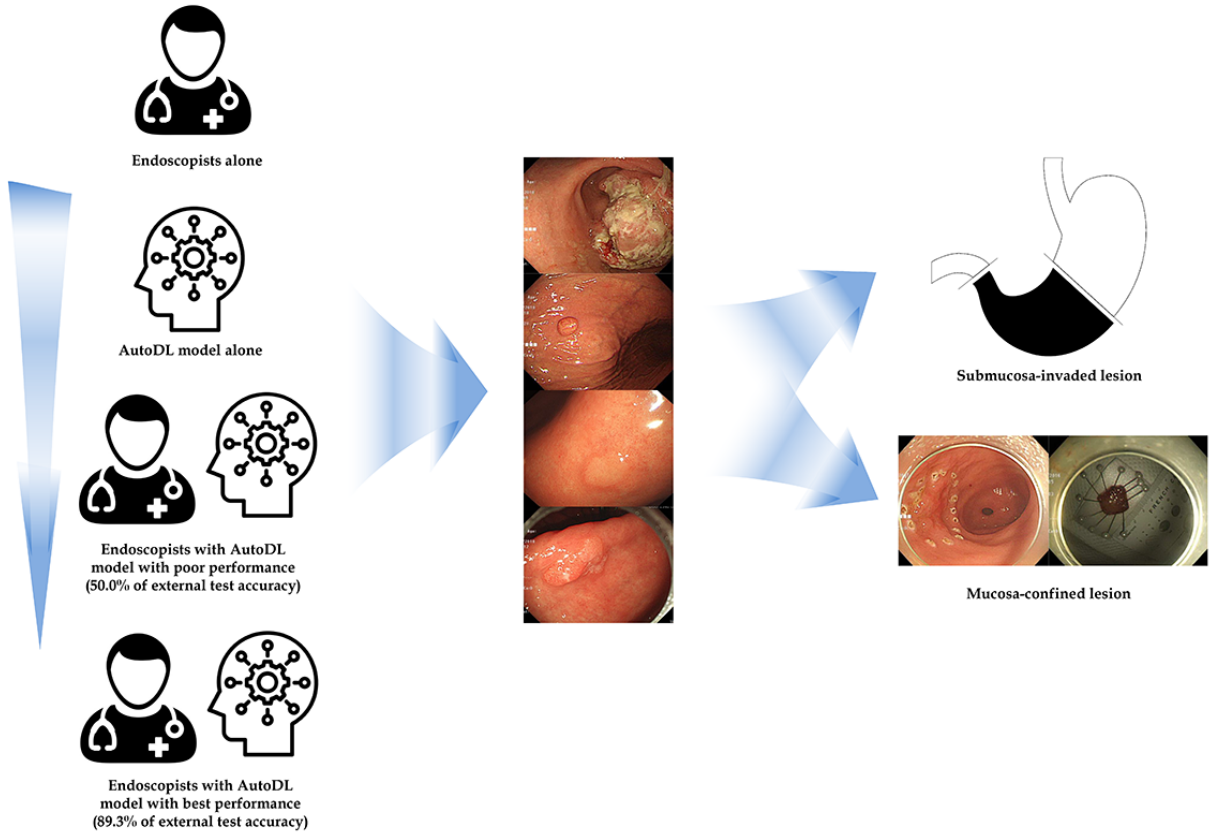
The hardware system used for the training included NVIDIA GeForce RTX 2080 Ti graphics processing units (GPUs), dual Intel Xeon central processing units (CPUs), and 256 GB RAM for the Neuro-T–based AutoDL models. Create ML–based models were established on both the MacBook Pro laptop (2019 version, AMD Radeon Pro 5500M GPU, Intel Core i9 CPU, and 32 GB RAM) and the Mac Pro workstation (2019 version, AMD Radeon Pro Vega II GPU, Intel Xeon W CPU, and 192 GB RAM) environments in order to compare the training time according to the hardware system.

Endoscopist-AI Interactions

Three doctors with different levels of endoscopy expertise were asked to classify the invasion depth of gastric neoplasms for each endoscopic image without AutoDL support, with faulty AutoDL support, and with the support of the best performance AutoDL in sequence (Figure 3). A board-certified endoscopist with more than 7 years of endoscopic submucosal dissection experience, an endoscopy trainee, and a general physician with minimal endoscopy expertise participated in the blind test. Endoscopic images (206 images from the external test data set) without information about invasion depth were used. The first test was conducted without AI support. To find the interaction between endoscopists and poor-quality AI, blind testing was conducted while providing the poor-quality model’s (faulty AI) answers with only 50.0% external test accuracy. Another round of blind testing was performed while providing the answers from the best performance AutoDL model (89.3% external test accuracy). The confidence of the raters in their answers was recorded for each test.

A detailed description of the primary outcome and statistics are described in Multimedia Appendix 1.

Figure 3. Schematic flow for the establishment of automated deep learning (AutoDL) models in endoscopist–artificial intelligence interaction test.



Results

Characteristics of the Included Images

The detailed characteristics of the images included in this study were described in the previous report [9]. In brief, 65.5% of the images were mucosa-confined lesions and 34.5% were determined to be submucosa-invaded lesions (2899 images for the training and internal validation set). For the first external test (206 images), 61.2% and 38.8% of the images were determined to be mucosa-confined lesions and submucosa-invaded lesions, respectively. For the second external test (1597 images), 84.2% and 15.8% were identified to be mucosa-confined lesions and submucosa-invaded lesions, respectively. Table 1 shows a summary of the images used in this study.

Diagnostic Performance of AutoDL Models for the First External Test

The Neuro-T–based classification model reached 89.3% (95% CI 85.1%-93.5%) accuracy, 89.1% (95% CI 84.8%-93.4%) average precision, 88.4% (95% CI 84.0%-92.8%) average recall, and 88.7% (95% CI 84.4%-93.0%) F1 score in the external test. The total training time was approximately 13 hours. The external test accuracy of the Neuro-T–based model was significantly higher than that of the previous CNN model with the best performance (ie, 77.3%, 95% CI 75.4%-79.3%; *P*=.005). The confusion matrix for the Neuro-T–based model in the external test is illustrated in Multimedia Appendix 2A. The detailed information of the established model is as follows: batch size, 80; epochs, 84; number of layers, 53; optimizer, Adam; and input height and width, 480×512 pixels. All images were resized with interlinear interpolation, and the initial learning rate was 0.002.

The anomaly detection model established by Neuro-T was able to reach 49.5% (95% CI 42.7%-56.3%) accuracy, 51.0% (95% CI 44.2%-57.8%) average precision, 51.0% (95% CI 44.2%-57.8%) average recall, and 51.0% (95% CI 44.2%-57.8%) F1 score in the external test. The training time was approximately 20 minutes. The confusion matrix for the anomaly detection model is illustrated in Multimedia Appendix 2B.

For the Create ML Image Classifier, data augmentation options combining “blur” and “rotate” provided the best performance after 25 iterations. The external test accuracy reached 83.5% (95% CI 78.4%-88.6%). The training time was approximately 76 minutes in the laptop environment, which was determined to be not significantly different from the training time in the workstation environment. Furthermore, the external test accuracy of the Create ML–based model was not statistically different from that of the previous model (*P*=.26). The confusion matrix for the Create ML–based AutoDL model with the best performance is presented in Multimedia Appendix 2C.

The fastest model establishment with high performance was achieved by data augmentation options combining “add noise” and “blur” after 25 iterations. The external test accuracy reached 82.0% (95% CI 76.8%-87.2%). The training time was determined to be only about 13 minutes in the laptop

environment, which was not different from the training time in the workstation environment. Also, the external test accuracy of the Create ML–based model was not statistically different from that of the previously established CNN model (*P*=.45). The confusion matrix for the Create ML–based AutoDL model with the fastest building time and high performance is illustrated in Multimedia Appendix 2D.

For the Google Cloud AutoML Vision model, external test accuracy reached 83.0% (95% CI 77.9%-88.1%). The training time was only 25 minutes (web cloud–based environment). The external test accuracy for the AutoML Vision–based model was not statistically different from that of the previous CNN model (*P*=.31). The confusion matrix for the AutoML Vision–based model is illustrated in Multimedia Appendix 2E.

The summary statistics of external test accuracy with internal validation accuracy are shown in Table 2.

Table 2. Summary of external test accuracy with internal validation accuracy for each automated deep learning (AutoDL) model.

AutoDL model	Accuracy, % (95% CI)	Precision, % (95% CI)	Recall, % (95% CI)	F1 score, % (95% CI)	Training time (minutes)
Neuro-T–based model					826
Internal validation performance (n=290)	92.4 (89.3-95.5)	M[a]: 92.0 (88.1-95.9); SM[b]: 93.3 (88.4-98.2)	M: 96.8 (94.3-99.3); SM: 84 (76.8-91.2)	M: 94.4 (91.1-97.7); SM: 88.4 (82.1-94.7)	
External test performance (n=290)	89.3 (85.1-93.5)	M: 89.9 (84.6-95.2); SM: 88.3 (81.3-95.3)	M: 92.8 (88.3-97.3); SM: 84.0 (76.0-92.0)	M: 91.3 (86.4-96.2); SM: 86.1 (78.6-93.6)	
Neuro-T–based anomaly detection model					20
Internal validation performance (n=1185)	80.2 (77.9-82.5)	M: 33.3 (26.6-40.0); SM: 86.2 (84.1-88.3)	M: 23.7 (17.7-29.7); SM: 91.0 (89.2-92.8)	M: 27.7 (21.3-34.1); SM: 88.6 (86.6-90.6)	
External test performance (n=206)	49.5 (42.7-56.3)	M: 61.8 (53.3-70.3); SM: 40.2 (29.5-50.9)	M: 44.0 (35.3-52.7); SM: 58.0 (47.3-68.7)	M: 51.4 (42.6-60.2); SM: 47.5 (36.6-58.4)	
Create ML–based model 1					76
Internal validation performance (n=149)	81.9 (75.7-88.1)	M: 80.6 (72.2-89.0); SM: 83.9 (75.0-92.8)	M: 89.3 (82.7-95.9); SM: 72.3 (61.4-83.2)	M: 84.7 (77.0-92.4); SM: 77.7 (67.6-87.8)	
External test performance (n=206)	83.5 (78.4-88.6)	M: 82.3 (75.6-89.0); SM: 86.2 (78.7-93.7)	M: 92.8 (88.3-97.3); SM: 69.1 (59.0-79.2)	M: 87.2 (81.3-93.1); SM: 76.7 (67.5-85.9)	
Create ML–based model 2					13
Internal validation performance (n=149)	81.9 (75.7-88.1)	M: 82.0 (73.8-90.2); SM: 81.7 (72.3-91.1)	M: 86.9 (79.7-94.1); SM: 75.4 (64.9-85.9)	M: 84.4 (76.6-92.2); SM: 78.4 (68.4-88.4)	
External test performance (n=206)	82.0 (76.8-87.2)	M: 79.7 (72.6-86.8); SM: 87.9 (80.8-95.0)	M: 94.4 (90.4-98.4); SM: 63.0 (52.5-73.5)	M: 86.1 (80.0-92.2); SM: 73.4 (63.8-83.0)	
AutoML Vision–based model					25
Internal validation performance (n=295)	84.7 (80.6-88.8)	M: 87.0 (82.3-91.7); SM: 80 (72.2-87.8)	M: 90.2 (86.0-94.4); SM: 74.5 (66.0-83.0)	M: 88.6 (84.0-93.1); SM: 77.2 (69.1-85.3)	
External test performance (n=206)	83.0 (77.9-88.1)	M: 80.0 (73.0-87.0); SM: 91.1 (84.9-97.3)	M: 96.0 (92.6-99.4); SM: 63.0 (52.9-73.1)	M: 87.3 (81.5-93.1); SM: 74.5 (65.0-84.0)	

[a]M: mucosa-confined lesions.

[b]SM: submucosa-invaded lesions.

Additional Performance Verification to Gain Generalization Potential in the Second External Test

For the 1597 images in the second external test, the AutoDL model was determined to perform the best (Neuro-T–based), reaching 88.6% (95% CI 87.0%-90.2%) accuracy, 83.7% (95% CI 81.9%-85.5%) average precision, 68.5% (95% CI 66.2%-70.8%) average recall, and 75.4% (95% CI 73.3%-77.5%) F1 score.

Hypothetical Application of AutoDL Model

Hypothetical clinical application of the established AutoDL model with the best performance was conducted using the first external test data set, assuming that the AutoDL was applied in order to determine the treatment (ie, endoscopic resection or surgical resection), based on the invasion depth of the lesion. Among the lesions with endoscopic resection (n=117), 2 lesions (1.7%) were found to invade the submucosa and were resected with additional surgery after endoscopic resection. The AutoDL model correctly determined that these were submucosa-invaded lesions. Thus, the model has the potential to prevent unnecessary endoscopic procedures. Among the lesions with surgical resection (n=89), 11 lesions (12.4%) were identified to be mucosa-confined, having the potential for endoscopic resection. The AutoDL model correctly determined the mucosa-confined lesions in 7 of the 11 patients (7.9%). Thus, the model has the potential to prevent unnecessary surgeries (Multimedia Appendix 3).

Endoscopist-AI Interactions

Figure 4 and Table 3 show the external test accuracy of the 3 raters. The expert endoscopist's decision was determined to have not been influenced by the support of the AutoDL in the consecutive tests. The faulty AutoDL model misled the decisions of the endoscopy trainee and general physician, but the difference was not statistically significant. However, support from the best performance AutoDL model corrected this misdirection. The endoscopy trainee benefited the most from the support of the AutoDL model (P=.002). In the analysis of whether the raters were sure of their answers, confident answers showed a similar pattern to that of the overall results. In the "unconfident answers" subgroup, the expert endoscopist's decisions were not influenced, and the trainee gained a statistically significant benefit, even with the support of the poor performance AutoDL model (P<.05). However, the general physician did not benefit from the support of the AutoDL models.

Figure 4. Endoscopist–artificial intelligence interactions. AutoDL: automated deep learning.

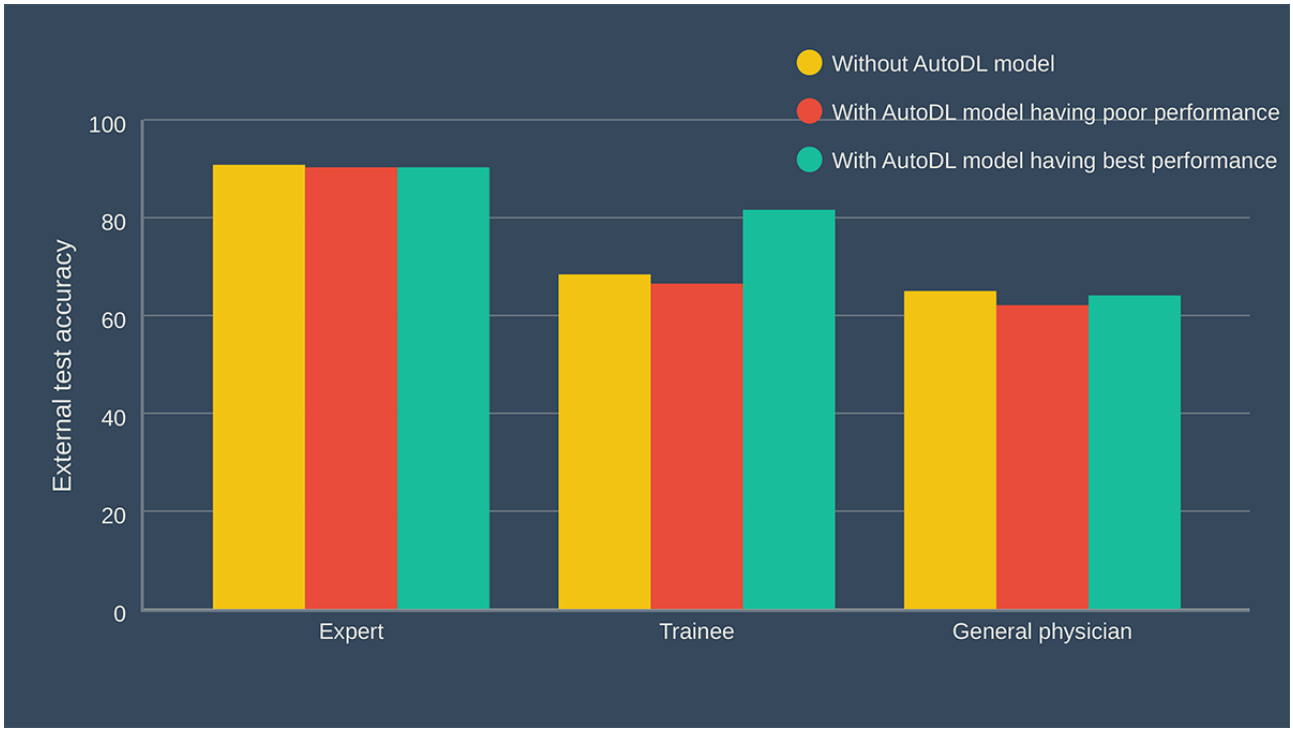


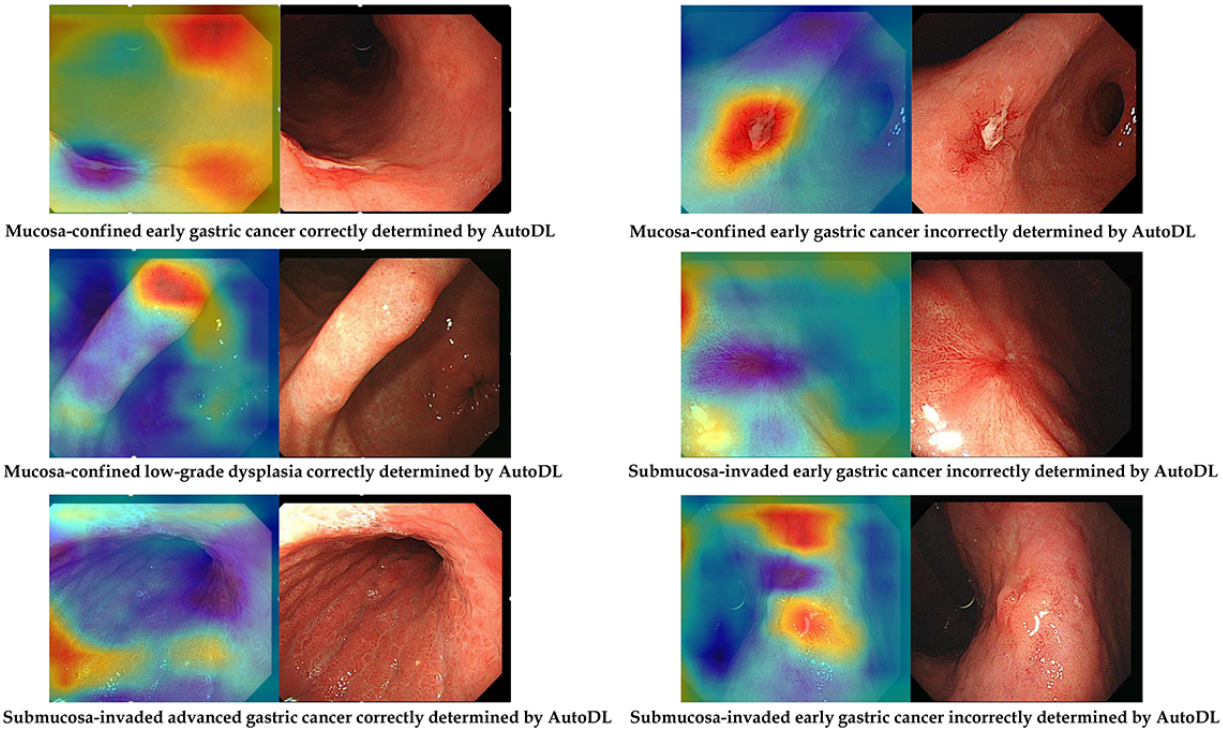
Table 3. Summary of external test accuracy in endoscopist–artificial intelligence (AI) interactions.

Rater	External test accuracy, % (95% CI)			*P* values	
	First test: endoscopist alone	Second test: endoscopist with faulty AI	Third test: endoscopist with best performance AI	First test vs second test	First test vs third test
Expert endoscopist					
All answers	90.8 (86.8-94.8), (187/206)	90.3 (86.3-94.3), (186/206)	90.3 (86.3-94.3), (186/206)	.87	.87
Confident answer	93.8 (91.7-95.9), (181/193)	90.6 (86.6-94.6), (182/201)	90.6 (86.6-94.6), (184/203)	.23	.24
Unconfident answer	46.2 (19.1-73.3), (6/13)	80.0 (44.9-99.9), (4/5)	66.7 (13.4-99.9), (2/3)	.20	.52
Endoscopy trainee					
All answers	68.4 (62.1-74.7), (141/206)	66.5 (60.1-72.9), (137/206)	81.6 (76.3-86.9), (168/206)	.67	.002
Confident answer	80.6 (74.4-86.8), (125/155)	71.1 (63.2-79.0), (91/128)	92.2 (87.8-96.6), (130/141)	.06	.004
Unconfident answer	31.4 (18.7-44.1), (16/51)	59.0 (48.1-69.9), (46/78)	58.5 (46.5-70.5), (38/65)	.002	.004
General physician					
All answers	65.0 (58.5-71.5), (134/206)	62.1 (55.5-68.7), (128/206)	64.1 (57.5-70.7), (132/206)	.38	.84
Confident answer	77.4 (70.0-84.8), (96/124)	71.6 (64.5-78.7), (111/155)	74.2 (67.5-80.9), (121/163)	.27	.53
Unconfident answer	46.3 (35.5-57.1), (38/82)	33.3 (20.4-46.2), (17/51)	25.6 (12.6-38.6), (11/43)	.14	.02

Attention Map for Explainability

A class activation map to localize the discriminative regions used by the AutoDL model to determine a specific class in the image is presented for the Neuro-T model. Figure 5 shows the correctly and incorrectly determined samples in the external test using the Neuro-T–based AutoDL model.

Figure 5. Representative samples of the attention map (Neuro-T–based model). AutoDL: automated deep learning.



Discussion

AutoDL models were established using only GUI-based systems, which surpassed or had a similar accuracy to that of the previous model for the determination of the invasion depth of gastric neoplasms using endoscopic images. The best performance model showed an external test accuracy of 89.3%, which is deemed much higher than the previous CNN model built in the traditional manner. Furthermore, the authors performed an external test to determine the generalizability, and the external test was found to exhibit robust performance. As far as the authors know, this is the highest performance achieved by an AutoDL model in the context of gastric neoplasms. In previous studies, the internal validation accuracy was 64.7% to 94.5%

for the discrimination of invasion depth in gastric cancers [13-16]. However, these models can only be applied after a definite diagnosis of gastric cancer, which has limited value due to various types of gastric neoplasms in real-world clinical practice [9]. Moreover, no single study evaluated the external test accuracy; thus, the previous tests lacked generalizability.

Creating a classification model using medical images through transfer learning based on high performance CNNs is a representative establishment method for AI models. However, health care researchers often lack the AI expertise to directly apply the models to clinical data to create the AI model [17]. Data scientists and endoscopists can collaborate to create AI models. However, this process usually takes a substantial amount of time and does not immediately reflect the unmet needs of clinical practice.

AutoDL technology makes it possible for nonexperts to create high-quality DL models even without AI expertise. These tools are easy-to-use and only require data uploading to the on-premise software or web cloud platform and simple labeling for the correct classification [18]. After self-learning and model fitting, AutoDL tools provide models ready for direct inference or deployment in real-world practice. Establishing a model requires considerably less time than traditional platform-based model generation. In this study, the fastest establishment time was only 13 minutes and provided an external test accuracy (82.0%) that was similar to that of the previous CNN models based on the laptop environment. In addition, Apple's Create ML and Google Cloud AutoML Vision are publicly available platforms that anyone can download from the App Store or access from the web cloud platform.

Endoscopists produce an enormous amount of image data in their daily practice; often, they are forced to make instantaneous medical judgments even during endoscopic procedures. However, the burnout phenomenon of endoscopists is a serious concern that needs to be addressed, as it affects concentration and, possibly, medical judgment [19]. A previous study on human-AI interactions suggested that applying a high-quality DL model to clinical decision making improves diagnostic performance compared with either DL models or physicians alone; thus, it was deemed particularly beneficial for less-experienced doctors [7].

Although AI is potentially efficacious in clinical practice, data regarding endoscopist-AI interactions remain to be scarce. Based on the findings in this study, inexperienced endoscopists with at least a certain level of expertise can benefit from AutoDL support. The endoscopy trainee—but not the general physician or expert endoscopist—benefited from AI support in this study. Most of the answers by the expert endoscopist were rated as confident and were not influenced by the AI support (ie, possible disagreement with or disregard for the AI answers). The proportion of unconfident answers in the first test was the highest for the general physician, and this proportion was markedly decreased with the support of the faulty AI in the second test. In the third test, the general physician appeared to be confused by changes in the AI answers (from the faulty AI answers to the best performance AI answers). The sequential support by the faulty AI and then the best performance AI confused the general physicians because of their minimal endoscopy expertise. This highlights not only the importance of robust answers provided by the AI but also the importance of the baseline level of experience of AI users. Therefore, the conclusion from the previous study [6] that inexperienced doctors would benefit the most from AI support was not reproduced in this study. Rather, endoscopists having at least a certain level of expertise benefited from AI support.

Although this study established a high performance AutoDL model and rigorously validated the model's performance, this analysis has several inevitable limitations originating from potential bias in data sets. First, the training images were retrieved from a single institution, which might infer a selection or spectrum bias. Because of the unique characteristics of patients in each institution, medical AI models developed from a single institution usually have limitations for widespread implementation, indicating the importance of the external test [5]. To compensate for this pitfall, we performed two rounds of prospective validations and included images from another institution. Second, the efficacy of inference for each established model was not measured in clinical practice. Each established AutoDL model employs a specified inference method, such as website-based inference or edge computing–based application inference. The efficacy of inference includes inference speed, accuracy, easy applicability, simple control flow, energy efficiency, and model size. Because inference is a different field from that of this study, another comparative study will have to be conducted for the best inference AutoDL model. Recently developed machine learning or DL models in gastrointestinal endoscopy are focused on improving the effectiveness rather than the interpretability or efficiency. The most accurate model in our study also showed longer establishment time than the other models. There have been efficiency-effectiveness trade-offs in the field of DL models. Although real-world clinical application or inference time was not the primary outcome in this study, efficiency-effectiveness trade-offs should be considered in the context of real-world settings. Third, a relatively small number of raters were included in the endoscopist-AI interaction test. We only included one representative physician in each endoscopy expertise level. Large-scale studies evaluating more discrete expertise levels would elucidate the future perspectives for the implementation of AutoDL models in the clinical setting.

AutoDL has been considered as a useful tool for the on-site establishment of customized DL models, and anyone can create an AI model with the help of AutoDL. An inexperienced endoscopist with at least a certain level of expertise can benefit from AutoDL support.

Acknowledgments

Funding for this research was provided by the Bio and Medical Technology Development Program of the National Research Foundation (NRF) and by the Korean government, Ministry of Science and ICT (MSIT; grant number NRF2020R1F1A1071494).

All investigators have access to the final data set. Data and analysis codes are available upon request from the corresponding author by email.

Authors' Contributions

Conceptualization: CSB. Data curation: CSB, HL, HMJ, and SHH. Formal analysis: CSB. Funding acquisition: CSB. Investigation: CSB. Methodology: CSB. Project administration: CSB. Resources: CSB. Writing of original draft: CSB. Review and editing: CSB.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Primary outcome and statistics.
[DOCX File , 13 KB - jmir_v23i4e25167_app1.docx]

Multimedia Appendix 2
Confusion matrices for the automated deep learning models in the external test. (A) Neuro-T–based model. (B) Neuro-T–based anomaly detection model. (C) Create ML–based model with the best performance. (D) Create ML–based model with the fastest building time and high performance. (E) AutoML Vision–based model.
[DOCX File , 529 KB - jmir_v23i4e25167_app2.docx]

Multimedia Appendix 3
Hypothetical clinical application of the automated deep learning model for the determination of treatments based on the invasion depth of the lesion in an external test.
[DOCX File , 311 KB - jmir_v23i4e25167_app3.docx]

References

1. Choi J, Shin K, Jung J, Bae H, Kim DH, Byeon J, et al. Convolutional Neural Network Technology in Endoscopic Imaging: Artificial Intelligence for Endoscopy. Clin Endosc 2020 Mar;53(2):117-126 [FREE Full text] [doi: 10.5946/ce.2020.054] [Medline: 32252504]
2. Cho B, Bang CS. Artificial Intelligence for the Determination of a Management Strategy for Diminutive Colorectal Polyps: Hype, Hope, or Help. Am J Gastroenterol 2020 Jan;115(1):70-72. [doi: 10.14309/ajg.0000000000000476] [Medline: 31770118]
3. Yang YJ, Bang CS. Application of artificial intelligence in gastroenterology. World J Gastroenterol 2019 Apr 14;25(14):1666-1683 [FREE Full text] [doi: 10.3748/wjg.v25.i14.1666] [Medline: 31011253]
4. Bang CS. [Deep Learning in Upper Gastrointestinal Disorders: Status and Future Perspectives]. Korean J Gastroenterol 2020 Mar 25;75(3):120-131 [FREE Full text] [doi: 10.4166/kjg.2020.75.3.120] [Medline: 32209800]
5. Abadir AP, Ali MF, Karnes W, Samarasena JB. Artificial Intelligence in Gastrointestinal Endoscopy. Clin Endosc 2020 Mar;53(2):132-141 [FREE Full text] [doi: 10.5946/ce.2020.038] [Medline: 32252506]
6. Bang CS, Lee JJ, Baik GH. Artificial Intelligence for the Prediction of Helicobacter Pylori Infection in Endoscopic Images: Systematic Review and Meta-Analysis Of Diagnostic Test Accuracy. J Med Internet Res 2020 Sep 16;22(9):e21983 [FREE Full text] [doi: 10.2196/21983] [Medline: 32936088]
7. Tschandl P, Rinner C, Apalla Z, Argenziano G, Codella N, Halpern A, et al. Human-computer collaboration for skin cancer recognition. Nat Med 2020 Aug;26(8):1229-1234. [doi: 10.1038/s41591-020-0942-0] [Medline: 32572267]
8. Cho B, Bang CS, Park SW, Yang YJ, Seo SI, Lim H, et al. Automated classification of gastric neoplasms in endoscopic images using a convolutional neural network. Endoscopy 2019 Dec;51(12):1121-1129. [doi: 10.1055/a-0981-6133] [Medline: 31443108]
9. Cho B, Bang CS, Lee JJ, Seo CW, Kim JH. Prediction of Submucosal Invasion for Gastric Neoplasms in Endoscopic Images Using Deep-Learning. J Clin Med 2020 Jun 15;9(6):1858 [FREE Full text] [doi: 10.3390/jcm9061858] [Medline: 32549190]
10. Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, et al. PyTorch: An Imperative Style, High-Performance Deep Learning Library. 2019 Presented at: 2019 Conference on Neural Information Processing Systems; December 8 to 14, 2019; Vancouver, BC, Canada.

11. Waring J, Lindvall C, Umeton R. Automated machine learning: Review of the state-of-the-art and opportunities for healthcare. Artif Intell Med 2020 Apr;104:101822 [FREE Full text] [doi: 10.1016/j.artmed.2020.101822] [Medline: 32499001]
12. Hoogenboom SA, Bagci U, Wallace MB. Artificial intelligence in gastroenterology. The current state of play and the potential. How will it affect our practice and when? Tech Innov Gastrointest Endosc 2020 Apr;22(2):42-47. [doi: 10.1016/j.tgie.2019.150634]
13. Kubota K, Kuroda J, Yoshida M, Ohta K, Kitajima M. Medical image analysis: computer-aided diagnosis of gastric cancer invasion on endoscopic images. Surg Endosc 2012 May;26(5):1485-1489. [doi: 10.1007/s00464-011-2036-z] [Medline: 22083334]
14. Yoon HJ, Kim S, Kim J, Keum J, Oh S, Jo J, et al. A Lesion-Based Convolutional Neural Network Improves Endoscopic Detection and Depth Prediction of Early Gastric Cancer. J Clin Med 2019 Aug 26;8(9):1310 [FREE Full text] [doi: 10.3390/jcm8091310] [Medline: 31454949]
15. Zhu Y, Wang Q, Xu M, Zhang Z, Cheng J, Zhong Y, et al. Application of convolutional neural network in the diagnosis of the invasion depth of gastric cancer based on conventional endoscopy. Gastrointest Endosc 2019 Apr;89(4):806-815.e1. [doi: 10.1016/j.gie.2018.11.011] [Medline: 30452913]
16. Nagao S, Tsuji Y, Sakaguchi Y, Takahashi Y, Minatsuki C, Niimi K, et al. Highly accurate artificial intelligence systems to predict the invasion depth of gastric cancer: efficacy of conventional white-light imaging, nonmagnifying narrow-band imaging, and indigo-carmine dye contrast imaging. Gastrointest Endosc 2020 Oct;92(4):866-873.e1. [doi: 10.1016/j.gie.2020.06.047] [Medline: 32592776]
17. Luo G, Stone BL, Johnson MD, Tarczy-Hornoch P, Wilcox AB, Mooney SD, et al. Automating Construction of Machine Learning Models With Clinical Big Data: Proposal Rationale and Methods. JMIR Res Protoc 2017 Aug 29;6(8):e175 [FREE Full text] [doi: 10.2196/resprot.7757] [Medline: 28851678]
18. Kim IK, Lee K, Park JH, Baek J, Lee WK. Classification of pachychoroid disease on ultrawide-field indocyanine green angiography using auto-machine learning platform. Br J Ophthalmol 2020 Jul 03:Online ahead of print. [doi: 10.1136/bjophthalmol-2020-316108] [Medline: 32620684]
19. Jang ES, Park SM, Park YS, Lee JC, Kim N. Work-Life Conflict and Its Health Effects on Korean Gastroenterologists According to Age and Sex. Dig Dis Sci 2020 Jan;65(1):86-95. [doi: 10.1007/s10620-019-05842-w] [Medline: 31549335]

Abbreviations

AI: artificial intelligence
AutoDL: automated deep learning
CNN: convolutional neural network
CPU: central processing unit
DL: deep learning
GPU: graphics processing unit
GUI: graphical user interface

Edited by R Kukafka; submitted 20.10.20; peer-reviewed by S Pang, J Kim; comments to author 26.11.20; revised version received 09.12.20; accepted 16.03.21; published 15.04.21.

Please cite as:
Bang CS, Lim H, Jeong HM, Hwang SH
Use of Endoscopic Images in the Prediction of Submucosal Invasion of Gastric Neoplasms: Automated Deep Learning Model Development and Usability Study
J Med Internet Res 2021;23(4):e25167
URL: https://www.jmir.org/2021/4/e25167
doi:10.2196/25167
PMID:33856356

Original Paper

Establishing Machine Learning Models to Predict Curative Resection in Early Gastric Cancer with Undifferentiated Histology: Development and Usability Study

Chang Seok Bang[1,2,3,4*], MD, PhD; Ji Yong Ahn[5], MD, PhD; Jie-Hyun Kim[6], MD, PhD; Young-Il Kim[7], MD, PhD; Il Ju Choi[7], MD, PhD; Woon Geon Shin[1,2*], MD, PhD

[1]Department of Internal Medicine, Hallym University College of Medicine, Chuncheon, Republic of Korea

[2]Institute for Liver and Digestive Diseases, Hallym University, Chuncheon, Republic of Korea

[3]Institute of New Frontier Research, Hallym University College of Medicine, Chuncheon, Republic of Korea

[4]Division of Big Data and Artificial Intelligence, Chuncheon Sacred Heart Hospital, Chuncheon, Republic of Korea

[5]Division of Gastroenterology, Department of Internal Medicine, University of Ulsan College of Medicine, Asan Medical Center, Seoul, Republic of Korea

[6]Division of Gastroenterology, Department of Internal Medicine, Gangnam Severance Hospital, Yonsei University College of Medicine, Seoul, Republic of Korea

[7]Center for Gastric Cancer, Research Institute and Hospital, National Cancer Center, Goyang, Republic of Korea

[*]these authors contributed equally

Corresponding Author:
Chang Seok Bang, MD, PhD
Department of Internal Medicine
Hallym University College of Medicine
Sakju-ro 77
Gangwon-do
Chuncheon, 24253
Republic of Korea
Phone: 82 33 240 5821
Email: csbang@hallym.ac.kr

Abstract

Background: Undifferentiated type of early gastric cancer (U-EGC) is included among the expanded indications of endoscopic submucosal dissection (ESD); however, the rate of curative resection remains unsatisfactory. Endoscopists predict the probability of curative resection by considering the size and shape of the lesion and whether ulcers are present or not. The location of the lesion, indicating the likely technical difficulty, is also considered.

Objective: The aim of this study was to establish machine learning (ML) models to better predict the possibility of curative resection in U-EGC prior to ESD.

Methods: A nationwide cohort of 2703 U-EGCs treated by ESD or surgery were adopted for the training and internal validation cohorts. Separately, an independent data set of the Korean ESD registry (n=275) and an Asan medical center data set (n=127) treated by ESD were chosen for external validation. Eighteen ML classifiers were selected to establish prediction models of curative resection with the following variables: age; sex; location, size, and shape of the lesion; and whether ulcers were present or not.

Results: Among the 18 models, the extreme gradient boosting classifier showed the best performance (internal validation accuracy 93.4%, 95% CI 90.4%-96.4%; precision 92.6%, 95% CI 89.5%-95.7%; recall 99.0%, 95% CI 97.8%-99.9%; and F1 score 95.7%, 95% CI 93.3%-98.1%). Attempts at external validation showed substantial accuracy (first external validation 81.5%, 95% CI 76.9%-86.1% and second external validation 89.8%, 95% CI 84.5%-95.1%). Lesion size was the most important feature in each explainable artificial intelligence analysis.

Conclusions: We established an ML model capable of accurately predicting the curative resection of U-EGC before ESD by considering the morphological and ecological characteristics of the lesions.

(J Med Internet Res 2021;23(4):e25053) doi:10.2196/25053

KEYWORDS

early gastric cancer; artificial intelligence; machine learning; endoscopic submucosal dissection; undifferentiated; gastric cancer; endoscopy; dissection

Introduction

Endoscopic submucosal dissection (ESD) is indicated for the treatment of patients with early gastric cancer (EGC) satisfying prespecified criteria, including histology, according to the differentiation, specific lesion size, morphology, and whether ulcers are present or not in the target lesion. The long-term prognosis following ESD for cases of EGC meeting the ESD criteria (achievement of curative resection) is comparable to that achieved with surgical resection [1,2]. In the context of histology, the undifferentiated type of EGC (U-EGC) generally refers to poorly differentiated adenocarcinoma, signet-ring cell carcinoma, or mucinous adenocarcinoma [3,4]. Although U-EGC is included among the expanded indications of ESD (mucosal U-EGC<2 cm without ulceration and without evidence of lymphovascular invasion), the rate of curative resection in U-EGC has remained very low—reported previously as 61.4% in a meta-analysis and 36.4% in a nationwide cohort study in Korea [5,6]. This implies that an unmet need persists regarding the accurate prediction of curative resection in U-EGC (ie, difficulty in adopting a precise ESD indication). Therefore, proper candidate selection prior to ESD is important.

Endoscopists predict the probability of curative resection by considering the size and shape of the lesion and whether ulcers are present or not. These components together compose the indications of ESD. In addition, lesion location, which can suggest the expected technical difficulty during the procedure and hint at the general condition of the patient, is also considered prior to conducting ESD. However, U-EGC has distinctive growth patterns relative to differentiated-type EGC [3,4,6,7]. U-EGC is known to extend laterally along the proliferative zone in the intermediate layer of the mucosa (subepithelial spreading), and the development pattern from the intermediate layer could lead to nonexposure to the surface mucosa, limiting the precise measurement of lesion size [5,8]. Subepithelial-spreading signet-ring cell carcinoma is more prevalent than the epithelial-spreading type in cases with background atrophy or intestinal metaplasia of the gastric mucosa [9,10]. Further, ESD of poorly differentiated adenocarcinoma presents a stronger association with submucosal invasion relative to that of signet-ring cell carcinoma [6]. Although adopting a precise indication is a key ability of endoscopists, U-EGC itself is a risk factor for a greater out-of-indication rate, leading to noncurative resection [11,12].

With the extensive production and collection of ongoing medical data, the application of artificial intelligence has been attempted in clinical practice [13]. Machine learning (ML) is a mathematical artificial intelligence algorithm automatically built from given data to predict precise outcomes in uncertain conditions without being explicitly programmed [14]. Examples of ML include Bayesian inferences, decision trees, support vector machines, deep neural networks, or ensemble methods (bagging or boosting) [14]. In short, ML is a type of applied statistical technique and is characterized by high accuracy. We aimed to establish an ML model to better predict the possibility of curative resection in U-EGC prior to ESD.

Methods

Ethical Statement

This study was approved by the Institutional Review Board of the Chuncheon Sacred Heart Hospital, Korea (no. 2020-07-019). It adhered to the principles expressed in the Declaration of Helsinki.

Data Sets

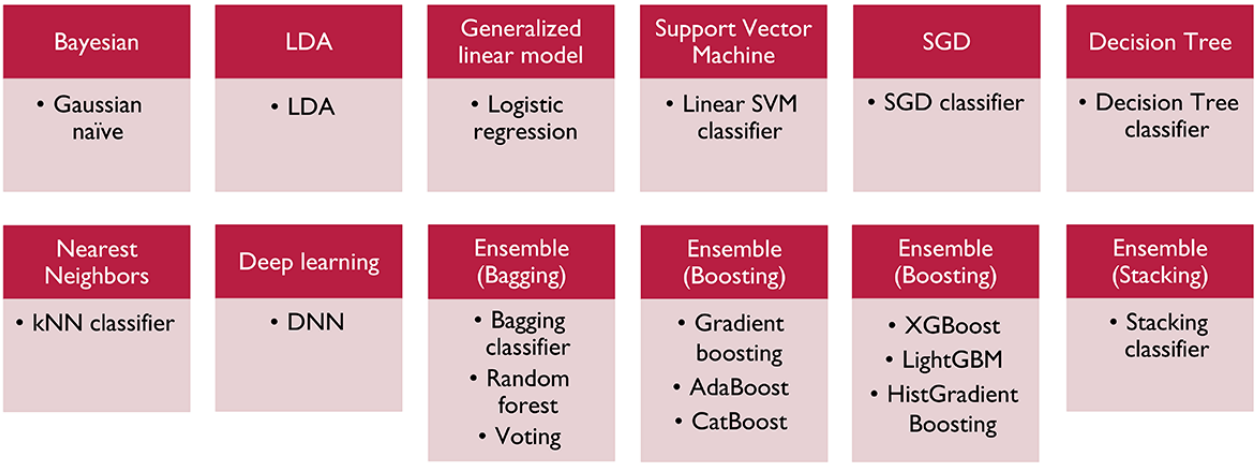
A nationwide cohort of 2703 U-EGCs treated by ESD (n=967) or surgery (n=1736) from 2006 to 2015 composed the training and internal validation groups. Eligible subjects were retrospectively enrolled from 18 university hospitals in Korea. Separately, an independent data set involving the Korean ESD registry with 275 U-EGCs and an Asan medical center data set with 127 U-EGCs treated by ESD were used for external validation. Subjects in the Korean ESD registry data set were retrospectively identified from 8 institutions of Korea [6], having been treated with ESD from 2006 to 2015, while subjects in the Asan medical center data set were treated by ESD from 2007 to 2013. All these data sets were mutually exclusive.

ML Models

All the currently available types of supervised ML classifiers were tested for the establishment of a curative resection prediction model in U-EGC. In total, 18 ML classifiers were assessed, including naïve Bayes in Bayesian inferences, linear-discriminant analysis, logistic regression in generalized linear modeling, linear support vector machine, stochastic gradient descent, decision tree, k-nearest neighbors, deep neural networks, bagging ensemble methods (bagging classifier, random forest, and voting classifier), boosting ensemble methods (gradient boosting, adaptive boosting, categorical Boosting, extreme gradient boosting [XGBoost], light gradient boosting machine, histogram-based gradient boosting), and a stacking ensemble method (stacking classifier). The Gaussian Naïve Bayes classifier is a model based on the Bayes' theorem encompassing the assumption that there is independence between the features. A generalized linear model is the extension of a linear model set up to include cases where the dependent variable is not normally distributed. We adopted the logistic regression classifier for this study. The support vector machine is a model that defines a decision boundary (hyperplane), that is, a reference line for classification. The stochastic gradient descent is a model for linear classifiers under convex loss functions such as support vector machine and logistic regression [15]. The decision tree is an algorithm that automatically finds rules in the data and creates tree-based classification rules.

k-nearest neighbors is a classification or clustering algorithm that relies on distance metrics measures for similarity. Deep neural networks refer to an artificial neural network with multiple hidden layers between the input and output layers that learns from input data and optimizes the output classification with mathematical calculations. Ensemble algorithms combine multiple classification models to achieve better performance and can be classified as either bagging, boosting, or stacking methods. Bagging is a parallel ensemble method that fits individual random samples of the data set and aggregates the predictions of each model for the final classification (bootstrap aggregation) [15]. This meta-estimator can reduce the variance of each classification model by introducing randomization for the model establishment and then creating an ensemble out of it. As such, bagging reduces overfitting of the ML model [15]. Separately, boosting algorithms attempt to conduct ensemble modeling sequentially by learning from the errors of the previous model and updating the weight of subsequent models to optimize the loss functions and reduce the overall bias. In contrast with learning from homogenous weak models in the bagging and boosting algorithms, stacking algorithms learn from heterogeneous models, creating a meta-model for the final classification. For the current ML analysis of this study, we used bagging classification, random forest, and voting classification for the bagging ensemble methods and gradient boosting, adaptive boosting, categorical boosting, XGBoost, light gradient boosting machine, and histogram-based gradient boosting for the boosting methods. For the stacking algorithm, we chose stacking classification. All the ML classifiers were imported from the scikit-learn package version 0.23.2 using the Python programming language (version 3.8.5, Python Software Foundation). Figure 1 shows the types of ML classifiers examined in this study.

Figure 1. Machine learning classifiers used in this study. AdaBoost: adaptive boosting; CatBoost: categorical boosting; DNN: deep neural network; HistGradientBoosting: histogram-based gradient boosting; kNN: k-nearest neighbors; ML: machine learning; LightGBM: light gradient boosting machine; LDA: linear discriminants analysis; SGD: stochastic gradient descent; SVM: support vector machine; XGBoost: extreme gradient boosting.



Variables, Primary Outcome, and Data Splitting

A total of 18 ML classifiers were used for the establishment of prediction models of curative resection with the following variables: age; sex; location, size, and shape of the lesion; and whether ulcers were present or not. The primary outcome was the accuracy of the established ML models for the prediction of curative resection with the given variables of the lesions. Thus, the main metric was the classifying accuracy. Each data set was prepared in the .csv file format. After uploading .csv files to the Google Colaboratory analysis platform, 2703 U-EGC data points were randomly split into training and internal validation sets according to a ratio of 9:1.

Definitions of the Variables

Among the variables used in this study, patient age and the size of the lesion were the continuous variables and the others were considered as categorical variables. The location of the lesion was categorized by both longitudinal location (lower-third, mid-third, and upper-third) and circular location (lesser curvature, greater curvature, posterior wall, and anterior wall). The shape of the lesion was defined in accordance with the Japanese classification: elevated, flat, or depressed according to the morphological characteristics. According to this system, type I (protruded) and type IIa (superficial elevated) were considered as elevated, type IIb (flat) and type IIc (superficial depressed) were considered as flat, and type III (excavated) was considered as depressed [4]. Curative resection was defined as complete resection of U-EGC with a diameter of 2 cm or less and a lesion confined to the mucosa, with negative lateral and deep resection margins and lymphovascular invasion. Noncurative resection referred to cases in which the resected lesion did not fulfill these criteria.

Statistical Analysis and Explainable Artificial Intelligence

Continuous variables were expressed as mean (SD) and categorical variables were expressed as numbers and percentages. Descriptive synthesis was conducted to reveal the baseline characteristics of the training and internal validation data set and external validation data set. To add to the interpretability of the established ML model, we performed an explainable artificial intelligence analysis. To elucidate the variables associated with lesions either accurately or inaccurately determined by the ML model, univariable analysis was conducted (Student *t* test and Fisher exact test for continuous and categorical variables, respectively). A two-tailed *P* value

of less than .05 was adopted as the threshold for statistical significance. These analyses were performed using SPSS version 24.0. (IBM Corporation). Additionally, a feature importance (or permutation importance) analysis was completed to reveal which variables primarily contributed to the model's decision process [16,17]. This assessment measures the predictive error when a certain feature value is randomly shuffled; therefore, insignificant features do not affect the performance of the model [15]. Feature importance is measured by the F-score, which represents the ratio between the explained and the unexplained variance [17]. A decision process tree was plotted to visualize the step-by-step process of the decision making of the established ML model using the Graphviz package (version 0.14.1; AT&T Labs Research). A partial-dependence plot tool box (version 0.2.0) in the scikit-learn package to visualize the important features for the ML model was adopted and the target plot and interaction plot were visualized [18,19]. A Shapley additive explanations (version 0.35.0) analysis is an approach used to explain the output of any ML model using Shapley values and the degree of independence between features. The Shapley value expresses how much each feature contributes to creating the overall performance and represents feature importance while maintaining consistent and locally accurate additive feature attribution for a particular prediction [20].

Results

Characteristics of the Training, Internal Validation, and External Validation Data Sets

The training and internal validation data sets contained not only endoscopically resected cases but also surgically removed cases of U-EGC. The first external validation data set was composed of a nationwide cohort of cases of ESD performed for U-EGC, while the second external validation data set consisted of cases of ESD performed for U-EGC from a single hospital with the largest degree of ESD experience to date in Korea. Therefore, the included data sets were marked by different clinical characteristics. Table 1 presents the detailed clinical characteristics of the included lesions in this study. A male sex predominance was consistently observed in all data sets. Patient age ranged from 64.1 (SD 13.0) years to 67.8 (SD 12.0) years. In the context of endoscopic findings, the lower-third part in the longitudinal location (2069/2703, 76.5% and 214/275, 77.8%) and lesser curvature in the circular location (97/275, 34.5% and 945/2703, 34.9%) were the most frequent lesion positions in the training and internal validation dataset and first external validation data set, respectively. Meanwhile, the mid-third part was the most frequent lesion location in the longitudinal location (61/127, 48.1%) for U-EGC in the second external validation data set. The mean endoscopic size of the included lesions ranged from 21.7 (SD 12.5) mm to 27.9 (SD 16.2) mm. Depressed lesions (type IIc) were observed as the most frequent morphological type in the training and internal validation data set and second external validation data set (1762/2703, 65.2% and 62/127, 48.8%, respectively), while the first external validation presented an even distribution of elevated, flat, and depressed lesion morphologies. Meanwhile, 63 (22.9%) and 16 (12.6%) cases had ulcers in the first and second external validation data sets, respectively. The overall rate of curative resection ranged from 36.4% (100/275) to 74.4% (2010/2703).

Table 1. Baseline characteristics of the included data sets.

Characteristics	Training and internal validation set (n=2703)	First external validation set (n=275)	Second external validation set (n=127)
Sex, n (%)			
Male	1427 (52.8)	165 (60.0)	80 (62.9)
Female	1276 (47.2)	110 (40.0)	47 (37.0)
Age (years), mean (SD)	65.9 (12.4)	67.8 (12.0)	64.1 (13.0)
Longitudinal location, n (%)			
Lower-third	2069 (76.5)	214 (77.8)	53 (41.7)
Mid-third	336 (12.4)	28 (10.2)	61 (48.1)
Upper-third	298 (11.0)	33 (12.0)	13 (10.2)
Circular location, n (%)			
Lesser curvature	945 (34.9)	95 (34.5)	49 (38.6)
Greater curvature	557 (20.6)	58 (21.1)	27 (21.3)
Posterior wall	585 (21.6)	68 (24.7)	22 (17.3)
Anterior wall	607 (22.5)	54 (19.6)	29 (22.8)
More than 2 areas involved	9 (0.3)	0 (0)	0 (0)
Endoscopic size of the lesion (mm), mean (SD)	21.7 (12.5)	27.9 (16.2)	21.7 (12.6)
Morphology, n (%)			
Elevated	375 (13.9)	101 (36.7)	28 (22.1)
Flat	566 (20.9)	98 (35.6)	37 (29.1)
Depressed	1762 (65.2)	76 (27.6)	62 (48.8)
Ulcer, n (%)			
Present	504 (18.6)	63 (22.9)	16 (12.6)
None	2199 (81.4)	212 (77.1)	111 (87.4)
Curative resection, n (%)			
Yes	2010 (74.4)	100 (36.4)	87 (68.5)
No	693 (25.6)	175 (63.6)	40 (31.5)

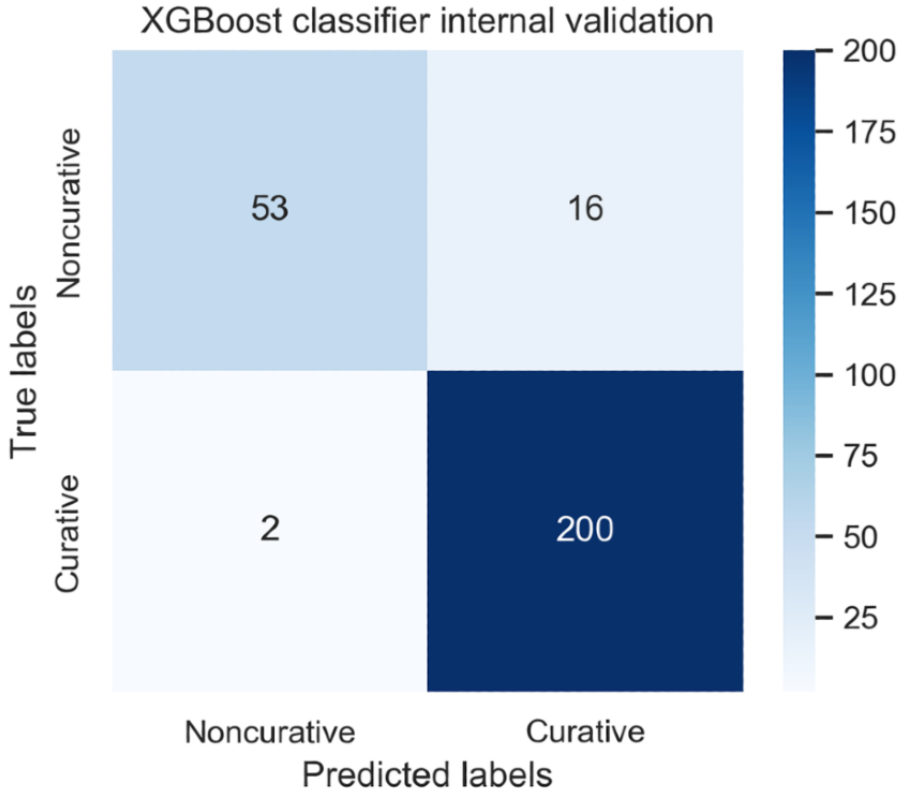
Internal Validation Performance

Table 2 shows the prediction performance of 18 ML classifiers for internal validation. The XGBoost classifier demonstrated the best performance as follows: internal validation accuracy 93.4%, 95% CI 90.4%-96.4%; precision 92.6%, 95% CI 89.5%–95.7%; recall 99.0%, 95% CI 97.8%-99.9%; and F1 score 95.7%, 95% CI 93.3%-98.1%. In detail, the XGBoost classifier required several parameter settings for the establishment of the ML model. The initial classifying performance of the XGBoost classifier established by us was as follows: internal validation accuracy 79.0%, 95% CI 74.1%-83.9%; precision 80.9%, 95% CI 76.2%-85.6%; recall 94.1%, 95% CI 91.3%-96.9%; and F1 score 87.0%, 95% CI 83.0%-91.0%. To discern the optimal hyperparameter setting for the establishment of the ML model, we relied on the GridSearchCV library (version 0.22) [15] to automatically search among multiple optimal parameter values to fit estimators of an ML model. By using the GridSearchCV analysis, we found the optimal hyperparameters for the best performance as follows: learning rate 0.4, maximum depth 6, and number of estimators 100. Figure 2 shows the confusion matrix for the XGBoost classifier in the internal validation data set.

Table 2. Internal validation performance for the prediction of curative resection of undifferentiated type of early gastric cancer by using 18 machine learning classifiers.

Machine learning classifier	Accuracy (%) (95% CI)	Precision (%) (95% CI)	Recall (%) (95% CI)	F1 score (%) (95% CI)
Gaussian Naïve Bayes	73.8 (68.6-79.0)	86.2 (82.1-90.3)	77.2 (72.2-82.2)	81.5 (76.9-86.1)
Linear discriminant analysis classifier	76.4 (71.3-81.5)	77.4 (72.4-82.4)	96.5 (94.3-98.7)	85.9 (81.8-90.0)
Logistic regression classifier	77.5 (72.5-82.5)	80.5 (75.8-85.2)	92.1 (88.9-95.3)	85.9 (81.8-90.0)
Linear support vector machine classifier	74.5 (69.3-79.7)	74.5 (69.3-79.7)	99.9 (98.8-99.9)	85.4 (81.2-89.6)
Stochastic gradient descent classifier	74.5 (69.3-79.7)	77.6 (72.6-82.6)	92.6 (89.5-95.7)	84.4 (80.1-88.7)
Decision tree classifier	74.5 (69.3-79.7)	74.5 (69.3-79.7)	99.9 (98.8-99.9)	85.4 (81.2-89.6)
k-nearest neighbors classifier	72.0 (66.7-77.3)	78.1 (73.2-83.0)	86.6 (82.5-90.7)	82.2 (77.6-86.8)
Deep neural network	77.9 (73.0-82.8)	80.6 (75.9-85.3)	92.6 (89.5-95.7)	86.2 (82.1-90.3)
Ensemble (bagging)				
Bagging classifier	72.0 (66.7-77.3)	81.2 (76.5-85.9)	81.2 (76.5-85.9)	81.2 (76.5-85.9)
Random forest classifier	72.7 (67.4-78.0)	80.2 (75.5-84.9)	84.2 (79.9-88.5)	82.1 (77.5-86.7)
Voting classifier	84.5 (80.2-88.8)	88.1 (84.2-92.0)	91.6 (88.3-94.9)	89.8 (86.2-93.4)
Ensemble (boosting)				
Gradient boosting classifier	77.5 (72.5-82.5)	80.5 (75.8-85.2)	92.1 (88.9-95.3)	85.9 (81.8-90.0)
Adaptive boosting classifier	77.9 (73.0-82.8)	81.1 (76.4-85.8)	91.6 (88.3-94.9)	86.0 (81.9-90.1)
Categorical boosting classifier	84.1 (79.7-88.5)	83.8 (79.4-88.2)	97.5 (95.6-99.4)	90.2 (86.7-93.7)
Extreme gradient boosting classifier	93.4 (90.4-96.4)	92.6 (89.5-95.7)	99.0 (97.8-99.9)	95.7 (93.3-98.1)
Light gradient boosting machine classifier	75.6 (70.6-80.8)	80.9 (76.2-85.6)	88.1 (84.2-92.0)	84.4 (80.1-88.7)
Histogram-based gradient boosting classifier	85.2 (81.0-89.4)	84.9 (80.689.2)	97.5 (95.6-99.4)	90.8 (87.4-94.2)
Ensemble (stacking)	75.6 (70.5-80.7)	78.6 (73.7-83.5)	92.6 (89.5-95.7)	85.0 (80.7-89.3)

Figure 2. Confusion matrix for the extreme gradient boosting classifier in the internal validation cohort. XGBoost: extreme gradient boosting.



External Validation Performance in the XGBoost Classifier

For the first external validation data set, the XGBoost classifier demonstrated its performance as follows: external validation accuracy 81.5%, 95% CI 76.9%-86.1%; precision 83.6%, 95% CI 79.2%-88.0%; recall 61.0%, 95% CI 55.2%-66.8%; and F1 score 70.5%, 95% CI 65.1%-75.9%. Then, for the second external validation data set, the XGBoost classifier demonstrated its performance as follows: external validation accuracy 89.8%, 95% CI 84.5%-95.1%; precision 90.2%, 95% CI 85.0%-95.4%; recall 95.4%, 95% CI 91.8%-99.0%; and F1 score 92.7%, 95% CI 88.2%-97.2%. Figure 3 and Figure 4 show the confusion matrices for the XGBoost classifier in the first and second external validation data sets, respectively.

Figure 3. Confusion matrix for the extreme gradient boosting classifier in the first external validation cohort. XGBoost: extreme gradient boosting.

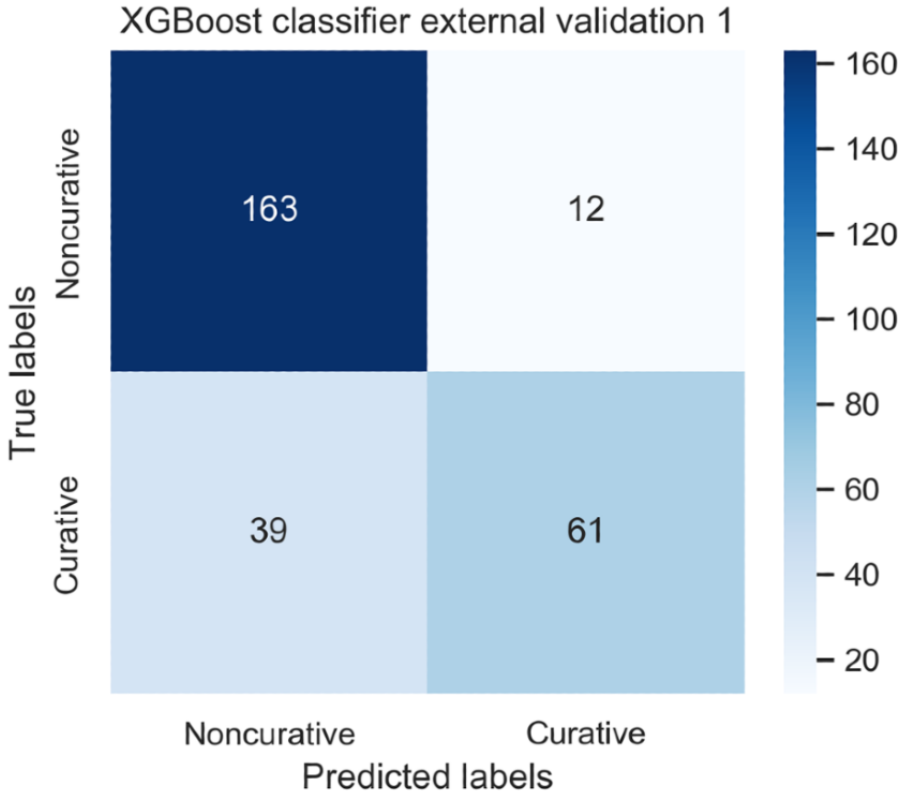
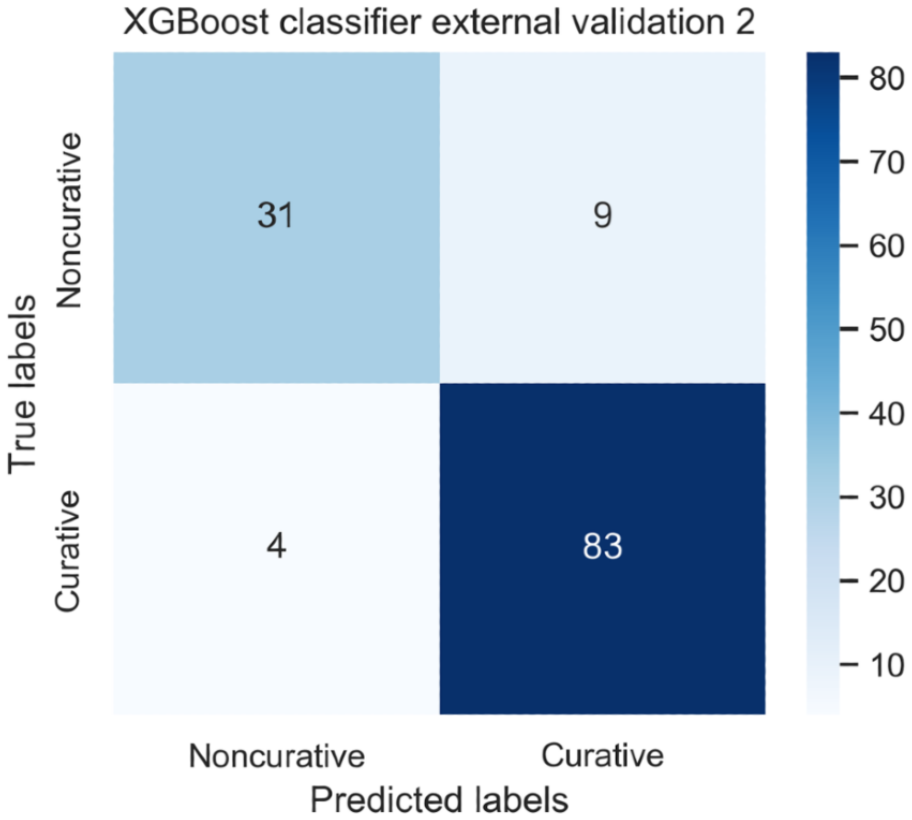


Figure 4. Confusion matrix for the extreme gradient boosting in the second external validation cohort. XGBoost: extreme gradient boosting.



Explainable Artificial Intelligence

Table 3 shows the univariable analysis for the associated factors of lesions determined accurately or inaccurately in the curative resection of U-EGC by the XGBoost classifier. Notably, there was no single significant factor associated with lesions determined either accurately or inaccurately by the XGBoost classifier.

Table 3. Univariable analysis of the associated factors of lesions determined accurately or inaccurately in the curative resection of undifferentiated type of early gastric cancer by the extreme gradient boosting classifier.

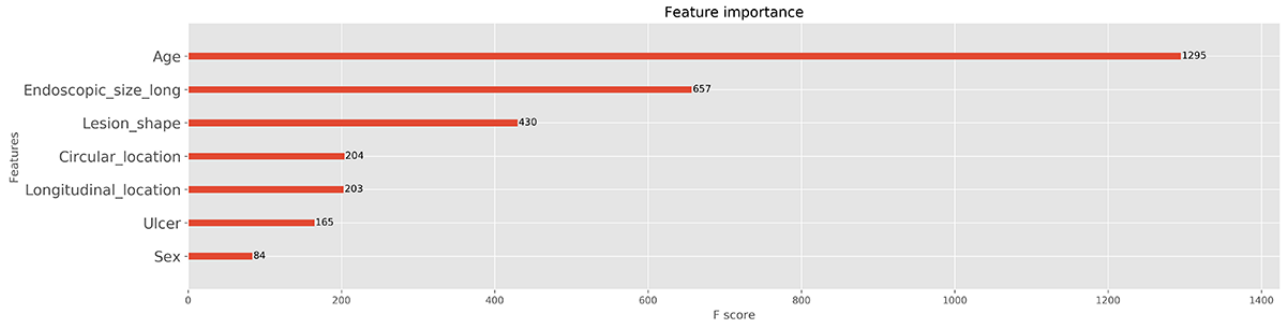
Characteristics	First external validation set			Second external validation set		
	Accurately determined by XGBoost[a] classifier (n=224)	Inaccurately determined by XGBoost classifier (n=51)	*P* value	Accurately determined by XGBoost classifier (n=114)	Inaccurately determined by XGBoost classifier (n=13)	*P* value
Sex, n (%)			.06			.37
Male	128 (57.1)	37 (73)		70 (61.4)	10 (77)	
Female	96 (42.9)	14 (28)		44 (38.6)	3 (23)	
Age (years), mean (SD)	67.3 (12.6)	70.0 (9.2)	.09	63.9 (13.3)	65.5 (10.4)	.66
Longitudinal location, n (%)			.22			.33
Lower-third	173 (77.2)	41 (80)		50 (43.9)	3 (23)	
Mid-third	21 (9.4)	7 (14)		53 (46.5)	8 (62)	
Upper-third	30 (13.4)	3 (6)		11 (9.6)	2 (15)	
Circular location, n (%)			.38			.29
Lesser curvature	74 (33.0)	21 (41)		46 (40.4)	3 (23)	
Greater curvature	45 (20.1)	13 (26)		23 (20.2)	4 (31)	
Posterior wall	58 (25.9)	10 (20)		21 (18.4)	1 (8)	
Anterior wall	47 (20.9)	7 (14)		24 (21.1)	5 (39)	
Endoscopic size of the lesion (cm), mean (SD)	28.4 (16.4)	25.5 (14.7)	.25	22.2 (12.7)	17.1 (11.4)	.16
Morphology, n (%)			.36			.93
Elevated (I, IIa, and IIa+IIc)	78 (34.8)	23 (45)		25 (21.9)	3 (23)	
Flat (IIb)	81 (36.2)	17 (33)		34 (29.8)	3 (23)	
Depressed (IIc)	65 (29)	11 (22)		55 (48.2)	7 (54)	
Ulcer, n (%)			.86			.21
Present	52 (23.2)	11 (22)		13 (11.4)	3 (23)	
None	172 (76.8)	40 (78)		101 (88.6)	10 (77)	

[a]XGBoost: extreme gradient boosting.

Figure 5 shows the feature importance plot for the XGBoost classifier. Age, endoscopic size, and morphology of the lesions were the three most significant factors for the establishment of the ML model, in sequence. Multimedia Appendix 1 illustrates the decision process tree for the XGBoost classifier prior to adopting the GridSearchCV library. This simplified tree shows the step-by-step determination process of the ML model. The final leaf score is inserted in the following equation: $p(x) = 1 / 1 + e^{-leaf\ score}$. Any value over 0.5 (50%) indicates curative resection and any value less than 0.5 indicates noncurative resection, as predicted by the XGBoost classifier [21]. Multimedia Appendix 2 shows the final decision process tree for the XGBoost classifier after adopting the GridSearchCV library, which presented the best performance in the internal validation. Endoscopic size of the lesion, patient age, and longitudinal location of the lesion were the important factors, in sequence. Multimedia Appendix 3 shows the partial-dependence target plot for the feature of endoscopic size of the lesion in the first external validation assessment. The probability of curative resection for the lesions with sizes ranging from 4 mm to 10 mm reached 80%. Meanwhile, U-EGC lesions with sizes ranging from 20.78 mm to 26.22 mm showed the lowest probability of curative resection at 16.1%. Multimedia Appendix 4 presents the two-way partial-dependence target plot for the features of endoscopic size of the lesion and patient age in the first external validation cohort. Given that the color of the circle above the imaginary line of Y=X is darker than that below the line, the endoscopic size and age are suggested to be correlated with curative resection of U-EGC. Multimedia Appendix 5 shows the partial-dependence interaction plot for the features of endoscopic size of the lesion and age in the first external validation group. Given that the contour lines are generally parallel to the Y-axis, the probability of curative resection is more dependent on the endoscopic size of the lesion. Since the feature importance analysis measures the prediction error after permutating the features' values, the results can be skewed when the said features exhibit dependency. However, the Shapley value considers the influence of the features on each other. Multimedia Appendix 6 and Multimedia Appendix 7 demonstrate the summary plot and bar plot of the Shapley additive explanations analysis, respectively, where endoscopic

size of the lesion and age are the important features for the model output.

Figure 5. Feature importance plot for the extreme gradient boosting classifier in the internal validation cohort. The average F-score was calculated through 50 repetitions of five-fold cross-validation in the training data set.



Discussion

This study introduces the good performance of an ML model applied to the prediction of curative resection of U-EGC prior to ESD, suggesting the possibility of a beneficial effect of ML modeling for decision making in this part of clinical practice [22]. Moreover, thorough external validations confirmed the higher rate of curative resection predicted by ML modeling as compared with curative resection rates reported by clinicians. To our knowledge, this is the first study to establish and confirm the predictive performance of an artificial intelligence model for the therapeutic outcomes of ESD for U-EGC. Indeed, ML is characterized as a computer-aided prediction method and its most important benefit in this context consists of the improvement in predictive accuracy for curative resection prior to ESD. The proper selection of candidates for ESD is essential before beginning ESD. The most fundamental hypothesis is that endoscopic resection can be performed with curative intent in cases of EGC without lymph node metastasis. Therefore, indications of ESD were established using a combination of factors associated with a negligible lymph-node metastasis rate from the retrospective analysis of surgically resected specimens [3]. These indications are categorized by differentiated-type EGC and U-EGC according to the differentiation, specific size, and morphological and histological conditions of the involved lesion. However, optical endoscopic determination of the factors stated above involves operator-dependent characteristics. In the study of a Korean multicenter registry of ESD for U-EGC, there was a discrepancy between pre-ESD indications and post-ESD criteria in 36.7% of all the lesions [6]. Underestimation of the size was the most common reason for noncurative resection (71.4%), followed by underestimation of the depth of invasion (32%) and unpredictability of lymphovascular invasion (14.9%) [6]. Although adopting a precise indication is important, U-EGC itself is a risk factor for an enhanced out-of-indication rate, leading to noncurative resection; therefore, more strict indications might be necessary for pursuing the ESD of U-EGC [11,12].

Another important finding of this study is the presentation of the determination reason or process of the ML model through the explainable artificial intelligence analysis. Notably, there is a tradeoff between accuracy and interpretability in the classification model of ML [14]. Although the ML approach exhibited high degrees of accuracy based on complex calculations, it is characterized by low interpretability (artificial intelligence is more generally characterized as being of a “black-box nature”) [14]. Conventional statistical analyses such as univariate or multivariate logistic regression analyses in previous studies have shown the reasons underlying the lower curative resection rate of ESD for U-EGC [5,6]. However, there is a limitation in the explanatory power of the overall model (low accuracy) in these studies. The XGBoost classifier used parallel-tree boosting analysis to provide highly efficient and accurate predictions. Through the ensemble model and extensive explainable artificial intelligence analysis, we identified the size of the lesion as being the most important feature for the successful prediction of curative resection in the ESD of U-EGC. Although a prospective trial of ESD for U-EGC that satisfied the expanded indication reported an excellent long-term survival rate [6,23,24], more cautious application or restriction of ESD indications has been recommended, especially regarding the size categorization [3,25]. Most recently published studies have also indicated that small intramucosal U-EGC lesions measuring less than 1.0 cm or 1.5 cm without lymphovascular invasion should be considered as the ESD candidate [26,27]. The explainable artificial intelligence analysis in our study also revealed that U-EGC lesions of less than 1 cm have the greatest probability of curative resection (Multimedia Appendix 3). Considering that the aim of this study was not the validation of current ESD criteria, further studies with robust analysis would elucidate the value of these findings.

In the context of ecological factors, age and gender have been tested with the endoscopic factors for the potential variable for the curative resection rate prediction. However, these variables were not consistently identified as important indicators for predicting curative resection [28-30]. Although feature importance analysis (Figure 5) or Shapley additive explanations analysis (Multimedia Appendix 6) in our study revealed that age is an important variable for the ML determination process, explainable artificial intelligence analysis is currently an experimental method to understand how ML judges. It is presumed that the reason ML shows higher accuracy than traditional statistics is that it performs a complex operation that considers all variables. It is true that age is an important factor influencing ML judgment, but further explainable artificial intelligence statistics can explain how much it affects the actual curative resection.

Although this study established and rigorously validated the predictive performance of the designed ML model, several inevitable limitations became apparent. First, there was some discrepancy in the validation performance between the first and second external data sets. The indications of ESD for U-EGC have not been approved by all endoscopists. Therefore, practice patterns adopting ESD indications for U-EGC have been heterogenous depending on the institution. The first external validation data set was more heterogenous with respect to the baseline characteristics and therapeutic outcomes. However, the second data set was collected from a single institution, thus providing a more discrete application pattern of the ESD indication for U-EGC. Second, patient age was an important feature in the explainable artificial intelligence analysis; however, this feature does not perfectly reflect the general condition of the patient. Further, there is no age factor for ESD indications. However, the general condition of the patients is frequently considered in the determination of whether to pursue ESD. Therefore, clinical factors that reliably reflect patients' health status other than age should be developed and considered so as to attain the most favorable therapeutic outcomes of ESD. Third, the training and internal validation data sets included cases that were surgically resected as well as endoscopically resected cases. Endoscopists decide whether to perform ESD or surgery when they detect U-EGC. In other words, it has not been determined which U-EGC is a candidate for ESD or surgery. All the U-EGCs resected with surgery or ESD were included as it was not always accurate and appropriate for the endoscopists to differentiate between ESD or surgery. If only U-EGCs that were resected by ESD were collected, a clear ESD candidate would have been collected, which in itself may be a selection bias. In conclusion, we established an ML model capable of accurately predicting the curative resection of U-EGC prior to ESD by considering the morphological and ecological characteristics of the lesions. A clinical application study in a randomized controlled manner would elucidate the real value of this ML model.

Acknowledgments

Funding for this research was provided by the Bio & Medical Technology Development Program of the National Research Foundation, the Korean government, Ministry of Science and ICT (grant number NRF2017M3A9E8033253), and the Korean College of Helicobacter and Upper Gastrointestinal Research Foundation Grant (2017-08). All investigators have access to the final data set. All the data are accessible and are available upon request by the corresponding author.

Authors' Contributions

CSB, WGS, JYA, JK, and IJC curated the data. CSB acquired the funds. CSB and WGS conceptualized this study, conducted the investigations, designed the methodology for this study, administered the project, performed the formal analysis, supervised the study, wrote the original draft, and reviewed and edited the manuscript. Resources were provided by CSB, WGS, JYA, JK, and IJC.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Decision process tree for the extreme gradient boosting classifier prior to adopting the GridSearchCV library in the internal validation assessment.
[DOCX File , 178 KB - jmir_v23i4e25053_app1.docx]

Multimedia Appendix 2
Decision process tree for the extreme gradient boosting classifier after adopting GridSearchCV library in the internal validation assessment.
[DOCX File , 659 KB - jmir_v23i4e25053_app2.docx]

Multimedia Appendix 3
Partial-dependence target plot for the feature of endoscopic size of the lesion in the first external validation cohort.
[DOCX File , 422 KB - jmir_v23i4e25053_app3.docx]

Multimedia Appendix 4
Two-way partial-dependence target plot for the features of endoscopic size of the lesion and patient age in the first external validation cohort.
[DOCX File , 469 KB - jmir_v23i4e25053_app4.docx]

Multimedia Appendix 5

Partial-dependence interaction plot for the features of endoscopic size of the lesion and patient age in the first external validation cohort.
[DOCX File , 816 KB - jmir_v23i4e25053_app5.docx]

Multimedia Appendix 6
Summary plot of the Shapley additive explanations analysis. Endoscopic size of the lesion and patient age are the important features for the model output.
[DOCX File , 507 KB - jmir_v23i4e25053_app6.docx]

Multimedia Appendix 7
Bar plot of the Shapley additive explanations analysis. Endoscopic size of the lesion and patient age are the important features for the model output.
[DOCX File , 265 KB - jmir_v23i4e25053_app7.docx]

References

1. Choi IJ, Lee JH, Kim Y, Kim CG, Cho S, Lee JY, et al. Long-term outcome comparison of endoscopic resection and surgery in early gastric cancer meeting the absolute indication for endoscopic resection. Gastrointest Endosc 2015 Feb;81(2):333-41.e1. [doi: 10.1016/j.gie.2014.07.047] [Medline: 25281498]
2. Jeon HK, Kim GH, Lee BE, Park DY, Song GA, Kim DH, et al. Long-term outcome of endoscopic submucosal dissection is comparable to that of surgery for early gastric cancer: a propensity-matched analysis. Gastric Cancer 2018 Jan;21(1):133-143. [doi: 10.1007/s10120-017-0719-4] [Medline: 28397011]
3. Bang CS, Baik GH. Pitfalls in the Interpretation of Publications about Endoscopic Submucosal Dissection of Early Gastric Cancer with Undifferentiated-Type Histology. Clin Endosc 2019 Jan;52(1):30-35 [FREE Full text] [doi: 10.5946/ce.2018.158] [Medline: 30650945]
4. Japanese Gastric Cancer Association. Japanese gastric cancer treatment guidelines 2018 (5th edition). Gastric Cancer 2021 Jan;24(1):1-21 [FREE Full text] [doi: 10.1007/s10120-020-01042-y] [Medline: 32060757]
5. Bang CS, Baik GH, Shin IS, Kim JB, Suk KT, Yoon JH, et al. Endoscopic submucosal dissection for early gastric cancer with undifferentiated-type histology: A meta-analysis. World J Gastroenterol 2015 May 21;21(19):6032-6043 [FREE Full text] [doi: 10.3748/wjg.v21.i19.6032] [Medline: 26019470]
6. Bang CS, Park JM, Baik GH, Park JJ, Joo MK, Jang JY, et al. Therapeutic Outcomes of Endoscopic Resection of Early Gastric Cancer with Undifferentiated-Type Histology: A Korean ESD Registry Database Analysis. Clin Endosc 2017 Nov;50(6):569-577 [FREE Full text] [doi: 10.5946/ce.2017.017] [Medline: 28743132]
7. Lee JH, Kim JG, Jung H, Kim JH, Jeong WK, Jeon TJ, et al. Clinical practice guidelines for gastric cancer in Korea: an evidence-based approach. J Gastric Cancer 2014 Jun;14(2):87-104 [FREE Full text] [doi: 10.5230/jgc.2014.14.2.87] [Medline: 25061536]
8. Sawada S, Fujisaki J, Yamamoto N, Kato Y, Ishiyama A, Ueki N, et al. Expansion of indications for endoscopic treatment of undifferentiated mucosal gastric cancer: analysis of intramucosal spread in resected specimens. Dig Dis Sci 2010 May;55(5):1376-1380. [doi: 10.1007/s10620-009-0883-4] [Medline: 19657738]
9. Kim J. Important considerations when contemplating endoscopic resection of undifferentiated-type early gastric cancer. World J Gastroenterol 2016 Jan 21;22(3):1172-1178 [FREE Full text] [doi: 10.3748/wjg.v22.i3.1172] [Medline: 26811655]
10. Kim H, Kim J, Lee YC, Kim H, Youn YH, Park H, et al. Growth Patterns of Signet Ring Cell Carcinoma of the Stomach for Endoscopic Resection. Gut Liver 2015 Nov 23;9(6):720-726 [FREE Full text] [doi: 10.5009/gnl14203] [Medline: 25473081]
11. Horiuchi Y, Fujisaki J, Yamamoto N, Ishizuka N, Omae M, Ishiyama A, et al. Undifferentiated-type component mixed with differentiated-type early gastric cancer is a significant risk factor for endoscopic non-curative resection. Dig Endosc 2018 Sep;30(5):624-632. [doi: 10.1111/den.13059] [Medline: 29570860]
12. Bang CS, Yang YJ, Lee JJ, Baik GH. Endoscopic Submucosal Dissection of Early Gastric Cancer with Mixed-Type Histology: A Systematic Review. Dig Dis Sci 2020 Jan;65(1):276-291. [doi: 10.1007/s10620-019-05761-w] [Medline: 31367880]
13. Cho B, Bang CS. Artificial Intelligence for the Determination of a Management Strategy for Diminutive Colorectal Polyps: Hype, Hope, or Help. Am J Gastroenterol 2020 Jan;115(1):70-72. [doi: 10.14309/ajg.0000000000000476] [Medline: 31770118]
14. Yang YJ, Bang CS. Application of artificial intelligence in gastroenterology. World J Gastroenterol 2019 Apr 14;25(14):1666-1683 [FREE Full text] [doi: 10.3748/wjg.v25.i14.1666] [Medline: 31011253]
15. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O. Scikit-learn: Machine learning in Python. J Mach Learn Res. 2011 Oct. URL: https://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf [accessed 2021-03-22]
16. Fisher A, Rudin C, Dominic F. Model class reliance: Variable importance measures for any machine learning model class, from the "Rashomon" perspective. Deep AI. 2018. URL: https://deepai.org/publication/

model-class-reliance-variable-importance-measures-for-any-machine-learning-model-class-from-the-rashomon-perspective [accessed 2021-03-29]
17. Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. In: Association for Computing Machinery Digital Library. 2016 Presented at: KDD 16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; August 2016; San Francisco, California p. 785-794. [doi: 10.1145/2939672.2939785]
18. Christoph M. Partial dependence plot. Interpretable Machine Learning: A Guide for Making Black Box Models Explainable. URL: https://christophm.github.io/interpretable-ml-book/pdp.html [accessed 2021-03-22]
19. Christoph M. Individual conditional expectation. Interpretable Machine Learning: A Guide for Making Black Box Models Explainable. URL: https://christophm.github.io/interpretable-ml-book/ice.html [accessed 2021-03-22]
20. Lundberg SM, Nair B, Vavilala MS, Horibe M, Eisses MJ, Adams T, et al. Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. Nat Biomed Eng 2018 Oct;2(10):749-760 [FREE Full text] [doi: 10.1038/s41551-018-0304-0] [Medline: 31001455]
21. XGboost parameters. URL: https://xgboost.readthedocs.io/en/latest/parameter.html [accessed 2021-03-22]
22. Triantafyllidis AK, Tsanas A. Applications of Machine Learning in Real-Life Digital Health Interventions: Review of the Literature. J Med Internet Res 2019 Apr 05;21(4):e12286 [FREE Full text] [doi: 10.2196/12286] [Medline: 30950797]
23. Jeon HK, Lee SJ, Kim GH, Park DY, Lee BE, Song GA. Endoscopic submucosal dissection for undifferentiated-type early gastric cancer: short- and long-term outcomes. Surg Endosc 2018 Apr;32(4):1963-1970. [doi: 10.1007/s00464-017-5892-3] [Medline: 29046960]
24. Lee S, Choi KD, Han M, Na HK, Ahn JY, Jung KW, et al. Long-term outcomes of endoscopic submucosal dissection versus surgery in early gastric cancer meeting expanded indication including undifferentiated-type tumors: a criteria-based analysis. Gastric Cancer 2018 May;21(3):490-499. [doi: 10.1007/s10120-017-0772-z] [Medline: 29052052]
25. Kim J. Strategy for Curative Endoscopic Resection of Undifferentiated-Type Early Gastric Cancer. Clin Endosc 2019 Jan;52(1):9-14 [FREE Full text] [doi: 10.5946/ce.2018.199] [Medline: 30674183]
26. Liang X, Wang Z, Li H, Ma G, Yu W, Zhou H, et al. Indication for endoscopic treatment based on the risk of lymph node metastasis in patients with undifferentiated early gastric cancer. Asian J Surg 2020 Oct;43(10):973-977 [FREE Full text] [doi: 10.1016/j.asjsur.2019.12.002] [Medline: 31964584]
27. Ryu DG, Choi CW, Kim SJ, Kang DH, Kim HW, Park SB, et al. Possible indication of endoscopic resection in undifferentiated early gastric cancer. Sci Rep 2019 Nov 14;9(1):16869 [FREE Full text] [doi: 10.1038/s41598-019-53374-0] [Medline: 31728024]
28. Xu P, Wang Y, Dang Y, Huang Q, Wang J, Zhang W, et al. Predictive Factors and Long-Term Outcomes of Early Gastric Carcinomas in Patients with Non-Curative Resection by Endoscopic Submucosal Dissection. Cancer Manag Res 2020;12:8037-8046 [FREE Full text] [doi: 10.2147/CMAR.S263525] [Medline: 32943936]
29. Libânio D, Pimentel-Nunes P, Afonso LP, Henrique R, Dinis-Ribeiro M. Long-Term Outcomes of Gastric Endoscopic Submucosal Dissection: Focus on Metachronous and Non-Curative Resection Management. GE Port J Gastroenterol 2017 Jan;24(1):31-39 [FREE Full text] [doi: 10.1159/000450874] [Medline: 28868336]
30. Lee SH, Kim MC, Jeon SW, Lee KN, Park JJ, Hong SJ. Risk Factors and Clinical Outcomes of Non-Curative Resection in Patients with Early Gastric Cancer Treated with Endoscopic Submucosal Dissection: A Retrospective Multicenter Study in Korea. Clin Endosc 2020 Mar;53(2):196-205 [FREE Full text] [doi: 10.5946/ce.2019.123] [Medline: 31648421]

Abbreviations

EGC: early gastric cancer
ESD: endoscopic submucosal dissection
ML: machine learning
U-EGC: undifferentiated type of early gastric cancer
XGBoost: extreme gradient boosting

Edited by R Kukafka; submitted 16.10.20; peer-reviewed by SC Park, JA Benítez-Andrades; comments to author 03.11.20; revised version received 15.11.20; accepted 21.03.21; published 15.04.21.

Please cite as:
Bang CS, Ahn JY, Kim JH, Kim YI, Choi IJ, Shin WG
Establishing Machine Learning Models to Predict Curative Resection in Early Gastric Cancer with Undifferentiated Histology: Development and Usability Study
J Med Internet Res 2021;23(4):e25053
URL: https://www.jmir.org/2021/4/e25053
doi:10.2196/25053
PMID:33856358

Review

Role of Artificial Intelligence Applications in Real-Life Clinical Practice: Systematic Review

Jiamin Yin[1*], BA; Kee Yuan Ngiam[2*], MBBS; Hock Hai Teo[1*], PhD

[1]Department of Information Systems and Analytics, School of Computing, National University of Singapore, Singapore, Singapore

[2]Department of Surgery, National University Hospital, Singapore, Singapore

[*]all authors contributed equally

Corresponding Author:
Hock Hai Teo, PhD
Department of Information Systems and Analytics
School of Computing
National University of Singapore
13 Computing Drive, NUS
Singapore, 117417
Singapore
Phone: 65 65162979
Email: teohh@comp.nus.edu.sg

Abstract

Background: Artificial intelligence (AI) applications are growing at an unprecedented pace in health care, including disease diagnosis, triage or screening, risk analysis, surgical operations, and so forth. Despite a great deal of research in the development and validation of health care AI, only few applications have been actually implemented at the frontlines of clinical practice.

Objective: The objective of this study was to systematically review AI applications that have been implemented in real-life clinical practice.

Methods: We conducted a literature search in PubMed, Embase, Cochrane Central, and CINAHL to identify relevant articles published between January 2010 and May 2020. We also hand searched premier computer science journals and conferences as well as registered clinical trials. Studies were included if they reported AI applications that had been implemented in real-world clinical settings.

Results: We identified 51 relevant studies that reported the implementation and evaluation of AI applications in clinical practice, of which 13 adopted a randomized controlled trial design and eight adopted an experimental design. The AI applications targeted various clinical tasks, such as screening or triage (n=16), disease diagnosis (n=16), risk analysis (n=14), and treatment (n=7). The most commonly addressed diseases and conditions were sepsis (n=6), breast cancer (n=5), diabetic retinopathy (n=4), and polyp and adenoma (n=4). Regarding the evaluation outcomes, we found that 26 studies examined the performance of AI applications in clinical settings, 33 studies examined the effect of AI applications on clinician outcomes, 14 studies examined the effect on patient outcomes, and one study examined the economic impact associated with AI implementation.

Conclusions: This review indicates that research on the clinical implementation of AI applications is still at an early stage despite the great potential. More research needs to assess the benefits and challenges associated with clinical AI applications through a more rigorous methodology.

(J Med Internet Res 2021;23(4):e25759) doi:10.2196/25759

KEYWORDS

artificial intelligence; machine learning; deep learning; system implementation; clinical practice; review

Introduction

Background

Artificial intelligence (AI) has greatly expanded in health care in the past decade. In particular, AI applications have been applied to uncover information from clinical data and assist health care providers in a wide range of clinical tasks, such as disease diagnosis, triage or screening, risk analysis, and surgical operations [1-4]. According to Accenture analysis, the global health AI market is expected to reach US $6.6 billion by 2021

and has the potential to grow more than 10 times in the next 5 years [5].

The term "AI" was coined by McCarthy in the 1950s and refers to a branch of computer science wherein algorithms are developed to emulate human cognitive functions, such as learning, reasoning, and problem solving [6]. It is a broadly encompassing term that includes, but is not limited to, machine learning (ML), deep learning (DL), natural language processing (NLP), and computer vision (CV).

Researchers have devoted a great deal of effort to the development of health care AI applications. The number of related articles in the Google Scholar database has grown exponentially since 2000. However, their implementation in real-life clinical practice is not widespread [1,7]. Several reasons may account for this research-practice gap. Specifically, AI algorithms may be subject to technical issues, such as data set shift, overfitting, bias, and lack of generalizability [8], limiting the safe translation of AI research into clinical practice. Further, practical implementation of AI applications can be incredibly challenging. Examples of key challenges that need to be addressed include data sharing and privacy issues, lack of algorithm transparency, the changing nature of health care work, financial concerns, and the demanding regulatory environment [1,3,9-13]. However, the huge potential of health care AI applications can only be realized when they have been integrated into clinical routine workflows.

Research Gap

To the best of our knowledge, this review is the first to systematically examine the role of AI applications in real-life clinical environments. We note that many reviews have been carried out in the area of health care AI. One stream of reviews provided an overview of the current status of AI technology in specific clinical domains, such as breast cancer diagnosis [14], melanoma diagnosis [15], pulmonary tuberculosis diagnosis [16], stroke diagnosis and prediction [17], and diabetes management [18]. Another stream of reviews focused on comparing clinician performance and AI performance to provide the evidence base needed for AI implementation [19-21]. In contrast, our work differs from previous reviews in at least three aspects. First, we review clinical AI applications that provide decision support more broadly and hence do not restrict our scope to a specific clinical domain. Second, we focus on studies that reported the evaluation of clinical AI applications in the real world. We hence exclude studies that discussed the development and validation of clinical AI algorithms without actual implementation. Finally, we report a wide range of evaluation outcomes associated with AI implementation, such as performance comparison, clinician and patient outcomes, and economic impact.

On the other hand, we note that several viewpoint articles have provided a general outlook of health care AI [1-3,7,9,22]. These articles mainly provided insights into the current status of health care AI and selected a few clinical AI applications as illustrative examples. They might have also discussed the challenges associated with the practical implementation of AI. However, these articles did not discuss the progress of AI implementation that had been made in detail. In contrast, our work aims to provide a comprehensive map of the literature on the evaluation of AI applications in real-life clinical settings. By doing so, we summarize empirical evidence of the benefits and challenges associated with AI implementation and provide suggestions for future research in this important and promising area.

Objective

The objective of this systematic review was to identify and summarize the existing research on AI applications that have been implemented in real-life clinical practice. This helps us better understand the benefits and challenges associated with AI implementation in routine care settings, such as augmenting clinical decision-making capacity, improving care processes and patient outcomes, and reducing health care costs. Specifically, we synthesize relevant studies based on (1) study characteristics, (2) AI application characteristics, and (3) evaluation outcomes and key findings. Considering the research-practice gap, we also provide suggestions for future research that examines and assesses the implementation of AI in clinical practice.

Methods

Search Strategies

The systematic review was conducted following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines [23]. We searched PubMed, Embase, Cochrane Central, and CINAHL in June 2020 to identify relevant articles on AI applications that had been implemented in clinical practice. We limited our search to English-written peer-reviewed journal articles published between January 2010 and May 2020. We chose 2010 as the start period because health care AI research has since taken off.

We used two groups of keywords to identify terms in the titles, abstracts, and keywords of the publications. The first group of keywords had AI-related terms, including "artificial intelligence," "machine learning," and "deep learning." It is worth noting here that AI is a broadly encompassing term and also includes specific AI techniques, such as neural networks, support vector machines, decision trees, and NLP. However, studies using these techniques are highly likely to use "artificial intelligence" or "machine learning" in abstracts or keywords [24]. The second group of keywords had terms related to clinical implementation, including "clinical," "health," "healthcare," "medical," "implement," "implementation," "deploy," "deployment," and "adoption." Details of the search strategy can be found in Multimedia Appendix 1.

Eligibility Criteria

We downloaded and imported all of the identified articles using EndNote X9 (Thomson Reuters) for citation management. After removing duplicates, two researchers (JY and KYN) independently screened the titles and abstracts of the identified articles to determine their eligibility. Disagreements were resolved by discussion between the authors until consensus was reached. The inclusion criteria were as follows: (1) the study implemented an AI application with patients or health care providers in a real-life clinical setting and (2) the AI application provided decision support by emulating clinical decision-making

processes of health care providers (eg, medical image interpretation and clinical risk assessment). Medical hardware devices, such as X-ray machines, ultrasound machines, surgery robots, and rehabilitation robots, were outside our scope.

The exclusion criteria were as follows: (1) the study discussed the development and validation of clinical AI algorithms without actual implementation; (2) the AI application provided automation (eg, automated insulin delivery and monitoring) rather than decision support; and (3) the AI application targeted nonclinical tasks, such as biomedical research, operational tasks, and epidemiological tasks. We also excluded conference abstracts, reviews, commentaries, simulation papers, and ongoing studies.

Data Extraction and Charting

Following article selection, we created a data-charting form to extract information from the included articles in the following aspects: (1) study characteristics, (2) AI application characteristics, and (3) evaluation outcomes and key findings (Textbox 1).

Textbox 1. Components of the data-charting form.

Study characteristics

- Author, year
- Study design
- Involved patient(s) and health care provider(s)
- Involved hospital(s) and country of the study

Artificial intelligence (AI) application characteristics

- Application description
- AI techniques used (eg, neural networks, random forests, and natural language processing)
- Targeted clinical tasks
- Targeted disease domains and conditions

Evaluation outcomes and key findings

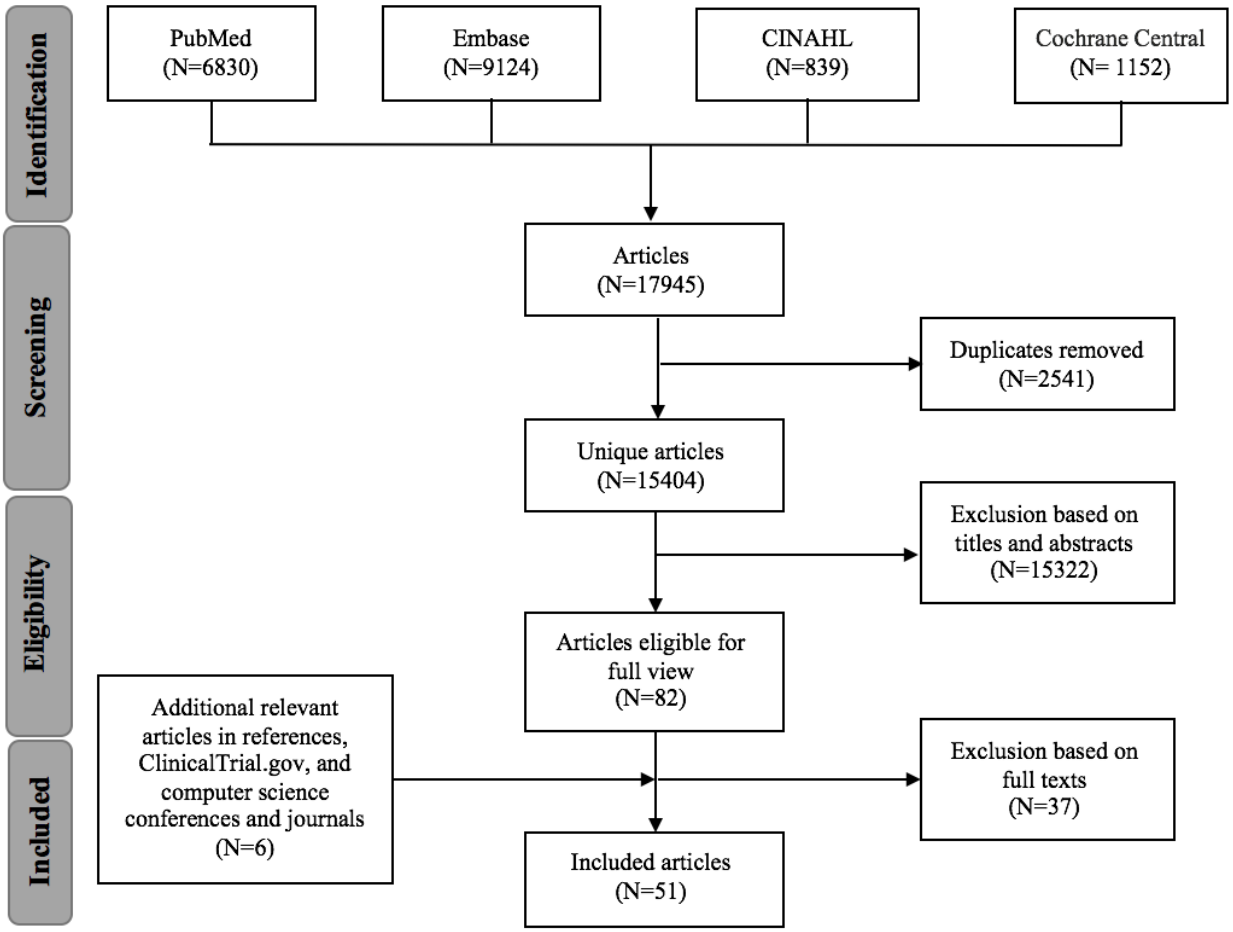
- Performance of AI applications
- Clinician outcomes
- Patient outcomes
- Cost-effectiveness

Results

Overview

Our initial search in June 2020 returned a total of 17,945 journal articles (6830 from PubMed, 9124 from Embase, 839 from CINAHL, and 1152 from Cochrane Central) (Figure 1). We first identified and excluded 2541 duplicates. After that, we excluded 15,322 articles after screening the titles and abstracts. Thus, 82 articles remained for full-text review, of which 45 were included in this review. Additionally, we identified six relevant articles by examining the references of the included articles, browsing through ClinicalTrial.gov using AI-related keywords, and hand searching premier computer science journals and conferences in AI (Multimedia Appendix 1). Finally, a total of 51 articles met our inclusion criteria.

Figure 1. Flow diagram of the literature search based on the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement.



Study Characteristics

Table 1 summarizes the authors, year of publication, study design, involved patients and health care providers, and involved hospitals [25-75]. As shown in Figure 2, there was a rising trend in the number of included studies in the last decade, with a recent peak in 2019, suggesting accelerated research activity in this area.

Regarding study design, the 51 studies included 20 observational studies (17 prospective studies and three retrospective studies), 13 randomized controlled trials (RCTs), eight experimental studies, four before-and-after studies, three surveys, one randomized crossover trial, one nonrandomized trial, and one structured interview. It is important to note that observational studies can be categorized into prospective and retrospective studies based on the timing of data collection. In prospective studies, researchers design the research and plan the data collection procedures before any of the subjects have the disease or develop other outcomes of interest. In retrospective studies, researchers collect existing data on current and past subjects, that is, subjects may have the disease or develop other outcomes of interest before researchers initiate research design and data collection.

Of the 51 studies, 29 (57%) explicitly mentioned the involved patients, two of which had a sample size smaller than 30. On the other hand, 28 (55%) studies provided information about the involved health care providers, of which 17 studies had 10 or fewer providers.

Additionally, 46 (90%) studies mentioned the involved hospitals or clinics (Figure 3). Of these, 36 studies were conducted in developed countries, with 20 conducted in the United States, five in the United Kingdom, two each in Australia, Canada, and Japan, one each in Germany, Israel, Spain, and the Netherlands, and one in the United States and South Korea. On the contrary, 10 studies were conducted in developing countries, with eight conducted in China, one in India, and one in India and Kenya.

Table 1. Characteristics of the included studies.

Author, year	Study design	Sample characteristics	Hospital (country)	Evaluation outcomes
Abràmoff et al, 2018 [25]	Observational study (prospective)	819 patients	10 primary care clinics (United States)	AP[a] (sensitivity, specificity, imageability rate)
Aoki et al, 2020 [26]	Experimental study (crossover design)	6 physicians	The University of Tokyo Hospital (Japan)	CO[b] (reading time, mucosal break detection rate)
Arbabshirani et al, 2018 [27]	Observational study (prospective)	347 routine head CT[c] scans of patients	Geisinger Health System (United States)	AP (AUC[d], accuracy, sensitivity, specificity) CO (time to diagnosis)
Bailey et al, 2013 [28]	Crossover RCT[e]	20,031 patients	Barnes-Jewish Hospital (United States)	PO[f] (ICU[g] transfer, hospital mortality, hospital LOS[h])
Barinov et al, 2019 [29]	Experiment (within subjects)	3 radiologists	NR[i]	AP (AUC) CO (diagnostic accuracy)
Beaudoin et al, 2016 [30]	Observational study (prospective)	350 patients (515 prescriptions)	Centre hospitalier universitaire de Sherbrooke (Canada)	AP (number of triggered recommendations, precision, recall, accuracy)
Bien et al, 2018 [31]	Experimental study (within subjects)	9 clinical experts	Stanford University Medical Center (United States)	AP (AUC) CO (specificity, sensitivity, accuracy)
Brennan et al, 2019 [32]	Nonrandomized trial	20 physicians	An academic quaternary care institution (United States)	AP (AUC) CO (risk assessment changes, AUC, usability)
Chen et al, 2020 [33]	RCT	437 patients	Renmin Hospital, Wuhan University (China)	CO (blind spot rate)
Connell et al, 2019 [34]	Before-after study	2642 patients	Royal Free Hospital, Barnet General Hospital (United Kingdom)	PO (renal recovery rate, other clinical outcomes, care process)
Eshel et al, 2017 [35]	Observational study (prospective)	6 expert microscopists	Apollo Hospital, Chennai (India); Aga Khan University Hospital (Kenya)	AP (sensitivity, specificity, species identification accuracy, device parasite count)
Giannini et al, 2019 [36]	Before-after study	22,280 patients in the silent period, 32,184 patients in the alert period	3 urban acute hospitals under University of Pennsylvania Health System (United States)	AP (sensitivity, specificity) PO (mortality, discharge disposition, ICU transfer, time to ICU transfer, clinical process measures)
Ginestra et al, 2019 [37]	Survey	43 nurses and 44 health care providers	A tertiary teaching hospital in Philadelphia (United States)	CO (nurse and provider perceptions)
Gómez-Vallejo et al, 2016 [38]	Observational study (retrospective)	1800 patients (2569 samples)	A Spanish National Health System hospital (Spain)	AP (accuracy) CO (system perceptions)
Grunwald et al, 2016 [39]	Observational study (retrospective)	15 patients, 3 neuroradiologists	A comprehensive stroke center (Germany)	AP (e-ASPECTS performance)
Kanagasingam et al, 2018 [40]	Observational study (prospective)	193 patients, 4 physicians	A primary care practice in Midland (Australia)	AP (sensitivity, specificity, PPV[j], NPV[k])
Keel et al, 2018 [41]	Survey	96 patients	St Vincent's Hospital, University Hospital Geelong (Australia)	AP (sensitivity and specificity, assessment time) PO (patient acceptability)
Kiani et al, 2020 [42]	Experimental study (within subjects)	11 pathologists	Stanford University Medical Center (United Kingdom)	AP (accuracy) CO (accuracy)
Lagani et al, 2015 [43]	Observational study (prospective)	2 health care providers	Chorleywood Health Centre (United Kingdom)	AP (system performance) CO (usability)

Author, year	Study design	Sample characteristics	Hospital (country)	Evaluation outcomes
Lin et al, 2019 [44]	RCT	350 patients	5 ophthalmic clinics (China)	AP (accuracy, PPV, NPV) CO (time to diagnosis) PO (patient satisfaction)
Lindsey et al, 2018 [45]	Experimental study (within subjects)	40 practicing emergency clinicians	Hospital for Special Surgery (United States)	AP (AUC) CO (sensitivity, specificity, misinterpretation rate)
Liu et al, 2020 [46]	RCT	1026 patients	No. 988 Hospital of Joint Logistic Support Force of PLA (China)	CO (ADR[l], PDR[m], number of detected adenomas and polyps)
Mango et al, 2020 [47]	Experimental study (within subjects)	15 physicians	13 different medical centers (United States)	AP (AUC, sensitivity, specificity) CO (AUC, interreliability, intrareliability)
Martin et al, 2012 [48]	Observational study (prospective)	214 patients	13 different medical centers (United States)	AP (sensitivity, PPV) PO (ACSC[n], care-supported activities)
McCoy and Das, 2017 [49]	Before-after study	1328 patients	Cape Regional Medical Center (United States)	PO (hospital mortality, hospital LOS, readmission rate)
McNamara et al, 2019 [50]	Observational study (prospective)	3 breast cancer experts	John Theurer Cancer Center (United States)	CO (decision making)
Mori et al, 2018 [51]	Observational study (prospective)	791 patients, 23 endoscopists	Showa University Northern Yokohama Hospital (Japan)	AP (NPV) CO (time to diagnosis)
Nagaratnam et al, 2020 [52]	Observational study (retrospective)	1 patient	Royal Berkshire Hospital (United Kingdom)	PO (patient care and clinical outcomes)
Natarajan et al, 2019 [53]	Observational study (prospective)	213 patients	Dispensaries under Municipal Corporation of Greater Mumbai (India)	AP (sensitivity, specificity)
Nicolae et al, 2020 [54]	RCT	41 patients	Sunnybrook Odette Cancer Centre (Canada)	AP (day 30 dosimetry) CO (planning time)
Park et al, 2019 [55]	Experimental study (within subjects)	8 clinicians	Stanford University Medical Center (United States)	CO (specificity, sensitivity, accuracy interrater agreement, time to diagnosis)
Romero-Brufau et al, 2020 [56]	Pre-post survey	81 clinical staff	3 primary-care clinics in Southwest Wisconsin (United States)	CO (attitudes about AI[o] in the workplace)
Rostill et al, 2018 [57]	RCT	204 patients, 204 caregivers	NHS, Surrey and Hampshire (United Kingdom)	CO (system evaluations) PO (early clinical interventions, patient evaluations)
Segal et al, 2014 [58]	Observational study (prospective)	16 pediatric neurologists	Boston Children's Hospital (United States)	CO (diagnostic errors, diagnosis relevance, number of workup items)
Segal et al, 2016 [59]	Observational study (prospective)	26 clinicians	Boston Children's Hospital (United States)	CO (diagnostic errors)
Segal et al, 2017 [60]	Structured interviews	10 medical specialists	Geisinger Health System and Intermountain Healthcare (United States)	CO (system perceptions)
Segal et al, 2019 [61]	Observational study (prospective)	3160 patients (315 prescription alerts)	Sheba Medical Center (Israel)	AP (accuracy, clinical validity, and usefulness) PO (changes in medical orders)
Shimabukuro et al, 2017 [62]	RCT	142 patients	University of California San Francisco Medical Center (United States)	PO (LOS, in-hospital mortality)

Author, year	Study design	Sample characteristics	Hospital (country)	Evaluation outcomes
Sim et al, 2020 [63]	Observational study (prospective)	12 radiologists	4 medical centers (United States and South Korea)	AP (sensitivity, FPPI[p]) CO (sensitivity, FPPI, decision change)
Steiner et al, 2018 [64]	Experimental study (within subjects)	6 anatomic pathologists	NR	CO (sensitivity, average review per image, interpretation difficulty)
Su et al, 2020 [65]	RCT	623 patients, 6 endoscopists	Qilu Hospital of Shandong University (China)	CO (ADR, PDR, number of adenomas and polyps, withdrawal time, adequate bowel preparation rate)
Titano et al, 2018 [66]	RCT	2 radiologists	NR	CO (time to diagnosis, queue of urgent cases)
Vandenberghe et al, 2017 [67]	Observational study (prospective)	1 pathologist and 2 HER2 raters	NR	CO (decision concordance, decision modification)
Voerman et al, 2019 [68]	Before-after study	NR	Five Rivers Medical Center, Pocahontas (United States)	CE[q] (average total costs per patient) PO (numbers of patients with clostridium difficile and antibiotic-resistant infections, LOS, antibiotic use)
Wang et al, 2019 [69]	RCT	1058 patients, 8 physicians	Sichuan Provincial People's Hospital (China)	CO (ADR, PDR, number of adenomas per patient)
Wang et al, 2019 [70]	RCT	75 patients	4 primary care clinics affiliated with Brigham and Women's Hospital (United States)	CO (anticoagulation prescriptions)
Wang et al, 2020 [71]	RCT	962 patients	Caotang branch hospital of Sichuan Provincial People's Hospital (China)	CO (ADR, PDR, number of adenomas and polyps per colonoscopy)
Wijnberge et al, 2020 [72]	RCT	68 patients	Amsterdam UMC (Netherlands)	PO (median time-weighted average of hypotension, median time of hypotension, treatment, time to intervention, adverse events)
Wu et al, 2019 [73]	Observational study (prospective)	3600 residents	3 ophthalmologists, community healthcare centers (China)	AP (AUC) CO (ophthalmologist-to-population service ratio)
Wu et al, 2019 [74]	RCT	303 patients, 6 endoscopists	Renmin hospital of Wuhan University (China)	AP (accuracy, completeness of photo documentation) CO (blind spot rate, number of ignored patients, inspection time) PO (adverse events)

Author, year	Study design	Sample characteristics	Hospital (country)	Evaluation outcomes
Yoo et al, 2018 [75]	Observational study (prospective)	50 patients, 1 radiologist	NR (Korea)	AP (sensitivity, specificity, PPV, NPV, accuracy) CO (sensitivity, specificity, PPV, NPV, accuracy)

[a]AP: application performance.
[b]CO: clinician outcomes.
[c]CT: computed tomography.
[d]AUC: area under the curve.
[e]RCT: randomized controlled trial.
[f]PO: patient outcomes.
[g]ICU: intensive care unit.
[h]LOS: length of stay.
[i]NR: not reported.
[j]PPV: positive-predictive value.
[k]NPV: negative-predictive value.
[l]ADR: adenoma detection rate.
[m]PDR: polyp detection rate.
[n]ACSC: ambulatory care sensitive admissions.
[o]AI: artificial intelligence.
[p]FFPI: false-positive per image.
[q]CE: cost-effectiveness.

Figure 2. Distribution of the included articles from 2010 to 2020.

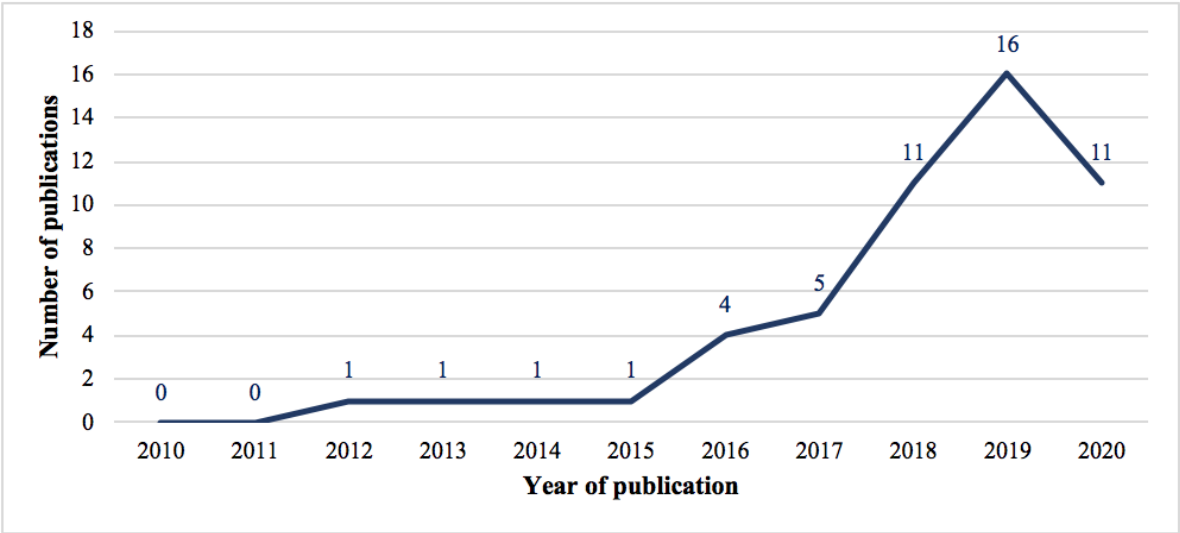
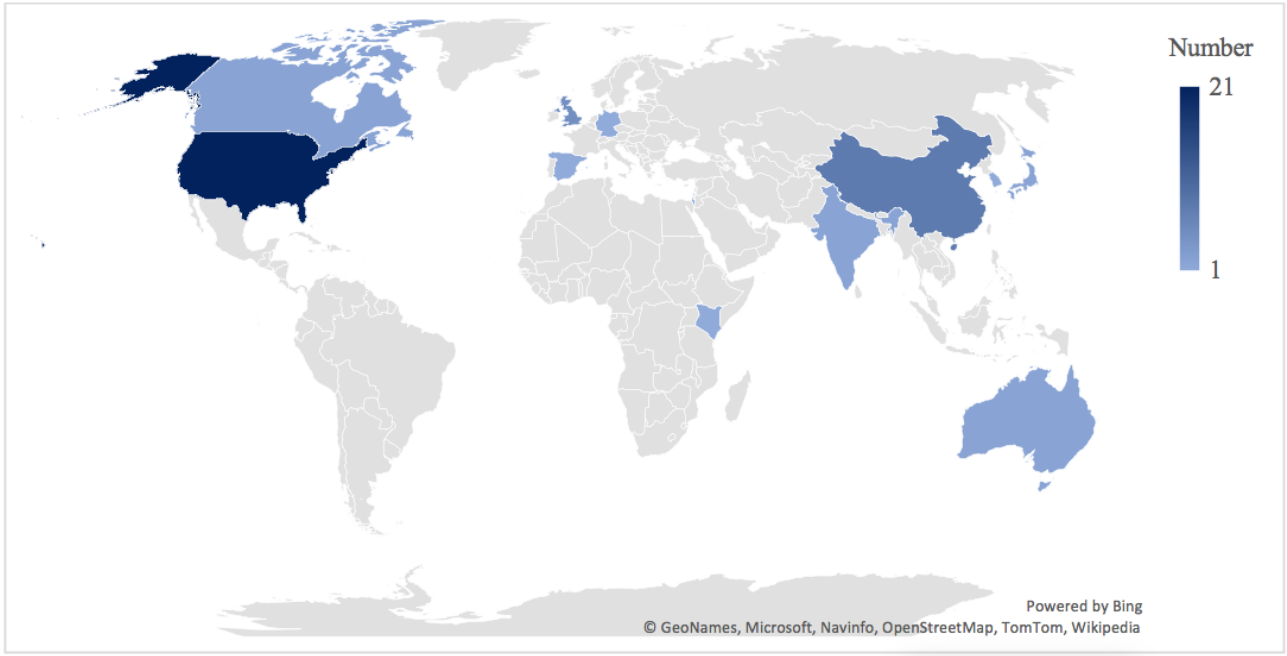


Figure 3. Country distribution of the involved hospitals.



Quality Assessment

Considering the heterogeneity of study types included in the review, we only assessed the risk of bias of 13 RCTs using the Cochrane Collaboration Risk of Bias tool (Multimedia Appendix 2) [76]. Overall, the total score of the RCTs ranged from 0 [57] to 6 [44,65,69], with a mean value of 3.84. Specifically, eight studies reported random sequence generation [33,44,62,65,69,71,72], and three studies explicitly stated that the allocation was concealed [62,65,72]. Only two studies were double blinded [66,71]. Blinding of participants was unsuccessful in two studies [62,72] and was unclear in six studies [44,46,54,57,69,70]. Blinding of outcome assessment was unsuccessful in seven studies [33,46,62,65,69,70,74] and was unclear in one study [57]. Three studies did not clearly state whether they had complete data for the enrolled participants [57,66,77]. All of the 13 studies had a low risk of selective reporting bias. Other potential sources of bias included a small sample size [62,70,72], a short study period [62], and a lack of detailed information regarding RCTs and follow-ups [57,66].

AI Application Characteristics

Among the 51 studies, two did not disclose any information regarding the AI techniques used. Among the remaining 49 studies, the most popular ML technique was neural networks (n=22), followed by random forests (n=3), Bayesian pattern matching (n=3), support vector machine (n=2), decision tree (n=2), and deep reinforcement learning (n=2). We also found that the included AI applications mainly provided decision support in the following four categories of clinical tasks: disease screening or triage (n=16), disease diagnosis (n=16), risk analysis (n=14), and treatment (n=7). Further, AI applications in 46 (94%) studies targeted one or more specific diseases and conditions. The most prevalent diseases and conditions were sepsis (n=6), breast cancer (n=5), diabetic retinopathy (n=4), polyp and adenoma (n=4), cataracts (n=2), and stroke (n=2). Details of AI application characteristics are provided in Multimedia Appendix 3.

Evaluation Outcomes

We categorized the evaluation outcomes in our review studies into the following four types: performance of AI applications, clinician outcomes, patient outcomes, and cost-effectiveness, as can be seen in Table 1 and Multimedia Appendix 4.

Performance of AI Applications

Twenty-six studies evaluated the performance of AI applications in real-life clinical settings [25,27,29-32,35,36,38-43,45,47,48,51,52,54,61,63,73-75,78]. Commonly used performance metrics included accuracy, area under the curve (AUC), specificity, sensitivity, positive-predictive value (PPV), and negative-predictive value (NPV). Of these, 24 studies reported acceptable and satisfactory performance of AI applications in practice. For example, one study [25] conducted a pivotal trial of the IDx-DR diagnostic system (IDx, LLC) to detect diabetic retinopathy in 10 primary clinic offices in the United States. They reported that IDx-DR had a sensitivity of 87.2%, a specificity of 90.7%, and an imageability rate of 96.1%, exceeding prespecified endpoints. Based on the results, IDx-DR became the first Food and Drug Administration (FDA)–authorized AI diagnostic system, with the potential to improve early detection of diabetic retinopathy and prevent vision loss in thousands of patients with diabetes.

On the contrary, two studies found that AI applications failed to outperform health care providers and needed further improvement [40,44]. In particular, one RCT [44] examined the performance of CC-Cruiser, an AI-based platform for childhood cataracts, in five ophthalmic clinics in China. The authors found that CC-Cruiser had considerably lower accuracy, PPV, and NPV than senior consultants in diagnosing childhood congenital cataracts and making treatment decisions. Another study [40] evaluated the performance of an AI-based diabetic retinopathy grading system in a primary care office in Australia

and found that the AI system had a high false-positive rate with a PPV of 12%. Specifically, of the 193 patients who consented to the study, the AI system identified 17 patients with severe diabetic retinopathy that required referral. However, only two patients were correctly identified, and the remaining 15 patients were false positives.

Clinician Outcomes

Thirty-three studies examined the effect of AI applications on clinician outcomes, that is, clinician decision making, clinician workflow and efficiency, and clinician evaluations and acceptance of AI applications [26,27,29,31-33,37,38,42-47,50,51,54-60,64,65,67,69-71,73-75].

AI applications have the potential to provide clinical decision support. From our review, 16 studies demonstrated that AI applications could enhance clinical decision-making capacity [31-33,45-47,50,55,58,59,63,64,67,69,71,74,75]. For example, Brennan et al [32] found that clinicians gained knowledge after interacting with MySurgery, an algorithm for preoperative risk assessments, and improved their risk assessment performance as a result. On the contrary, two studies did not find any evidence for enhanced decision-making [26,42]. One possible explanation is that AI may provide misleading recommendations, offsetting the benefits of AI. Specifically, Kiani et al [42] evaluated the effect of a DL-based system for live cancer classification on the diagnostic performance of 11 pathologists and found that AI use did not greatly improve the diagnostic accuracy. They further noted that AI improved accuracy when it provided correct predictions and harmed accuracy when it provided wrong predictions. Aoki et al [26] examined the impact of a DL-based system for mucosal break detection on endoscopists in reading small bowel capsule endoscopy. They found that the system failed to improve the mucosal break detection performance of endoscopists, particularly trainees.

Seven studies were aimed at clinician workflow and efficiency [26,27,44,51,54,66,73]. Of these, six studies found that AI accelerated the time needed for clinical tasks and improved the existing workflow [26,27,44,51,54,66]. For example, Titano et al [66] found that a DL-based cranial image triage algorithm processed and interpreted images 150 times faster than human radiologists (1.2 seconds vs 177 seconds) and appropriately escalated urgent cases, enhancing the triage of cases in the radiology workflow. The only exception is the work of Wu et al [74], which assessed the quality improvement system WISENSE for blind-spot monitoring and procedure timing during esophagogastroduodenoscopy. This study found that WISENSE helped endoscopists monitor and control their time on each procedure and increased inspection time as a result.

Finally, clinician perceptions and acceptance of AI applications were examined in seven studies [32,37,38,43,56,57,60]. Particularly, five out of the seven studies reported overall positive perceptions of AI applications [32,38,43,57,60]. For example, Brennan et al [32] asked 20 surgical intensivists to use and evaluate MySurgeryRisk for preoperative risk prediction in a simulated clinical workflow. Most respondents indicated that MySurgeryRisk was useful and easy to use and believed that it would be helpful for decision making. On the other hand, the remaining two studies reported mixed or even negative evaluations of AI [37,56]. Specifically, Ginestra et al [37] assessed physician evaluations of an ML-based sepsis prediction system in a tertiary teaching hospital and found that only 16% of health care providers perceived system-generated sepsis alerts to be helpful. The negative evaluations could be attributed to providers' low confidence in alerts, low algorithm transparency, and a lack of established actions after alerts. Romero-Brufau et al [56] reported survey results from implementing an AI-based clinical decision support system in a regional health system practice and found that only 14% of clinical staff were willing to recommend the system. Staff feedback revealed that some system-recommended interventions were inadequate and inappropriate.

Patient Outcomes

Fourteen studies reported patient outcomes [28,34,36,41,44,48,49,52,57,61,62,68,72,74]. In 11 of the 14 studies, researchers examined the effect of AI on clinical processes and outcomes, such as hospital length of stay, in-hospital mortality, intensive care unit (ICU) transfer, readmission, and time to intervention [28,34,36, 48,49,52,57,61,62,68,72,74]. The results were inconsistent. Most studies reported improved clinical outcomes (n=8) [36,48,52,57,61,62,68,72,74]. For example, one RCT [62] implemented and assessed an ML-based severe sepsis prediction algorithm (Dascena) in two ICUs at the University of California San Francisco Medical Center. They found that the algorithm implementation greatly decreased the hospital length of stay from 13.0 days to 10.3 days and decreased the in-hospital mortality rate from 21.3% to 8.96%. However, three of the studies did not find evidence for improved clinical outcomes, indicating the limited applicability of the algorithms in their current form [28,34,36]. In particular, Bailey et al [28] examined the effect of an ML-based algorithm that generated real-time alerts for clinical deterioration in hospitalized patients. They found that providing alerts alone could not reduce the hospital length of stay and the in-hospital mortality. Connell et al [34] examined the effect of a novel digitally enabled care pathway for acute kidney injury management and found no step changes in the renal recovery rate and other secondary clinical outcomes following the intervention. Giannini et al [36] developed and implemented a sepsis prediction algorithm in a tertiary teaching hospital system. The results showed that the algorithm-generated alerts had a limited impact on clinical processes and could not reduce mortality, discharge dispositions, or transfer to the ICU. Future algorithm optimization is thus needed.

Three studies examined how patients evaluated AI applications, and all of them reported positive results [41,44,57]. Keel et al [41] evaluated patient acceptability of an AI-based diabetic retinopathy screening tool in an endocrinology outpatient setting. They found that 96% (92/96) of the screened patients were satisfied with the AI tool and 78% (43/55) of the patients in the follow-up survey preferred AI screening over manual screening, suggesting that the AI tool was well-accepted by patients. Lin et al [44] assessed patient satisfaction with CC-Cruiser for childhood cataracts and found that patients were slightly more satisfied with CC-Cruiser in comparison with senior consultants. One explanation is that childhood cataracts may cause

irreversible vision impairment and even blindness without early intervention. Therefore, parents of patients appreciated the faster diagnosis of CC-Cruiser. Rostill et al [57] assessed an Internet of Things (IoT) system for dementia care and found that dementia patients trusted the system and would like to recommend it.

Cost-Effectiveness

The economic impact of AI implementation in clinical practice was addressed in only one study [68]. This study reported that the implementation of an ML-based system for antibiotic stewardship reduced costs by US $25,611 for sepsis and US $3630 for lower respiratory tract infections compared with usual care.

Discussion

Principal Findings

AI applications have huge potential to augment clinician decision making, improve clinical care processes and patient outcomes, and reduce health care costs. Our review seeks to identify and summarize the existing studies on AI applications that have been implemented in real-life clinical practice. It yields the following interesting findings.

First, we note that the number of included studies was surprisingly small considering the tremendous number of studies on health care AI. In particular, most of the health care AI studies were proof-of-concept studies that focused on AI algorithm development and validation using retrospective clinical data sets. In contrast, only a handful of studies implemented and evaluated AI in a clinical environment. To ensure safe adoption, however, an AI application should provide solid scientific evidence for its effectiveness relative to the standard of care. Therefore, we urge the health care AI research community to work closely with health care providers and institutions to demonstrate the potential of AI in real-life clinical settings.

Second, more than two-thirds of the included articles were from developed economies, of which more than half were from the United States, suggesting that developed countries are at the forefront of health care AI development and deployment. This is consistent with the fact that top health AI companies and start-ups (eg, Google Health, IBM Watson Health, and Babylon Health) are mainly located in the United States and Europe. This finding should be interpreted with caution because we excluded non-English–written articles, even though our search had identified 890 non-English publications. We did not include these non-English articles because it is difficult to conduct an unbiased analysis owing to translation difficulty and variation. The imbalanced distribution of articles by country or economic development status could be attributed to the fact that researchers from low-income countries have a very low publication rate.

However, it is worth noting that 8 (16%) of our articles were from China, suggesting that China has been extensively applying health care AI and conducting health care AI research. Indeed, hospitals, technology companies, and the Chinese government have been driving clinical AI deployment with the aim to alleviate doctor shortages, relieve medical resource inequality, and reduce health care costs [79-82], and Chinese researchers have acquired the capability to publish in international English journals.

Third, the quality of research on clinical AI evaluation needs to be improved in the future. Our review revealed that only 13 (26%) studies were RCTs and most of them suffered from moderate to high risk of bias. Eight studies were experimental studies, and all of them adopted a cross-over design or within-subjects design and were hence susceptible to confounding effects. With respect to sample information, only 8 (16%) studies provided information on both patients and health care providers, and 14 (28%) studies used a sample size smaller than 20 (Table 1), limiting the generalizability of their results. Regarding the evaluation design, one-third of the studies (n=17, 33%) did not include a comparison group (Multimedia Appendix 4), limiting the ability to identify the added value of AI applications compared with the current best practice. Given that health care providers may hold different perceptions toward different AI systems of varying performance and reliability, it would be helpful if the studies provide a transparent description of the AI system's architecture, accuracy or reliability performance, and possible risks. Unfortunately, in our review, 21 studies did not provide adequate information about the architecture of the AI applications [25,29,32,34,37,44,46,50,52,54,56-62,65,68,70,75] and 22 studies did not reveal the performance and possible risks of AI under evaluation [26,29,34,37,39,46,48-50,52,54,56-62,64,65,68,69]. Further, considering that some self-evolving adaptive clinical AI applications continuously incorporate the latest clinical practice data and published evidence, it is important to undertake periodic monitoring and recalibration of AI applications to ensure that they are working as expected. Finally, we found that more than half of the studies (n=29, 57%) investigated only one aspect of evaluation outcome (Multimedia Appendix 4). We encourage future research to conduct a more comprehensive assessment of the quality of clinical AI applications as well as their impacts on clinicians, patients, and health care institutions. This will facilitate the comparison and selection of alternative AI solutions in the same clinical domain.

Fourth, our analysis indicated that AI applications could provide effective decision support, albeit in certain contexts. For instance, the augmenting role of AI in clinical decision-making capacity can be affected by the level of expertise. In particular, two studies suggested that junior physicians were more likely to benefit from AI than senior physicians because they had a higher tendency to reconsider and modify their clinical decisions when encountering disconfirming AI suggestions [38,47]. However, it is worth noting that AI can be misleading sometimes. For example, one study from our review speculated that trainee endoscopists may feel confused about false-positive results from an AI screening tool owing to limited reading experience and, as a result, ignore AI-marked lesions of small-bowel mucosal breaks [26]. It is therefore important for future research to examine under what circumstances physicians could benefit more from AI applications. However, we are sanguine that when AI technology is sufficiently mature and

accurate to become the evidence-based best practice, its use would become part of routine clinical care in the future.

With respect to AI acceptance, we observed that health care providers expressed negative feelings toward AI in two studies [37,56], indicating that barriers existed in the incorporation of AI into the routine workflow. However, an elaboration of AI implementation barriers will be lengthy and is beyond the scope of this work, and we refer interested readers to the reports by Kelly et al [8], Ngiam and Khor [83], Lysaght et al [84], Shaw et al [10], and Yu and Kohane [12] for more details.

Fifth, for most of the included studies on patient outcomes, we found that they did not examine the clinical processes and interventions in detail. However, AI applications without appropriate and useful interventions may be ineffective at improving patient outcomes. For example, Bailey et al [28] found that simply notifying the nursing staff of clinical deterioration risks was not able to improve the outcomes of high-risk patients. Effective patient-specific interventions are needed. Therefore, future research may design and evaluate patient-directed interventions to enhance the clinical effectiveness of AI applications.

Moreover, three of the included studies suggested that patients and their families were highly satisfied with health care AI owing to its convenience and efficiency [41,44,57]. However, this may not always be the case. Prior research has shown that patients preferred to receive primary care from a human provider than AI even if the care from the health provider entailed a higher misdiagnosis risk [85]. The reason is that they perceived AI to be less capable in considering their unique circumstances. Additionally, patients may disparage physicians aided by a clinical decision support system and perceive them as less capable and professional than their unaided counterparts [86]. Further studies to explore the possible patient concerns and resistance toward health AI applications should be considered.

Finally, according to an Accenture survey, more than half of health care institutions are optimistic that AI will reduce costs and improve revenue despite the high initial costs associated with AI implementation [87]. However, only one included study documented the economic outcomes of AI implementation. This highlights the need to conduct more cost-effectiveness analyses of AI applications in clinical practice.

Limitations

This review has several limitations. First, we only included peer-reviewed English-written journal articles. It is plausible that some relevant articles were written in other languages or published in conferences, workshops, and news reports. As noted earlier, this may partly explain the imbalanced country distribution of the reviewed articles. Moreover, we did not include articles that were published before 2010 because AI only started to make in-roads in the clinical field in the last decade, as evident in our search results. Moreover, we only reviewed premium computer science conferences and journals without comprehensively examining engineering and computer science databases. This should be less of a concern here because we found that computer science conferences and journals mainly focus on the training and validation of novel AI algorithms without actual deployment. Still, future research can expand the search scope to gain deeper insights into state-of-the-art clinical AI algorithms.

Another concern is that some AI applications may have been implemented in real-world clinical practice without any openly accessible publications. For example, IDx-DR, the first FDA-approved AI system, has been implemented in more than 20 health care institutions such as University of Iowa Health Care [88]. However, our search only identified one related published result [25]. Clinical practitioners should take a more active role in reporting AI evaluation and use results in their daily practice in the future.

Conclusions

AI applications have tremendous potential to improve patient outcomes and improve care processes. Based on the literature presented in this review, there is great interest to develop AI tools to support clinical workflows, with increasing high-quality evidence being generated. However, there is currently insufficient level 1 evidence to advocate the routine use of health care AI for decision support, hindering the growth of health care AI and presenting potential risks to patient safety. We thus conclude that it is important to conduct robust RCTs to benchmark AI-aided care processes and outcomes to the current best practice. A rigorous, robust, and comprehensive evaluation of health care AI will help move from theory to clinical practice.

Acknowledgments

This study was supported by NSCP (grant no. N-171-000-499-001) from the National University of Singapore. It was also supported by Dean Strategic Fund-Health Informatics (grant no. C-251-000-068-001) from the National University of Singapore.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy.
[PDF File (Adobe PDF File), 86 KB - jmir_v23i4e25759_app1.pdf]

Multimedia Appendix 2

Quality assessments of randomized controlled trials based on the Cochrane Collaboration Risk of Bias Tool.
[XLSX File (Microsoft Excel File), 10 KB - jmir_v23i4e25759_app2.xlsx]

Multimedia Appendix 3
Artificial intelligence application characteristics.
[PDF File (Adobe PDF File), 112 KB - jmir_v23i4e25759_app3.pdf]

Multimedia Appendix 4
Evaluation outcomes and main results of the included studies.
[PDF File (Adobe PDF File), 153 KB - jmir_v23i4e25759_app4.pdf]

References

1. He J, Baxter SL, Xu J, Xu J, Zhou X, Zhang K. The practical implementation of artificial intelligence technologies in medicine. Nat Med 2019 Jan;25(1):30-36 [FREE Full text] [doi: 10.1038/s41591-018-0307-0] [Medline: 30617336]
2. Jiang F, Jiang Y, Zhi H, Dong Y, Li H, Ma S, et al. Artificial intelligence in healthcare: past, present and future. Stroke Vasc Neurol 2017 Dec;2(4):230-243 [FREE Full text] [doi: 10.1136/svn-2017-000101] [Medline: 29507784]
3. Yu K, Beam AL, Kohane IS. Artificial intelligence in healthcare. Nat Biomed Eng 2018 Oct;2(10):719-731 [FREE Full text] [doi: 10.1038/s41551-018-0305-z] [Medline: 31015651]
4. Triantafyllidis AK, Tsanas A. Applications of Machine Learning in Real-Life Digital Health Interventions: Review of the Literature. J Med Internet Res 2019 Apr 05;21(4):e12286 [FREE Full text] [doi: 10.2196/12286] [Medline: 30950797]
5. Artificial intelligence (AI): healthcare's new nervous system. Accenture. URL: https://www.accenture.com/us-en/insight-artificial-intelligence-healthcare%C2%A0 [accessed 2021-04-05]
6. McCarthy J, Minsky M, Rochester N, Shannon C. A proposal for the Dartmouth summer research project on artificial intelligence. AI Magazine 2006;27(4):12. [doi: 10.1609/aimag.v27i4.1904]
7. Panch T, Mattie H, Celi LA. The "inconvenient truth" about AI in healthcare. NPJ Digit Med 2019;2:77 [FREE Full text] [doi: 10.1038/s41746-019-0155-4] [Medline: 31453372]
8. Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. BMC Med 2019 Oct 29;17(1):195 [FREE Full text] [doi: 10.1186/s12916-019-1426-2] [Medline: 31665002]
9. Shah P, Kendall F, Khozin S, Goosen R, Hu J, Laramie J, et al. Artificial intelligence and machine learning in clinical development: a translational perspective. NPJ Digit Med 2019;2:69 [FREE Full text] [doi: 10.1038/s41746-019-0148-3] [Medline: 31372505]
10. Shaw J, Rudzicz F, Jamieson T, Goldfarb A. Artificial Intelligence and the Implementation Challenge. J Med Internet Res 2019 Jul 10;21(7):e13659 [FREE Full text] [doi: 10.2196/13659] [Medline: 31293245]
11. Cohen IG, Amarasingham R, Shah A, Xie B, Lo B. The legal and ethical concerns that arise from using complex predictive analytics in health care. Health Aff (Millwood) 2014 Jul;33(7):1139-1147 [FREE Full text] [doi: 10.1377/hlthaff.2014.0048] [Medline: 25006139]
12. Yu K, Kohane IS. Framing the challenges of artificial intelligence in medicine. BMJ Qual Saf 2019 Mar;28(3):238-241. [doi: 10.1136/bmjqs-2018-008551] [Medline: 30291179]
13. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med 2019 Jan;25(1):44-56. [doi: 10.1038/s41591-018-0300-7] [Medline: 30617339]
14. Gardezi SJS, Elazab A, Lei B, Wang T. Breast Cancer Detection and Diagnosis Using Mammographic Data: Systematic Review. J Med Internet Res 2019 Jul 26;21(7):e14464 [FREE Full text] [doi: 10.2196/14464] [Medline: 31350843]
15. Rajpara SM, Botello AP, Townend J, Ormerod AD. Systematic review of dermoscopy and digital dermoscopy/ artificial intelligence for the diagnosis of melanoma. Br J Dermatol 2009 Sep;161(3):591-604. [doi: 10.1111/j.1365-2133.2009.09093.x] [Medline: 19302072]
16. Harris M, Qi A, Jeagal L, Torabi N, Menzies D, Korobitsyn A, et al. A systematic review of the diagnostic accuracy of artificial intelligence-based computer programs to analyze chest x-rays for pulmonary tuberculosis. PLoS One 2019;14(9):e0221339 [FREE Full text] [doi: 10.1371/journal.pone.0221339] [Medline: 31479448]
17. Wang W, Kiik M, Peek N, Curcin V, Marshall IJ, Rudd AG, et al. A systematic review of machine learning models for predicting outcomes of stroke with structured data. PLoS One 2020;15(6):e0234722 [FREE Full text] [doi: 10.1371/journal.pone.0234722] [Medline: 32530947]
18. Contreras I, Vehi J. Artificial Intelligence for Diabetes Management and Decision Support: Literature Review. J Med Internet Res 2018 May 30;20(5):e10775 [FREE Full text] [doi: 10.2196/10775] [Medline: 29848472]
19. Nagendran M, Chen Y, Lovejoy CA, Gordon AC, Komorowski M, Harvey H, et al. Artificial intelligence versus clinicians: systematic review of design, reporting standards, and claims of deep learning studies. BMJ 2020 Mar 25;368:m689 [FREE Full text] [doi: 10.1136/bmj.m689] [Medline: 32213531]
20. Shen J, Zhang CJP, Jiang B, Chen J, Song J, Liu Z, et al. Artificial Intelligence Versus Clinicians in Disease Diagnosis: Systematic Review. JMIR Med Inform 2019 Aug 16;7(3):e10010 [FREE Full text] [doi: 10.2196/10010] [Medline: 31420959]

21. Liu X, Faes L, Kale AU, Wagner SK, Fu DJ, Bruynseels A, et al. A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. The Lancet Digital Health 2019 Oct;1(6):e271-e297 [FREE Full text] [doi: 10.1016/S2589-7500(19)30123-2] [Medline: 33323251]
22. Quer G, Muse ED, Nikzad N, Topol EJ, Steinhubl SR. Augmenting diagnostic vision with AI. The Lancet 2017 Jul;390(10091):221 [FREE Full text] [doi: 10.1016/S0140-6736(17)31764-6] [Medline: 30078917]
23. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
24. Wolff J, Pauling J, Keck A, Baumbach J. The Economic Impact of Artificial Intelligence in Health Care: Systematic Review. J Med Internet Res 2020 Feb 20;22(2):e16866 [FREE Full text] [doi: 10.2196/16866] [Medline: 32130134]
25. Abràmoff MD, Lavin PT, Birch M, Shah N, Folk JC. Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices. NPJ Digit Med 2018;1:39 [FREE Full text] [doi: 10.1038/s41746-018-0040-6] [Medline: 31304320]
26. Aoki T, Yamada A, Aoyama K, Saito H, Fujisawa G, Odawara N, et al. Clinical usefulness of a deep learning-based system as the first screening on small-bowel capsule endoscopy reading. Dig Endosc 2020 May;32(4):585-591. [doi: 10.1111/den.13517] [Medline: 31441972]
27. Arbabshirani MR, Fornwalt BK, Mongelluzzo GJ, Suever JD, Geise BD, Patel AA, et al. Advanced machine learning in action: identification of intracranial hemorrhage on computed tomography scans of the head with clinical workflow integration. NPJ Digit Med 2018;1:9 [FREE Full text] [doi: 10.1038/s41746-017-0015-z] [Medline: 31304294]
28. Bailey TC, Chen Y, Mao Y, Lu C, Hackmann G, Micek ST, et al. A trial of a real-time alert for clinical deterioration in patients hospitalized on general medical wards. J Hosp Med 2013 May;8(5):236-242. [doi: 10.1002/jhm.2009] [Medline: 23440923]
29. Barinov L, Jairaj A, Becker M, Seymour S, Lee E, Schram A, et al. Impact of Data Presentation on Physician Performance Utilizing Artificial Intelligence-Based Computer-Aided Diagnosis and Decision Support Systems. J Digit Imaging 2019 Jun;32(3):408-416 [FREE Full text] [doi: 10.1007/s10278-018-0132-5] [Medline: 30324429]
30. Beaudoin M, Kabanza F, Nault V, Valiquette L. Evaluation of a machine learning capability for a clinical decision support system to enhance antimicrobial stewardship programs. Artif Intell Med 2016 Mar;68:29-36. [doi: 10.1016/j.artmed.2016.02.001] [Medline: 26947174]
31. Bien N, Rajpurkar P, Ball RL, Irvin J, Park A, Jones E, et al. Deep-learning-assisted diagnosis for knee magnetic resonance imaging: Development and retrospective validation of MRNet. PLoS Med 2018 Nov;15(11):e1002699 [FREE Full text] [doi: 10.1371/journal.pmed.1002699] [Medline: 30481176]
32. Brennan M, Puri S, Ozrazgat-Baslanti T, Feng Z, Ruppert M, Hashemighouchani H, et al. Comparing clinical judgment with the MySurgeryRisk algorithm for preoperative risk assessment: A pilot usability study. Surgery 2019 May;165(5):1035-1045 [FREE Full text] [doi: 10.1016/j.surg.2019.01.002] [Medline: 30792011]
33. Chen D, Wu L, Li Y, Zhang J, Liu J, Huang L, et al. Comparing blind spots of unsedated ultrafine, sedated, and unsedated conventional gastroscopy with and without artificial intelligence: a prospective, single-blind, 3-parallel-group, randomized, single-center trial. Gastrointest Endosc 2020 Feb;91(2):332-339.e3. [doi: 10.1016/j.gie.2019.09.016] [Medline: 31541626]
34. Connell A, Montgomery H, Martin P, Nightingale C, Sadeghi-Alavijeh O, King D, et al. Evaluation of a digitally-enabled care pathway for acute kidney injury management in hospital emergency admissions. NPJ Digit Med 2019;2:67 [FREE Full text] [doi: 10.1038/s41746-019-0100-6] [Medline: 31396561]
35. Eshel Y, Houri-Yafin A, Benkuzari H, Lezmy N, Soni M, Charles M, et al. Evaluation of the Parasight Platform for Malaria Diagnosis. J Clin Microbiol 2017 Mar;55(3):768-775 [FREE Full text] [doi: 10.1128/JCM.02155-16] [Medline: 27974542]
36. Giannini HM, Ginestra JC, Chivers C, Draugelis M, Hanish A, Schweickert WD, et al. A Machine Learning Algorithm to Predict Severe Sepsis and Septic Shock: Development, Implementation, and Impact on Clinical Practice. Crit Care Med 2019 Nov;47(11):1485-1492. [doi: 10.1097/CCM.0000000000003891] [Medline: 31389839]
37. Ginestra JC, Giannini HM, Schweickert WD, Meadows L, Lynch MJ, Pavan K, et al. Clinician Perception of a Machine Learning–Based Early Warning System Designed to Predict Severe Sepsis and Septic Shock*. Critical Care Medicine 2019;47(11):1477-1484. [doi: 10.1097/ccm.0000000000003803]
38. Gómez-Vallejo H, Uriel-Latorre B, Sande-Meijide M, Villamarín-Bello B, Pavón R, Fdez-Riverola F, et al. A case-based reasoning system for aiding detection and classification of nosocomial infections. Decision Support Systems 2016 Apr;84:104-116. [doi: 10.1016/j.dss.2016.02.005]
39. Grunwald IQ, Ragoschke-Schumm A, Kettner M, Schwindling L, Roumia S, Helwig S, et al. First Automated Stroke Imaging Evaluation via Electronic Alberta Stroke Program Early CT Score in a Mobile Stroke Unit. Cerebrovasc Dis 2016;42(5-6):332-338. [doi: 10.1159/000446861] [Medline: 27304197]
40. Kanagasingam Y, Xiao D, Vignarajan J, Preetham A, Tay-Kearney M, Mehrotra A. Evaluation of Artificial Intelligence-Based Grading of Diabetic Retinopathy in Primary Care. JAMA Netw Open 2018 Sep 07;1(5):e182665 [FREE Full text] [doi: 10.1001/jamanetworkopen.2018.2665] [Medline: 30646178]

41. Keel S, Lee PY, Scheetz J, Li Z, Kotowicz MA, MacIsaac RJ, et al. Feasibility and patient acceptability of a novel artificial intelligence-based screening model for diabetic retinopathy at endocrinology outpatient services: a pilot study. Sci Rep 2018 Mar 12;8(1):4330 [FREE Full text] [doi: 10.1038/s41598-018-22612-2] [Medline: 29531299]
42. Kiani A, Uyumazturk B, Rajpurkar P, Wang A, Gao R, Jones E, et al. Impact of a deep learning assistant on the histopathologic classification of liver cancer. NPJ Digit Med 2020;3:23 [FREE Full text] [doi: 10.1038/s41746-020-0232-8] [Medline: 32140566]
43. Lagani V, Chiarugi F, Manousos D, Verma V, Fursse J, Marias K, et al. Realization of a service for the long-term risk assessment of diabetes-related complications. J Diabetes Complications 2015 Jul;29(5):691-698. [doi: 10.1016/j.jdiacomp.2015.03.011] [Medline: 25953402]
44. Lin H, Li R, Liu Z, Chen J, Yang Y, Chen H, et al. Diagnostic Efficacy and Therapeutic Decision-making Capacity of an Artificial Intelligence Platform for Childhood Cataracts in Eye Clinics: A Multicentre Randomized Controlled Trial. EClinicalMedicine 2019 Mar;9:52-59 [FREE Full text] [doi: 10.1016/j.eclinm.2019.03.001] [Medline: 31143882]
45. Lindsey R, Daluiski A, Chopra S, Lachapelle A, Mozer M, Sicular S, et al. Deep neural network improves fracture detection by clinicians. Proc Natl Acad Sci U S A 2018 Nov 06;115(45):11591-11596 [FREE Full text] [doi: 10.1073/pnas.1806905115] [Medline: 30348771]
46. Liu W, Zhang Y, Bian X, Wang L, Yang Q, Zhang X, et al. Study on detection rate of polyps and adenomas in artificial-intelligence-aided colonoscopy. Saudi J Gastroenterol 2020;26(1):13-19 [FREE Full text] [doi: 10.4103/sjg.SJG_377_19] [Medline: 31898644]
47. Mango VL, Sun M, Wynn RT, Ha R. Should We Ignore, Follow, or Biopsy? Impact of Artificial Intelligence Decision Support on Breast Ultrasound Lesion Assessment. AJR Am J Roentgenol 2020 Jun;214(6):1445-1452. [doi: 10.2214/AJR.19.21872] [Medline: 32319794]
48. Martin CM, Vogel C, Grady D, Zarabzadeh A, Hederman L, Kellett J, et al. Implementation of complex adaptive chronic care: the Patient Journey Record system (PaJR). J Eval Clin Pract 2012 Dec;18(6):1226-1234. [doi: 10.1111/j.1365-2753.2012.01880.x] [Medline: 22816797]
49. McCoy A, Das R. Reducing patient mortality, length of stay and readmissions through machine learning-based sepsis prediction in the emergency department, intensive care unit and hospital floor units. BMJ Open Qual 2017;6(2):e000158 [FREE Full text] [doi: 10.1136/bmjoq-2017-000158] [Medline: 29450295]
50. McNamara DM, Goldberg SL, Latts L, Atieh Graham DM, Waintraub SE, Norden AD, et al. Differential impact of cognitive computing augmented by real world evidence on novice and expert oncologists. Cancer Med 2019 Nov;8(15):6578-6584 [FREE Full text] [doi: 10.1002/cam4.2548] [Medline: 31509353]
51. Mori Y, Kudo S, Misawa M, Saito Y, Ikematsu H, Hotta K, et al. Real-Time Use of Artificial Intelligence in Identification of Diminutive Polyps During Colonoscopy: A Prospective Study. Ann Intern Med 2018 Sep 18;169(6):357-366. [doi: 10.7326/M18-0249] [Medline: 30105375]
52. Nagaratnam K, Harston G, Flossmann E, Canavan C, Geraldes RC, Edwards C. Innovative use of artificial intelligence and digital communication in acute stroke pathway in response to COVID-19. Future Healthc J 2020 Jun;7(2):169-173 [FREE Full text] [doi: 10.7861/fhj.2020-0034] [Medline: 32550287]
53. Natarajan S, Jain A, Krishnan R, Rogye A, Sivaprasad S. Diagnostic Accuracy of Community-Based Diabetic Retinopathy Screening With an Offline Artificial Intelligence System on a Smartphone. JAMA Ophthalmol 2019 Aug 08:1182-1188 [FREE Full text] [doi: 10.1001/jamaophthalmol.2019.2923] [Medline: 31393538]
54. Nicolae A, Semple M, Lu L, Smith M, Chung H, Loblaw A, et al. Conventional vs machine learning-based treatment planning in prostate brachytherapy: Results of a Phase I randomized controlled trial. Brachytherapy 2020;19(4):470-476 [FREE Full text] [doi: 10.1016/j.brachy.2020.03.004] [Medline: 32317241]
55. Park A, Chute C, Rajpurkar P, Lou J, Ball RL, Shpanskaya K, et al. Deep Learning-Assisted Diagnosis of Cerebral Aneurysms Using the HeadXNet Model. JAMA Netw Open 2019 Jun 05;2(6):e195600 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.5600] [Medline: 31173130]
56. Romero-Brufau S, Wyatt KD, Boyum P, Mickelson M, Moore M, Cognetta-Rieke C. A lesson in implementation: A pre-post study of providers' experience with artificial intelligence-based clinical decision support. Int J Med Inform 2020 May;137:104072. [doi: 10.1016/j.ijmedinf.2019.104072] [Medline: 32200295]
57. Rostill H, Nilforooshan R, Morgan A, Barnaghi P, Ream E, Chrysanthaki T. Technology integrated health management for dementia. Br J Community Nurs 2018 Oct 02;23(10):502-508. [doi: 10.12968/bjcn.2018.23.10.502] [Medline: 30290728]
58. Segal MM, Williams MS, Gropman AL, Torres AR, Forsyth R, Connolly AM, et al. Evidence-based decision support for neurological diagnosis reduces errors and unnecessary workup. J Child Neurol 2014 Apr;29(4):487-492. [doi: 10.1177/0883073813483365] [Medline: 23576414]
59. Segal MM, Athreya B, Son MBF, Tirosh I, Hausmann JS, Ang EYN, et al. Evidence-based decision support for pediatric rheumatology reduces diagnostic errors. Pediatr Rheumatol Online J 2016 Dec 13;14(1):67 [FREE Full text] [doi: 10.1186/s12969-016-0127-z] [Medline: 27964737]
60. Segal MM, Rahm AK, Hulse NC, Wood G, Williams JL, Feldman L, et al. Experience with Integrating Diagnostic Decision Support Software with Electronic Health Records: Benefits versus Risks of Information Sharing. EGEMS (Wash DC) 2017 Dec 06;5(1):23 [FREE Full text] [doi: 10.5334/egems.244] [Medline: 29930964]

61. Segal G, Segev A, Brom A, Lifshitz Y, Wasserstrum Y, Zimlichman E. Reducing drug prescription errors and adverse drug events by application of a probabilistic, machine-learning based clinical decision support system in an inpatient setting. J Am Med Inform Assoc 2019 Dec 01;26(12):1560-1565 [FREE Full text] [doi: 10.1093/jamia/ocz135] [Medline: 31390471]
62. Shimabukuro DW, Barton CW, Feldman MD, Mataraso SJ, Das R. Effect of a machine learning-based severe sepsis prediction algorithm on patient survival and hospital length of stay: a randomised clinical trial. BMJ Open Respir Res 2017;4(1):e000234 [FREE Full text] [doi: 10.1136/bmjresp-2017-000234] [Medline: 29435343]
63. Sim Y, Chung MJ, Kotter E, Yune S, Kim M, Do S, et al. Deep Convolutional Neural Network-based Software Improves Radiologist Detection of Malignant Lung Nodules on Chest Radiographs. Radiology 2020 Jan;294(1):199-209. [doi: 10.1148/radiol.2019182465] [Medline: 31714194]
64. Steiner DF, MacDonald R, Liu Y, Truszkowski P, Hipp JD, Gammage C, et al. Impact of Deep Learning Assistance on the Histopathologic Review of Lymph Nodes for Metastatic Breast Cancer. Am J Surg Pathol 2018 Dec;42(12):1636-1646 [FREE Full text] [doi: 10.1097/PAS.0000000000001151] [Medline: 30312179]
65. Su J, Li Z, Shao X, Ji C, Ji R, Zhou R, et al. Impact of a real-time automatic quality control system on colorectal polyp and adenoma detection: a prospective randomized controlled study (with videos). Gastrointest Endosc 2020 Feb;91(2):415-424.e4. [doi: 10.1016/j.gie.2019.08.026] [Medline: 31454493]
66. Titano JJ, Badgeley M, Schefflein J, Pain M, Su A, Cai M, et al. Automated deep-neural-network surveillance of cranial images for acute neurologic events. Nat Med 2018 Sep;24(9):1337-1341. [doi: 10.1038/s41591-018-0147-y] [Medline: 30104767]
67. Vandenberghe ME, Scott MLJ, Scorer PW, Söderberg M, Balcerzak D, Barker C. Relevance of deep learning to facilitate the diagnosis of HER2 status in breast cancer. Sci Rep 2017 Apr 05;7:45938 [FREE Full text] [doi: 10.1038/srep45938] [Medline: 28378829]
68. Voermans AM, Mewes JC, Broyles MR, Steuten LMG. Cost-Effectiveness Analysis of a Procalcitonin-Guided Decision Algorithm for Antibiotic Stewardship Using Real-World U.S. Hospital Data. OMICS 2019 Oct;23(10):508-515 [FREE Full text] [doi: 10.1089/omi.2019.0113] [Medline: 31509068]
69. Wang P, Berzin TM, Glissen Brown JR, Bharadwaj S, Becq A, Xiao X, et al. Real-time automatic detection system increases colonoscopic polyp and adenoma detection rates: a prospective randomised controlled study. Gut 2019 Oct;68(10):1813-1819 [FREE Full text] [doi: 10.1136/gutjnl-2018-317500] [Medline: 30814121]
70. Wang SV, Rogers JR, Jin Y, DeiCicchi D, Dejene S, Connors JM, et al. Stepped-wedge randomised trial to evaluate population health intervention designed to increase appropriate anticoagulation in patients with atrial fibrillation. BMJ Qual Saf 2019 Oct;28(10):835-842 [FREE Full text] [doi: 10.1136/bmjqs-2019-009367] [Medline: 31243156]
71. Wang P, Liu X, Berzin TM, Glissen Brown JR, Liu P, Zhou C, et al. Effect of a deep-learning computer-aided detection system on adenoma detection during colonoscopy (CADe-DB trial): a double-blind randomised study. The Lancet Gastroenterology & Hepatology 2020 Apr;5(4):343-351. [doi: 10.1016/S2468-1253(19)30411-X] [Medline: 31981517]
72. Wijnberge M, Geerts BF, Hol L, Lemmers N, Mulder MP, Berge P, et al. Effect of a Machine Learning-Derived Early Warning System for Intraoperative Hypotension vs Standard Care on Depth and Duration of Intraoperative Hypotension During Elective Noncardiac Surgery: The HYPE Randomized Clinical Trial. JAMA 2020 Mar 17;323(11):1052-1060 [FREE Full text] [doi: 10.1001/jama.2020.0592] [Medline: 32065827]
73. Wu X, Huang Y, Liu Z, Lai W, Long E, Zhang K, et al. Universal artificial intelligence platform for collaborative management of cataracts. Br J Ophthalmol 2019 Nov;103(11):1553-1560 [FREE Full text] [doi: 10.1136/bjophthalmol-2019-314729] [Medline: 31481392]
74. Wu L, Zhang J, Zhou W, An P, Shen L, Liu J, et al. Randomised controlled trial of WISENSE, a real-time quality improving system for monitoring blind spots during esophagogastroduodenoscopy. Gut 2019 Dec;68(12):2161-2169 [FREE Full text] [doi: 10.1136/gutjnl-2018-317366] [Medline: 30858305]
75. Yoo YJ, Ha EJ, Cho YJ, Kim HL, Han M, Kang SY. Computer-Aided Diagnosis of Thyroid Nodules via Ultrasonography: Initial Clinical Experience. Korean J Radiol 2018;19(4):665-672 [FREE Full text] [doi: 10.3348/kjr.2018.19.4.665] [Medline: 29962872]
76. Higgins J, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al, editors. Cochrane Handbook for Systematic Reviews of Interventions. Chichester, UK: John Wiley & Sons; 2019.
77. Auloge P, Cazzato RL, Ramamurthy N, de Marini P, Rousseau C, Garnon J, et al. Augmented reality and artificial intelligence-based navigation during percutaneous vertebroplasty: a pilot randomised clinical trial. Eur Spine J 2020 Jul;29(7):1580-1589. [doi: 10.1007/s00586-019-06054-6] [Medline: 31270676]
78. Liu C, Liu X, Wu F, Xie M, Feng Y, Hu C. Using Artificial Intelligence (Watson for Oncology) for Treatment Recommendations Amongst Chinese Patients with Lung Cancer: Feasibility Study. J Med Internet Res 2018 Sep 25;20(9):e11087 [FREE Full text] [doi: 10.2196/11087] [Medline: 30257820]
79. Embrace Healthcare Revolution: Strategically moving from volume to value. Deloitte. URL: https://www2.deloitte.com/cn/en/pages/life-sciences-and-healthcare/articles/embrace-healthcare-revolution.html [accessed 2021-04-05]
80. Sun Y. AI could alleviate China's doctor shortage. MIT Technology Review. 2018 Mar 21. URL: https://www.technologyreview.com/2018/03/21/144544/ai-could-alleviate-chinas-doctor-shortage/ [accessed 2021-04-05]

81. China Is Building The Ultimate Technological Health Paradise. Or Is It? The Medical Futurist. 2019 Feb 19. URL: https://medicalfuturist.com/china-digital-health/ [accessed 2021-04-05]
82. IDC Identifies China's Emerging Healthcare AI Trends in New CIO Perspective Report. IDC China. 2020 Apr 14. URL: https://www.idc.com/getdoc.jsp?containerId=prCHC46215920 [accessed 2021-04-05]
83. Ngiam KY, Khor IW. Big data and machine learning algorithms for health-care delivery. The Lancet Oncology 2019 May;20(5):e262-e273. [doi: 10.1016/S1470-2045(19)30149-4] [Medline: 31044724]
84. Lysaght T, Lim HY, Xafis V, Ngiam KY. AI-Assisted Decision-making in Healthcare: The Application of an Ethics Framework for Big Data in Health and Research. Asian Bioeth Rev 2019 Sep 12;11(3):299-314 [FREE Full text] [doi: 10.1007/s41649-019-00096-0] [Medline: 33717318]
85. Longoni C, Bonezzi A, Morewedge C. Resistance to medical artificial intelligence. Journal of Consumer Research 2019;46(4):629-650. [doi: 10.1093/jcr/ucz013]
86. Shaffer VA, Probst CA, Merkle EC, Arkes HR, Medow MA. Why do patients derogate physicians who use a computer-based diagnostic support system? Med Decis Making 2013 Jan;33(1):108-118. [doi: 10.1177/0272989X12453501] [Medline: 22820049]
87. Innovative solutions for digital health. Accenture. 2018. URL: https://www.accenture.com/us-en/insights/health/healthcare-walking-ai-talk [accessed 2021-04-05]
88. Carfagno J. IDx-DR, the First FDA-Approved AI System, is Growing Rapidly Internet. DocWire News. 2019. URL: https://www.docwirenews.com/docwire-pick/future-of-medicine-picks/idx-dr-the-first-fda-approved-ai-system-is-growing-rapidly/ [accessed 2021-04-05]

Abbreviations

AI: artificial intelligence
DL: deep learning
FDA: Food and Drug Administration
ICU: intensive care unit
ML: machine learning
NLP: natural language processing
NPV: negative-predictive value
PPV: positive-predictive value
RCT: randomized controlled trial

Edited by T Kool; submitted 14.11.20; peer-reviewed by S Guo, E Loh; comments to author 11.01.21; revised version received 08.03.21; accepted 09.03.21; published 22.04.21.

Please cite as:
Yin J, Ngiam KY, Teo HH
Role of Artificial Intelligence Applications in Real-Life Clinical Practice: Systematic Review
J Med Internet Res 2021;23(4):e25759
URL: https://www.jmir.org/2021/4/e25759
doi:10.2196/25759
PMID:33885365

Original Paper

Deep Convolutional Neural Network–Based Computer-Aided Detection System for COVID-19 Using Multiple Lung Scans: Design and Implementation Study

Mustafa Ghaderzadeh[1], PhD; Farkhondeh Asadi[1], PhD; Ramezan Jafari[2], MD; Davood Bashash[3], PhD; Hassan Abolghasemi[4], MD; Mehrad Aria[5], PhD

[1]Department of Health Information Technology and Management, School of Allied Medical Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran

[2]Department of Radiology, Baqiyatallah University of Medical Sciences, Tehran, Iran

[3]Department of Hematology and Blood Banking, School of Allied Medical Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran

[4]Pediatric Congenital Hematologic Disorders Research Center, Shahid Beheshti University of Medical Sciences, Tehran, Iran

[5]Department of Computer Engineering, Faculty of Electrical and Computer Engineering, Shiraz University, Shiraz, Iran

Corresponding Author:
Farkhondeh Asadi, PhD
Department of Health Information Technology and Management
School of Allied Medical Sciences
Shahid Beheshti University of Medical Sciences
Darband St
Ghods Square
Tehran
Iran
Phone: 98 9123187253
Email: Asadifar@sbmu.ac.ir

Abstract

Background: Owing to the COVID-19 pandemic and the imminent collapse of health care systems following the exhaustion of financial, hospital, and medicinal resources, the World Health Organization changed the alert level of the COVID-19 pandemic from high to very high. Meanwhile, more cost-effective and precise COVID-19 detection methods are being preferred worldwide.

Objective: Machine vision–based COVID-19 detection methods, especially deep learning as a diagnostic method in the early stages of the pandemic, have been assigned great importance during the pandemic. This study aimed to design a highly efficient computer-aided detection (CAD) system for COVID-19 by using a neural search architecture network (NASNet)–based algorithm.

Methods: NASNet, a state-of-the-art pretrained convolutional neural network for image feature extraction, was adopted to identify patients with COVID-19 in their early stages of the disease. A local data set, comprising 10,153 computed tomography scans of 190 patients with and 59 without COVID-19 was used.

Results: After fitting on the training data set, hyperparameter tuning, and topological alterations of the classifier block, the proposed NASNet-based model was evaluated on the test data set and yielded remarkable results. The proposed model's performance achieved a detection sensitivity, specificity, and accuracy of 0.999, 0.986, and 0.996, respectively.

Conclusions: The proposed model achieved acceptable results in the categorization of 2 data classes. Therefore, a CAD system was designed on the basis of this model for COVID-19 detection using multiple lung computed tomography scans. The system differentiated all COVID-19 cases from non–COVID-19 ones without any error in the application phase. Overall, the proposed deep learning–based CAD system can greatly help radiologists detect COVID-19 in its early stages. During the COVID-19 pandemic, the use of a CAD system as a screening tool would accelerate disease detection and prevent the loss of health care resources.

(J Med Internet Res 2021;23(4):e27468) doi:10.2196/27468

KEYWORDS

artificial intelligence; classification; computer-aided detection; computed tomography scan; convolutional neural network; coronavirus; COVID-19; deep learning; machine learning; machine vision; model; pandemic

Introduction

In 2020, the rapid global spread of COVID-19 made the World Health Organization declare the first pandemic of the 21st century, with the highest level of alert worldwide. Based on the WorldMeters statistics, until January 5, 2021, more than 86 million people worldwide contracted this disease, with more than 1,870,000 confirmed deaths due to COVID-19. Early detection of COVID-19 is essential not only for patient care but also for public health by ensuring patients' isolation and controlling disease spread [1,2]. The first and most important step to control this pandemic is the rapid detection of infected patients and monitoring of positive cases. Various diagnostic methods for the rapid detection of COVID-19 have been introduced by different studies and by the WHO, with the reverse transcription–polymerase chain reaction (RT–PCR) test being the most prominent diagnostic method. Although RT–PCR is the gold standard for COVID-19 detection, owing to its time-intensiveness and high cost, infected individuals, as a source of transmission, can transmit the virus to many people while they are waiting to receive the results of their RT–PCR test. Moreover, previous studies have reported that the RT–PCR test has a high false-negative rate; this is a major limitation of this diagnostic test and reduces its sensitivity. Furthermore, this leads to delayed detection, treatment, and—in advanced stages of the disease—an increased mortality rate [3-8]. A high influx of patients at diagnostic centers during the pandemic has led to excessive use of resources and a shortage of RT–PCR test kits. Regardless of the need of RT–PCR tests for suspected individuals, the multiple repeats of these tests for patients have imposed a heavy burden on health care sources. The time-consuming nature of laboratory tests, coupled with the molecular and nonspecific nature of serological tests, has necessitated the use of a cheaper test focusing on findings in the lung tissue. As major lung health monitoring tools, radiological tests have attracted the attention of clinical specialists. For COVID-19 evaluation, computed tomography (CT) is a more sensitive and specific detection method than chest X-ray imaging, and, in many cases, lung involvement and ground-glass opacities (GGO) can be viewed on CT even before the onset of clinical symptoms and before obtaining positive results on an RT–PCR test. This implies that, in many cases, before the emergence of the first clinical symptoms and a positive RT–PCR finding, the complications of COVID-19 can be detected in the lungs. Based on previous reports and the WHO's recommendations, chest CT has emerged as a valuable tool for the early detection and triaging of individuals suspected with COVID-19 [4,9,10]. In a study on 1014 patients with COVID-19, CT enabled more sensitive detection than RT–PCR [11]. Despite the success of this radiological modality in detecting COVID-19–related lung damage, certain problems are associated with its use. Despite the WHO's recommendations, chest CT findings are normal in some patients at the outset of the disease, and this lends a negative predictive value to CT alone. The low-specificity of CT can deter disease detection in non–COVID-19 cases. In addition, ionizing radiation from the CT scanner can cause problems to patients who require multiple CT scans during the course of their disease [12-16]. In the past decade, numerous computer-based methods have been employed for improving the efficiency of medical imaging techniques. One such method is the use of machine learning algorithms, which has had remarkable success in medical imaging. Among different types of machine learning methods, deep learning models have achieved high precision in machine vision tasks rapidly after the emergence of COVID-19. Convolutional neural networks (CNNs) have high potential for feature extraction and analysis. Upon the emergence of COVID-19, and owing to the limitations of diagnostic tests, numerous machine learning techniques have been adopted to improve the precision of diagnostic methods. Table 1 lists some relevant studies.

Table 1. Studies evaluating machine learning algorithms used for COVID-19 detection.

Study (country)	Study objective	Population	Models used	Evaluation results
Ni et al (China) [15]	Automatic detection	14,531	Convolutional multiview feature pyramid network with positron-aware attention and a 3D U-Net	• F_1 score=97% • Sensitivity=100%
Wang et al (China) [17]	Diagnostic and prognostic analysis	5372	Densenet121-feature pyramid network	• Area under the receiver operating characteristic curve=87%-88% • Sensitivity=80.3%-79.35%
Hasan et al (Iraq) [18]	Diagnosis (classification)	321	Long short-term memory classifier	• Accuracy=99.68%
Pathak et al (India) [19]	Classification (detection)	852	ResNet-50	• Accuracy=93.01%
Ardakani et al (Iran) [20]	Detection	194	10 pretrained convolutional neural networks: AlexNet, VGG-16, VGG-19, SqueezeNet, GoogleNet, MobileNet-V2, ResNet-18, ResNet-50, ResNet-101, and Xception	• Best performance: ResNet-101 and Xception • Sensitivity (ResNet-101)=100% • Sensitivity (Xception)=98.04% • Specificity (ResNet-101): 99.02% • Specificity (Xception)=100% • Accuracy (ResNet-101)=99.51% • Accuracy (Xception)=99.02%
Li et al (China) [21]	Automatic detection	4356	RestNet50 as the backbone of the main model	• Sensitivity=90% • Specificity=96%
Mei et al (United States) [22]	Rapid diagnosis	905	Inception-ResNet-V2	• Correctly identified 17 of 25 (68%) patients with COVID-19
Song et al (China) [23]	Diagnosis	227	Bidirectional generative adversarial network	• Sensitivity=85% • Specificity=88%

Methods

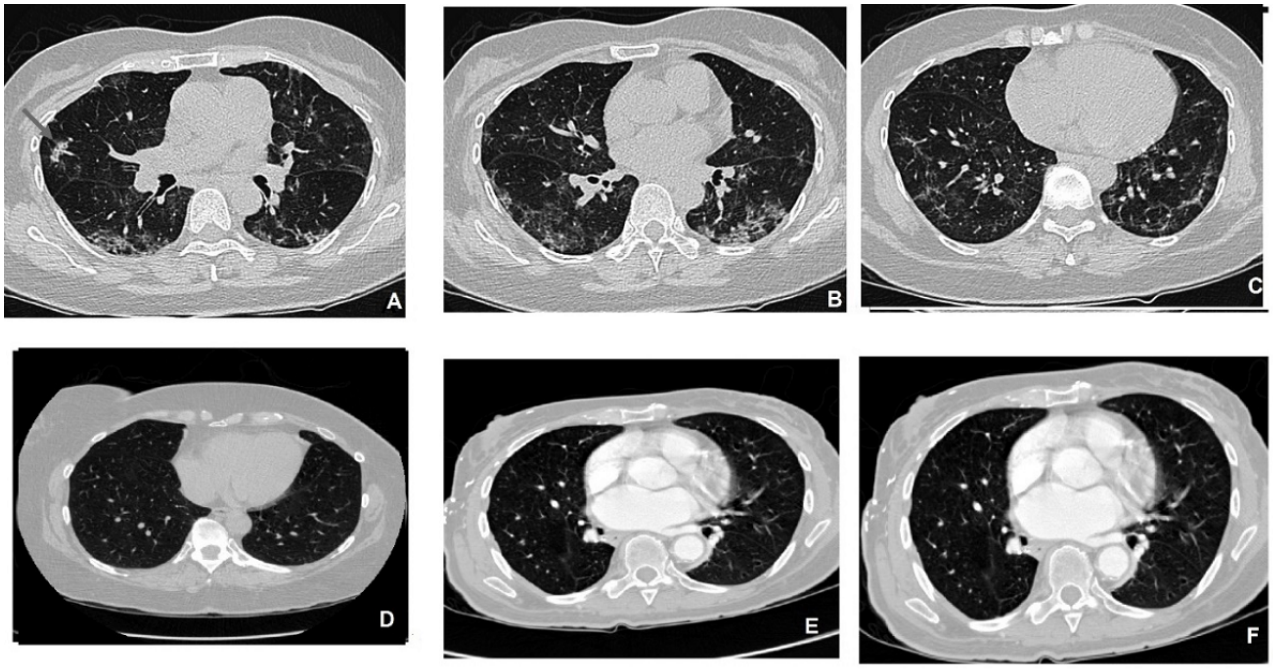
Study Overview

Based on the success of CNNs in machine vision tasks, we designed and implemented a model for the classification of CT images of individuals with and those without COVID-19 through a deep neural network based on a Neural Search Architecture Network (NASNet) [24] feature extractor.

Data Set

The data set comprised 10,153 CT scans, of which 7644 belong to 190 patients with COVID-19 and 2509 belong to 59 people without COVID-19, including those with pneumonia and otherwise healthy individuals who visited the hospital owing to a suspicion of COVID-19 [25]. All these images were collected from the radiology centers of teaching hospitals in Tehran, Iran. The disease status in suspected individuals was confirmed in this set after an RT–PCR test. Figure 1 shows the CT scans of some patients with COVID-19 and their counterparts with suspected disease.

Figure 1. Axial computed tomography scan slices of the lung. (A, B, C) Non–COVID-19 cases including those of pneumonia and healthy individuals; (D, E, F) infected lungs of patients with COVID-19.



Proposed Method

To detect COVID-19 in patients at an early stage of the disease from multiple lung CT scans, a state-of-the-art model based on a NASNet CNN feature extractor was proposed. Based on the proposed model, a computer-aided detection (CAD) system was designed.

Data Preparation and Preprocessing

For data preparation, lung CT scans were first received in the Digital Imaging and Communications in Medicine format as the output of the picture archiving and communications system of a diagnostic center. In the preprocessing stage, the images were converted to the commonly used JPG format, and the order of the color channels was changed from the default BGR to RGB to prepare the images for processing.

Based on the literature, the success of medical image visual tasks in deep learning is not merely attributed to CNN models; rather, a major part of this success results from image preprocessing [26]. Data normalization to maintain the integrity of the images was performed as the first step of preprocessing, which plays a key role in the analysis of CT scans [27]. To this end, first, the pixel-level global mean (SD) values were calculated for all the images; thereafter, the data were normalized using the following equation:

where  is the global mean of the image set X, σ is the SD, and ε=1e–10 is an insignificant value to prevent the denominator from becoming 0.

After normalization, for standardizing the images to achieve a unified scale for the input of the deep neural network, the values of the pixels of each image were scaled by transferring to (0,255) and then by transforming to the (0,1) interval, such that the images would be standard during training. Since CNNs depend on a large number of data to enhance their efficiency and prevent model over-fitting [28,29]; at this stage, data were augmented for the training data set through random rotation, contrast alteration, illumination alteration, and gamma correction.

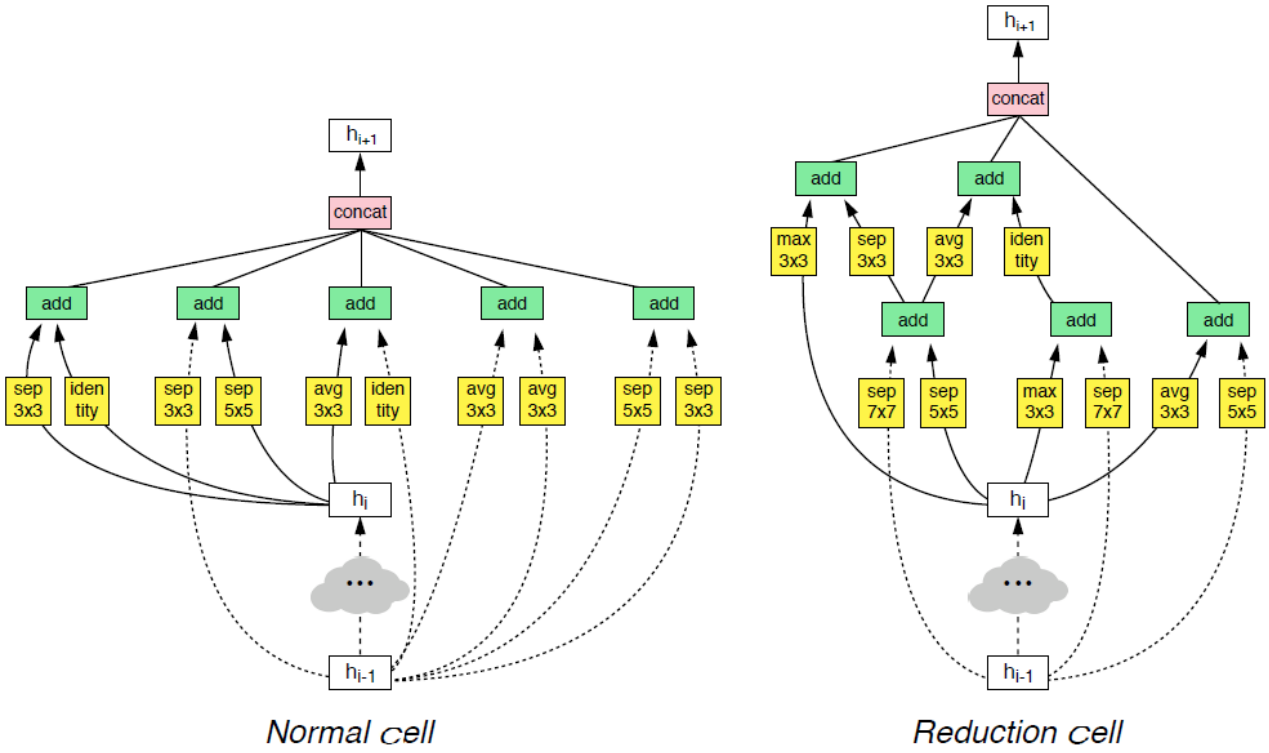
All the images of the sets were shuffled such that the network would not necessarily see the data of only a certain class during training, and each batch would include images with different labels belonging to both COVID-19 and non–COVID-19 classes. The dimensions of the input images were changed to 224×224×3; however, this method can be used on images with any other dimensions. The data set with the 64:16:20 ratio was randomly divided among the training, validation, and test sets, respectively, whereby 20% of the data were allocated to the test set, the remaining 80% to the training set, and 20% of the training set was assigned to the validation set.

Feature Extraction and Classification

Convolutional layers were used in the feature extraction block. Immediately after each Conv2D layer, the useful statistics were collected using the Max-Pooling module and, after normalizing them using batch normalization, passed to the next CNN block. To prevent model overfitting, in addition to batch normalization, weight regularization and dropout methods were also applied. For regularization, the Euclidean norm (L2) was used with different coefficient values in the (0.001-0.01) interval after activation using the LeakyReLU activation function and dropout of 20%-30% of the weights. Inspired by the transfer learning approach, for better feature extraction, the preliminary blocks of the pretrained NASNetLarge network were used. NASNet has a scalable architecture for image classification and consists of 2 repeated motifs termed the normal cell and the reduction cell. Figure 2 illustrates the architecture of these convolutional cells. All parameters were initialized using the weights obtained from fitting NASNetLarge on the ImageNet data set. After the

feature extraction block, the weights were transferred to 3 dense or fully connected layers by using a global average pooling flatten layer in the form of a 1-dimensional tensor.

Figure 2. Architecture of the NASNet's convolutional cells with B=5 blocks. The input (white) is the hidden state from previous activations (or input image). The output (pink) is the result of a concatenation operation across all resulting branches. Each convolutional cell is the result of B blocks. A single block corresponds to 2 primitive operations (yellow) and a combination operation (green) [23].



In these layers, batch normalization, regularization, and weight dropout were performed as well. The first dense layer used a ReLU activation function, the next layer utilized a LeakyReLU, and the last layer, which actually is the classifier layer, employed a Softmax multiclass activation function.

During the training process, in the first phase, the feature extraction block was frozen and contained nontrainable parameters, and Adam was used as the optimizer. In this phase, the initial learning rate of 1*e*–3 and the binary cross-entropy loss function were used. If validation loss remained stable in every 10 epochs, the learning rate would be reduced by 20% to a minimum of 1*e*–6 in case of no further improvement. If the validation loss remained stable up to 20 epochs, the process of training would stop. Eventually, only the best weights were saved. After training the dense layers, in the second phase the feature extraction block was unfreezed, and the network—this time fully trainable—was fitted once more on the same data; in this phase, the stochastic gradient descent optimizer with the initial learning rate of 1*e*–4 was utilized. The batch size was 32; the number of epochs in the first phase was assumed to be 200, and 1000 iterations were considered in the second phase.

CAD System Based on the Proposed Model

Many studies have recommended the use of CT scans for COVID-19 detection, many of which have used machine learning–based computer methods to enhance the results of chest CT. All machine learning methods have attempted to detect COVID-19 in the images with a single CT slice [10,12,30-33]; however, in real time, radiologists confirm or reject COVID-19 on the basis of overall slices of a patient's CT scan. This study aimed to design a computer-aided diagnostic system to detect COVID-19 with multiple CT images for each person. In the CAD system designed on the basis of the proposed model, 4 CT slices were obtained from a person suspected with COVID-19 and the system estimates the final result from the output average mean of classifying all the slices. However, the proposed system can receive a different number of slides and does not depend on the number of inputs. This increases the reliability of the results obtained from the proposed model. Figure 3 provides a schematic representation of the proposed model.

In the experiments, the proposed model successfully detected all the cases of COVID-19 with high accuracy and differentiated all the positive and negative cases without any discernible error. Figures 4 and 5 display the performance of the proposed model by presenting the results of detection on the first 25 samples and 25 random ones from the test set, respectively.

Figure 3. Proposed deep convolutional neural network–based CAD system for COVID-19 detection using multiple lung computed tomography scans. CT: computed tomography.

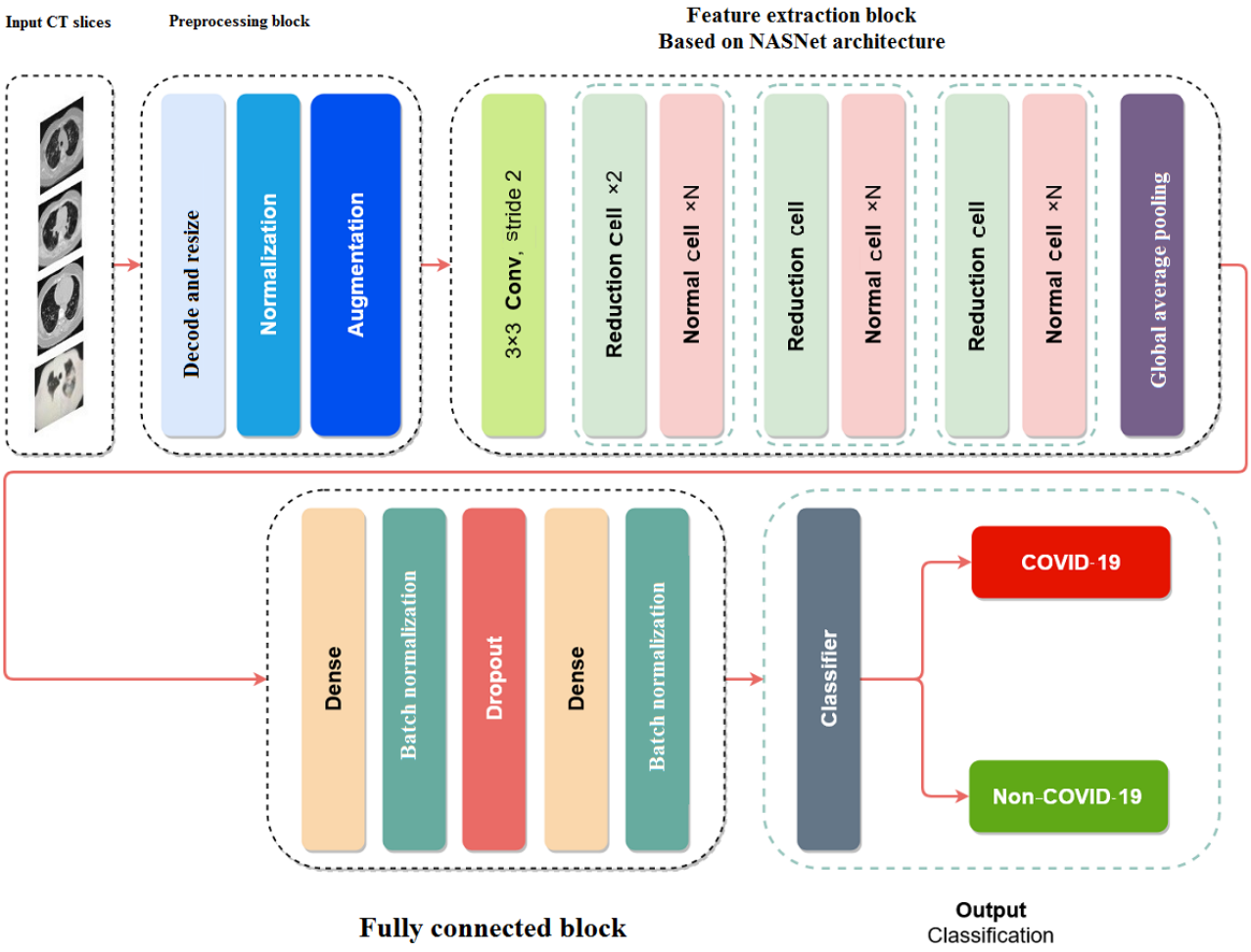


Figure 4. Results of the detection on the first 25 samples from the test set. “I” is the image index, “P” is the predicted value, and “L” is the grand truth label. Green indicates correct detection and red indicates incorrect detection.

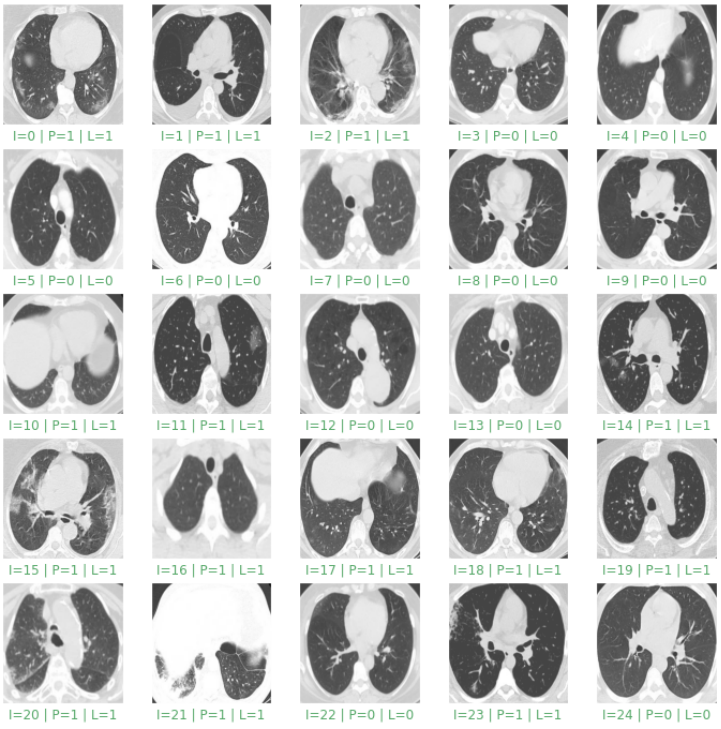
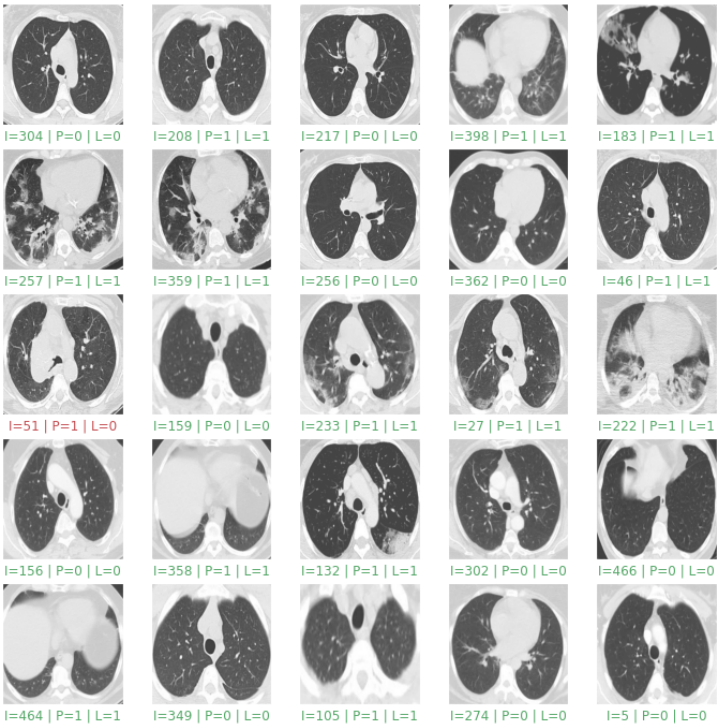


Figure 5. Results of detection on 25 random samples from the test set. “I” is the image index, “P” is the predicted value, and “L” is the grand truth label. Green indicates correct detection and red indicates incorrect detection.



In the case of COVID-19, the lack of an early and accurate diagnosis leads to the spread of the disease among other individuals, which has irreversible effects on the control of the pandemic. The proposed CAD system, which detects cases of COVID-19 from multiple CT slices can perform more accurately because, owing to the error of not showing the area of the GGO in some CT slices, the data of many slices may have been missed. Moreover, in models that detect COVID-19 from a single slice, there is a risk of error in viewing the infected area of the lung or ROI owing to operator errors, the angle of the slices, or problems with the CT scanner tube. Thus, by using a multislice CAD system, the disease can be detected in its early stages, and the initial signs of lung involvement can be discovered with maximum precision.

Implementation

The proposed method was implemented using the Python programming language by Keras, which is a high-level library for TensorFlow machine learning framework which also utilized a Compute Unified Device Architecture deep learning network library for parallel processing on the graphics processing unit. The computer system had an Intel Core i7 7700K CPU, 32 GB RAM, and an Nvidia T4 GPU accelerator. Implementation codes and the pretrained model are available on GitHub [34].

Results

Metrics

To quantitatively evaluate the performance of the proposed method, the sensitivity, specificity, accuracy, and F_1 score evaluation criteria were determined on the basis of the model’s performance by using a confusion matrix. Here, sensitivity was defined as the ratio of COVID-19 cases correctly detected by the model to all the actual COVID-19 cases. Specificity was defined as the ratio of the non–COVID-19 cases correctly detected by the model to all the actual non–COVID-19 cases. Moreover, accuracy was defined as the rate of all the COVID-19 and non–COVID-19 cases accurately detected on the basis of the CT images.

Experimental Results and Evaluation

In this study, we used traditional measures to evaluate the performance of the proposed model, using a confusion matrix. Based on this confusion matrix, the specificity and sensitivity in measuring and analyzing the performance of the proposed CAD model were calculated, where specificity was defined as the ability of the classifier to correctly identify individuals without COVID-19 (true-negative rate). Sensitivity was defined as the classifier’s ability to identify individuals with COVID-19 correctly (true-positive rate). These evaluations were performed using the following equations:

Figure 6 shows the confusion matrix for the evaluation in the test set for 2 classes. The evaluation criteria based on the confusion matrix are provided in Table 2.

Figure 6. (A) Confusion matrix and (B) normalized confusion matrix of the model performance for the test data.

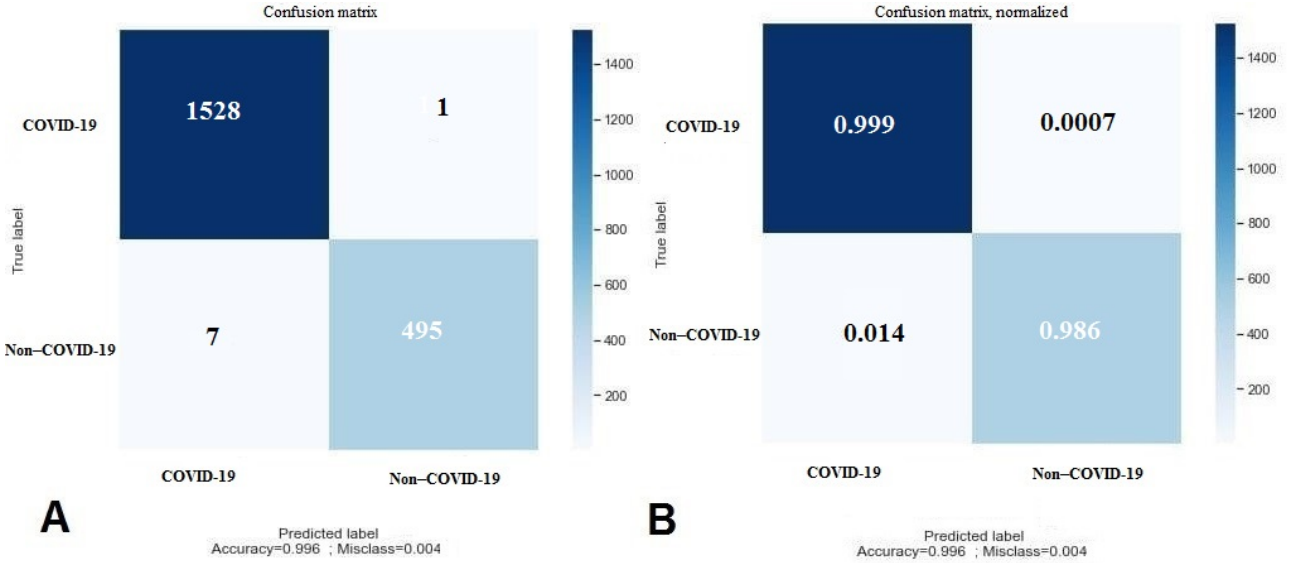


Table 2. Performance of the proposed method in the test data set.

Metric	Value
Sensitivity	99.9
Specificity	98.6
Accuracy	99.6

The learning curve of the proposed model for the training and validation sets is illustrated in Figure 7. On assessing the behavior of the proposed model in handling new validation data, we observed that with increased epochs, the model had a lower error rate, and thus enhanced accuracy for the unknown data, which suggests that the model has high potential for detecting new cases of COVID-19 from the CT scans. The mean square error in detecting all the COVID-19 and non–COVID-19 cases from among the test set images was 0.003938, which was significantly lower than that reported in previous studies.

The metrics of positive predictive value and negative predictive value, as well as the F_1 score, for the proposed model are shown in Table 3.

Figure 7. Training and validation loss and accuracy.

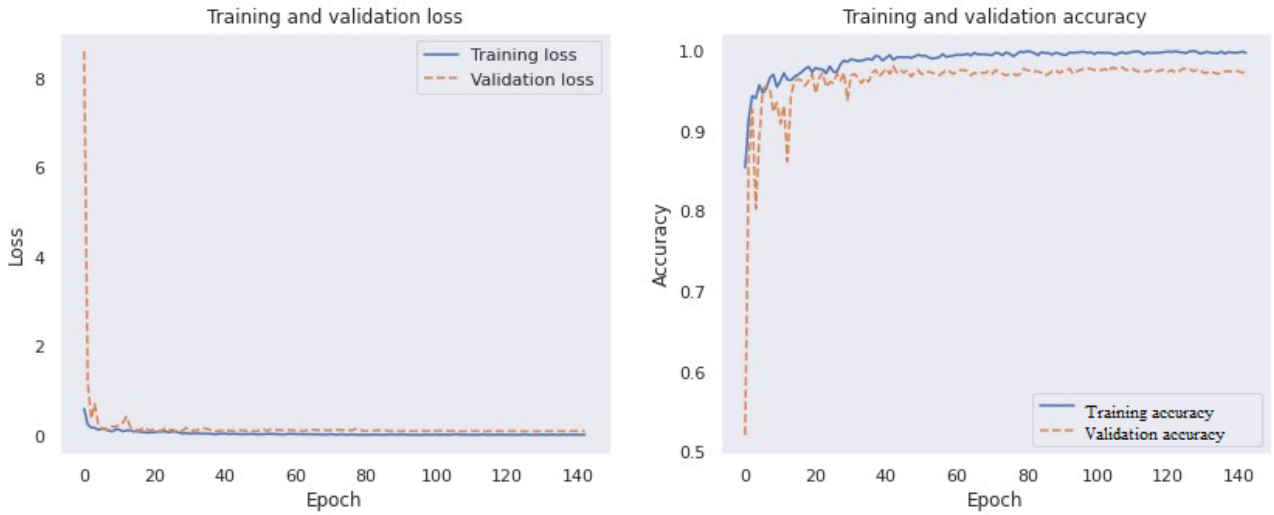


Table 3. Evaluation of the proposed model for the test set.

Metric	Value (%)
Positive predictive value = True positive / Positive calls	99.8%
Negative predictive value = True negative / Negative calls	99.7%
F_1 score = (True positive / True positive) + 0.5 (False positive + False negative)	99.8%

To evaluate the performance of the proposed model further in real-life applications and present a comparable evaluation, we tested our model on a publicly available and well-known data set [30] using the cross–data set evaluation approach. The results shown in Table 4 compare the proposed method with other state-of-the-art approaches, including traditional deep neural networks and pretrained networks [30-33,35].

Table 4. Comparison of the performance of different models for detecting COVID-19 using various evaluation metrics.

Model	Evaluation Metrics			
	Accuracy	Precision	Recall	F_1 score
SqueezeNet	95.1	94.2	96.2	95.2
ShuffleNet	97.5	96.1	99.0	97.5
GoogleNet	91.7	90.2	93.5	91.8
VGG-16	94.9	94.0	95.4	94.9
AlexNet	93.7	94.9	92.2	93.6
ResNet50	94.9	93.0	97.1	95.0
Xception	98.8	99.0	98.6	98.8
AdaBoost	95.1	93.6	96.7	95.1
Decision Tree	79.4	76.8	83.1	79.8
Explainable deep learning [30]	97.3	99.1	95.5	97.3
DenseNet201 [31]	96.2	96.2	96.2	96.2
Modied VGG19 [32]	95.0	95.3	94.0	94.3
COVID CT-Net [33]	90.7	88.5	85.0	90.0
Contrastive Learning [35]	90.8	95.7	85.8	90.8
Proposed	99.4	99.6	99.8	99.5

Discussion

Principal Findings

RT–PCR is the definitive method for diagnosing COVID-19. However, the nucleic acid test is very time-consuming, and sputum analysis may take several days. This test's high cost and low sensitivity have caused major problems to health care systems during the pandemic. Consequently, people with false-negative findings on RT–PCR have been a source of virus transmission and have spread the virus to others. When the WHO emphasized the need to increase diagnostic tests and comprehensively evaluate suspected individuals, physicians and health care systems were encouraged to utilize cheaper and faster tests [36-38]. When attempting to detect COVID-19 in its initial stages, a lung CT scan does not always demonstrate the lung consolidation areas, and no GGO findings are observed in many cases. Machine learning models can enhance the efficiency of radiological diagnostic methods and serve as a suitable alternative to the RT–PCR test. The core of the CAD system designed in this study is based on a deep CNN architecture and uses an input with 4 slices. NASNet was utilized here because it could determine the best architecture for feature engineering [24]. No previous study has employed this technological model for analyzing the CT scans of individuals suspected with COVID-19. Further examination of medical image processing revealed the remarkable performance of this model in image feature extraction. This study achieved maximum sensitivity and precision in detecting COVID-19 compared to previous studies. Considering the algorithm and the use of multiple chest CT scan slices for a single patient, the proposed system can be employed at diagnostic centers as a reliable method to detect individuals with COVID-19 with high precision in the early stages of the disease. In the future, this CAD can be included in the picture archiving and communications systems of radiology wards to achieve an automated and more efficient diagnosis.

Conclusions

Using the CAD system for detecting COVID-19 during the pandemic minimizes the time of image interpretation and consequently the number of patients waiting at radiology centers.

Furthermore, by increasing the number of images produced by the CT scanner and increasing the population size, better classification results for differentiating positive and negative cases can be expected.

Acknowledgments

This study was part of a research project conducted at Shahid Beheshti University of Medical Sciences (Tehran, Iran) and was approved by Iran National Committee for Ethics in Biomedical Research (approval ID IR.SBMU.RETECH.REC.1399.132).

Conflicts of Interest

None declared.

References

1. WHO Coronavirus Disease (COVID-19) Dashboard: Vaccination. World Health Organization. URL: https://covid19.who.int/ [accessed 2021-04-17]
2. WHO Coronavirus Disease (COVID-19) Dashboard. World Health Organization. URL: https://covid19.who.int/
3. Nishiura H, Jung S, Linton NM, Kinoshita R, Yang Y, Hayashi K, et al. The Extent of Transmission of Novel Coronavirus in Wuhan, China, 2020. J Clin Med 2020 Jan 24;9(2):330 [FREE Full text] [doi: 10.3390/jcm9020330] [Medline: 31991628]
4. Chung M, Bernheim A, Mei X, Zhang N, Huang M, Zeng X, et al. CT Imaging Features of 2019 Novel Coronavirus (2019-nCoV). Radiology 2020 Apr;295(1):202-207 [FREE Full text] [doi: 10.1148/radiol.2020200230] [Medline: 32017661]
5. Ye Z, Zhang Y, Wang Y, Huang Z, Song B. Chest CT manifestations of new coronavirus disease 2019 (COVID-19): a pictorial review. Eur Radiol 2020 Aug;30(8):4381-4389 [FREE Full text] [doi: 10.1007/s00330-020-06801-0] [Medline: 32193638]
6. Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia. N Engl J Med 2020 Mar 26;382(13):1199-1207 [FREE Full text] [doi: 10.1056/NEJMoa2001316] [Medline: 31995857]
7. Wang L, Lin Z, Wong A. COVID-Net: a tailored deep convolutional neural network design for detection of COVID-19 cases from chest X-ray images. Sci Rep 2020 Nov 11;10(1):19549 [FREE Full text] [doi: 10.1038/s41598-020-76550-z] [Medline: 33177550]
8. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet 2020 Feb 15;395(10223):497-506. [doi: 10.1016/S0140-6736(20)30183-5] [Medline: 31986264]
9. Choi H, Qi X, Yoon SH, Park SJ, Lee KH, Kim JY, et al. Extension of Coronavirus Disease 2019 on Chest CT and Implications for Chest Radiographic Interpretation. Radiol Cardiothorac Imaging 2020 Apr;2(2):e200107 [FREE Full text] [doi: 10.1148/ryct.2020200107] [Medline: 33778565]
10. Wang D, Hu B, Hu C, Zhu F, Liu X, Zhang J, et al. Clinical Characteristics of 138 Hospitalized Patients With 2019 Novel Coronavirus-Infected Pneumonia in Wuhan, China. JAMA 2020 Mar 17;323(11):1061-1069. [doi: 10.1001/jama.2020.1585] [Medline: 32031570]
11. Ai T, Yang Z, Hou H, Zhan C, Chen C, Lv W, et al. Correlation of Chest CT and RT-PCR Testing for Coronavirus Disease 2019 (COVID-19) in China: A Report of 1014 Cases. Radiology 2020 Aug;296(2):E32-E40 [FREE Full text] [doi: 10.1148/radiol.2020200642] [Medline: 32101510]
12. Gozes O, Frid-Adar M, Greenspan H, Browning PD, Zhang H, Ji W, et al. arXiv. Mar 24 Preprint posted online March 24, 2020.
13. Yamashita R, Nishio M, Do RKG, Togashi K. Convolutional neural networks: an overview and application in radiology. Insights Imaging 2018 Aug;9(4):611-629 [FREE Full text] [doi: 10.1007/s13244-018-0639-9] [Medline: 29934920]
14. Wang S, Kang B, Ma J, Zeng X, Xiao M, Guo J, et al. A deep learning algorithm using CT images to screen for Corona virus disease (COVID-19). Eur Radiol 2021 Feb 24:1-9 [FREE Full text] [doi: 10.1007/s00330-021-07715-1] [Medline: 33629156]
15. Ni Q, Sun ZY, Qi L, Chen W, Yang Y, Wang L, et al. A deep learning approach to characterize 2019 coronavirus disease (COVID-19) pneumonia in chest CT images. Eur Radiol 2020 Dec;30(12):6517-6527 [FREE Full text] [doi: 10.1007/s00330-020-07044-9] [Medline: 32617690]
16. Zhang K, Liu X, Shen J, Li Z, Sang Y, Wu X, et al. Clinically Applicable AI System for Accurate Diagnosis, Quantitative Measurements, and Prognosis of COVID-19 Pneumonia Using Computed Tomography. Cell 2020 Jun 11;181(6):1423-1433.e11 [FREE Full text] [doi: 10.1016/j.cell.2020.04.045] [Medline: 32416069]
17. Wang S, Zha Y, Li W, Wu Q, Li X, Niu M, et al. A fully automatic deep learning system for COVID-19 diagnostic and prognostic analysis. Eur Respir J 2020 Aug;56(2):2000775 [FREE Full text] [doi: 10.1183/13993003.00775-2020] [Medline: 32444412]

18. Hasan AM, Al-Jawad MM, Jalab HA, Shaiba H, Ibrahim RW, Al-Shamasneh AR. Classification of Covid-19 Coronavirus, Pneumonia and Healthy Lungs in CT Scans Using Q-Deformed Entropy and Deep Learning Features. Entropy (Basel) 2020 May 01;22(5):517 [FREE Full text] [doi: 10.3390/e22050517] [Medline: 33286289]
19. Pathak Y, Shukla P, Tiwari A, Stalin S, Singh S, Shukla P. Deep Transfer Learning Based Classification Model for COVID-19 Disease. Ing Rech Biomed 2020 May 20 [FREE Full text] [doi: 10.1016/j.irbm.2020.05.003] [Medline: 32837678]
20. Ardakani AA, Kanafi AR, Acharya UR, Khadem N, Mohammadi A. Application of deep learning technique to manage COVID-19 in routine clinical practice using CT images: Results of 10 convolutional neural networks. Comput Biol Med 2020 Jun;121:103795 [FREE Full text] [doi: 10.1016/j.compbiomed.2020.103795] [Medline: 32568676]
21. Li L, Qin L, Xu Z, Yin Y, Wang X, Kong B, et al. Using Artificial Intelligence to Detect COVID-19 and Community-acquired Pneumonia Based on Pulmonary CT: Evaluation of the Diagnostic Accuracy. Radiology 2020 Aug;296(2):E65-E71 [FREE Full text] [doi: 10.1148/radiol.2020200905] [Medline: 32191588]
22. Mei X, Lee H, Diao K, Huang M, Lin B, Liu C, et al. Artificial intelligence-enabled rapid diagnosis of patients with COVID-19. Nat Med 2020 Aug;26(8):1224-1228 [FREE Full text] [doi: 10.1038/s41591-020-0931-3] [Medline: 32427924]
23. Song J, Wang H, Liu Y, Wu W, Dai G, Wu Z, et al. End-to-end automatic differentiation of the coronavirus disease 2019 (COVID-19) from viral pneumonia based on chest CT. Eur J Nucl Med Mol Imaging 2020 Oct;47(11):2516-2524 [FREE Full text] [doi: 10.1007/s00259-020-04929-1] [Medline: 32567006]
24. Zoph B, Vasudevan V, Shlens J, Le QV. Learning Transferable Architectures for Scalable Image Recognition. 2018 Presented at: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition; 18-23 June 2018; Salt Lake City, UT. [doi: 10.1109/cvpr.2018.00907]
25. Aria M. COVID-19 Lung CT Scans: A large dataset of lung CT scans for COVID-19 (SARS-CoV-2) detection. Kaggle. URL: https://www.kaggle.com/mehradaria/covid19-lung-ct-scans [accessed 2021-04-20]
26. Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, et al. A survey on deep learning in medical image analysis. Med Image Anal 2017 Dec;42:60-88. [doi: 10.1016/j.media.2017.07.005] [Medline: 28778026]
27. Patro SK, sahu KK. Normalization: A Preprocessing Stage. Int Adv Res J Sci Eng Technol 2015 Mar 20;2(3):20-22. [doi: 10.17148/IARJSET.2015.2305]
28. Frühwirth-Schnatter S. Data Augmentation and Dynamic Linear Models. J Time Series Analysis 1994 Mar;15(2):183-202. [doi: 10.1111/j.1467-9892.1994.tb00184.x]
29. Shorten C, Khoshgoftaar TM. A survey on Image Data Augmentation for Deep Learning. J Big Data 2019 Jul 6;6(1). [doi: 10.1186/s40537-019-0197-0]
30. Soares E, Angelov P, Biaso S, Froes M, Abe D. SARS-CoV-2 CT-scan dataset: A large dataset of real patients CT scans for SARS-CoV-2 identification. medRxiv. Preprint posted online May 14, 2020.
31. Jaiswal A, Gianchandani N, Singh D, Kumar V, Kaur M. Classification of the COVID-19 infected patients using DenseNet201 based deep transfer learning. J Biomol Struct Dyn 2020 Jul 03:1-8. [doi: 10.1080/07391102.2020.1788642] [Medline: 32619398]
32. Panwar H, Gupta PK, Siddiqui MK, Morales-Menendez R, Singh V. Application of deep learning for fast detection of COVID-19 in X-Rays using nCOVnet. Chaos Solitons Fractals 2020 Sep;138:109944. [doi: 10.1016/j.chaos.2020.109944] [Medline: 32536759]
33. Yazdani S, Minaee S, Kafieh R, Saeedizadeh N, Sonka M. COVID CT-Net: Predicting Covid-19 From Chest CT Images Using Attentional Convolutional Network. arXiv. Preprint posted online September 10, 2020 [FREE Full text] [doi: 10.1016/j.media.2020.101794]
34. Deep CNN-Based CAD System for COVID-19 Detection Using Multiple Lung CT Scans. GitHub. URL: https://github.com/MehradAria/COVID-19-CAD [accessed 2021-04-16]
35. Wang Z, Liu Q, Dou Q. Contrastive Cross-Site Learning With Redesigned Net for COVID-19 CT Classification. IEEE J Biomed Health Inform 2020 Oct;24(10):2806-2813. [doi: 10.1109/JBHI.2020.3023246] [Medline: 32915751]
36. Milletari F, Navab N, Ahmadi SA. V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. 2016 Presented at: 2016 Fourth International Conference on 3D Vision (3DV); 25-28 October 2016; Stanford, CA. [doi: 10.1109/3dv.2016.79]
37. Li X, Zhou Y, Du P, Lang G, Xu M, Wu W. A deep learning system that generates quantitative CT reports for diagnosing pulmonary Tuberculosis. Appl Intell 2020 Nov 26. [doi: 10.1007/s10489-020-02051-1]
38. Ghaderzadeh M, Asadi F. Deep Learning in the Detection and Diagnosis of COVID-19 Using Radiology Modalities: A Systematic Review. J Healthc Eng 2021;2021:6677314 [FREE Full text] [doi: 10.1155/2021/6677314] [Medline: 33747419]

Abbreviations

CNN: convolutional neural network
CAD: computer-aided detection
CT: computed tomography
NASNet: neural search architecture network
GGO: ground-glass opacities

RT–PCR: reverse transcription–polymerase chain reaction
WHO: World Health Organization

Edited by C Basch; submitted 26.01.21; peer-reviewed by D Huang, N Ramezanghorbani, A Bordbar, S Almasi; comments to author 23.02.21; revised version received 26.02.21; accepted 03.04.21; published 26.04.21.

Please cite as:
Ghaderzadeh M, Asadi F, Jafari R, Bashash D, Abolghasemi H, Aria M
Deep Convolutional Neural Network–Based Computer-Aided Detection System for COVID-19 Using Multiple Lung Scans: Design and Implementation Study
J Med Internet Res 2021;23(4):e27468
URL: https://www.jmir.org/2021/4/e27468
doi:10.2196/27468
PMID:33848973

Original Paper

Healthfulness Assessment of Recipes Shared on Pinterest: Natural Language Processing and Content Analysis

Xiaolu Cheng[1*], PhD; Shuo-Yu Lin[1*], MS; Kevin Wang[1]; Y Alicia Hong[1], PhD; Xiaoquan Zhao[2], PhD; Dustin Gress[1], MS; Janusz Wojtusiak[1], PhD; Lawrence J Cheskin[3], MD; Hong Xue[1], PhD

[1]Department of Health Administration and Policy, College of Health and Human Services, George Mason University, Fairfax, VA, United States

[2]Department of Communication, College of Humanities and Social Sciences, George Mason University, Fairfax, VA, United States

[3]Department of Nutrition and Food Studies, College of Health and Human Services, George Mason University, Fairfax, VA, United States

[*]these authors contributed equally

Corresponding Author:
Hong Xue, PhD
Department of Health Administration and Policy
College of Health and Human Services
George Mason University
4400 University Dr
Fairfax, VA, 22030
United States
Phone: 1 703 993 9833
Email: hxue4@gmu.edu

Abstract

Background: Although Pinterest has become a popular platform for distributing influential information that shapes users' behaviors, the role of recipes pinned on Pinterest in these behaviors is not well understood.

Objective: This study aims to explore the patterns of food ingredients and the nutritional content of recipes posted on Pinterest and to examine the factors associated with recipes that engage more users.

Methods: Data were collected from Pinterest between June 28 and July 12, 2020 (207 recipes and 2818 comments). All samples were collected via 2 new user accounts with no search history. A codebook was developed with a raw agreement rate of 0.97 across all variables. Content analysis and natural language processing sentiment analysis techniques were employed.

Results: Recipes using seafood or vegetables as the main ingredient had, on average, fewer calories and less sodium, sugar, and cholesterol than meat- or poultry-based recipes. For recipes using meat as the main ingredient, more than half of the energy was obtained from fat (277/490, 56.6%). Although the most followed pinners tended to post recipes containing more poultry or seafood and less meat, recipes with higher fat content or providing more calories per serving were more popular, having more shared photos or videos and comments. The natural language processing–based sentiment analysis suggested that Pinterest users weighted *taste* more heavily than *complexity* (225/2818, 8.0%) and *health* (84/2828, 2.9%).

Conclusions: Although popular pinners tended to post recipes with more seafood or poultry or vegetables and less meat, recipes with higher fat and sugar content were more user-engaging, with more photo or video shares and comments. Data on Pinterest behaviors can inform the development and implementation of nutrition health interventions to promote healthy recipe sharing on social media platforms.

(J Med Internet Res 2021;23(4):e25757) doi:10.2196/25757

KEYWORDS

healthfulness assessment; recipes on Pinterest; social networks; natural language processing

Introduction

Background

Healthy eating patterns and their effect on disease prevention have been demonstrated robustly across the scientific literature [1-9]. The US government has dedicated enormous resources to improve Americans' eating patterns through programs such as the National School Lunch program and the Special Supplemental Nutrition Program for Women, Infants, and Children. In 2015, the Dietary Guidelines Advisory Committee published the 2015-2020 Dietary Guidelines for Americans, providing guidance for choosing a healthy diet. Despite such efforts, between 2003 and 2016, although the intake of sugar by Americans decreased by 4.8 teaspoons per day, no appreciable changes occurred in the intake of vegetables, total meat, poultry, and seafood [10]. Although from 2003-2004 to 2015-2016, Americans increased whole grain consumption, the mean intake of grains, vegetables, and dairy continued to be lower than the Dietary Guidelines recommendations [10]. Only 42% of the US population met dietary recommendations between 2013 and 2014, and less than half of the older adults met the recommendations between 2013 and 2016 [11,12].

Social media has become a new and efficient way to distribute and consume influential information that shapes people's dietary behaviors [13,14]. Several internet-based intervention programs have been implemented to enhance individuals' knowledge of healthy eating [15,16]. With the growing popularity of social media, there is an urgent need to assess the contents of healthy food and nutrition information on social media and their associations with user engagement among both posters and information seekers.

Pinterest, launched in 2010, is a unique social media platform where users can save images (*pins*) and upload it to the board (a collection of pins from different users) [17]. It has also become a popular social media site for users to share recipes. According to a survey conducted by the Pew Research Center in 2018, 28% of US adults reported that they had used Pinterest [18], and over 60% of active users made a new recipe inspired by Pinterest in 2015 [19]. Pinterest provides us with a platform and unprecedented opportunities to study the effect of social media on dietary behaviors and, consequently, public health [20]. This study aims to examine the patterns in which nutrition information on recipes is received and shared among Pinterest users and identify the key elements of recipes that influence the perceptions and preferences of Pinterest users. Our findings will shed light on future social media–based dietary intervention program design and implementation.

Objectives

To the best of our knowledge, no prior study has evaluated the recipe content on Pinterest. This study provides a first glimpse of this domain to advance the understanding of the relationship between social media use and dietary behavior. We aim to achieve the following 2 goals. First, we aim to examine the patterns of food ingredients and nutrients prescribed by recipes posted on Pinterest. Second, by employing both traditional content analysis and a natural language processing (NLP) technique, we sought to understand the factors that distinguish the most popular recipes among users.

Methods

Data Collection

Data were collected between June 28 and July 12, 2020. Although there is no "rule of thumb" on how long the data collection should persist, we adapted a proper time frame based on previous literature that specifically focused on Pinterest [21,22]. All samples were collected using the Pinterest search engine by 2 new user accounts with no search history, no posts and boards or pins, 0 followers, and 0 following. The keywords *recipe*, *breakfast*, *lunch*, and *dinner* were used to identify samples on Pinterest. Pins with recipes containing all the required information were selected by scrolling down the search results page for each keyword. This approach was developed based on previous studies on Pinterest content [21,22]. Pins that were duplicates or missing any of the following information were excluded: eating occasion, cooking method, cooking time, ingredients, and nutrition information. A total of 207 collected pins or boards that met our criteria and all comments (2818 comments) under the 207 recipes were included in the analysis. A codebook was developed in an Excel spreadsheet (Microsoft Corp) to document the URL, time of data collection, comments, number of replied photos and videos, poster's number of followers, eating occasion, cooking method, cooking time, ingredients, and nutrition information.

Data Analysis

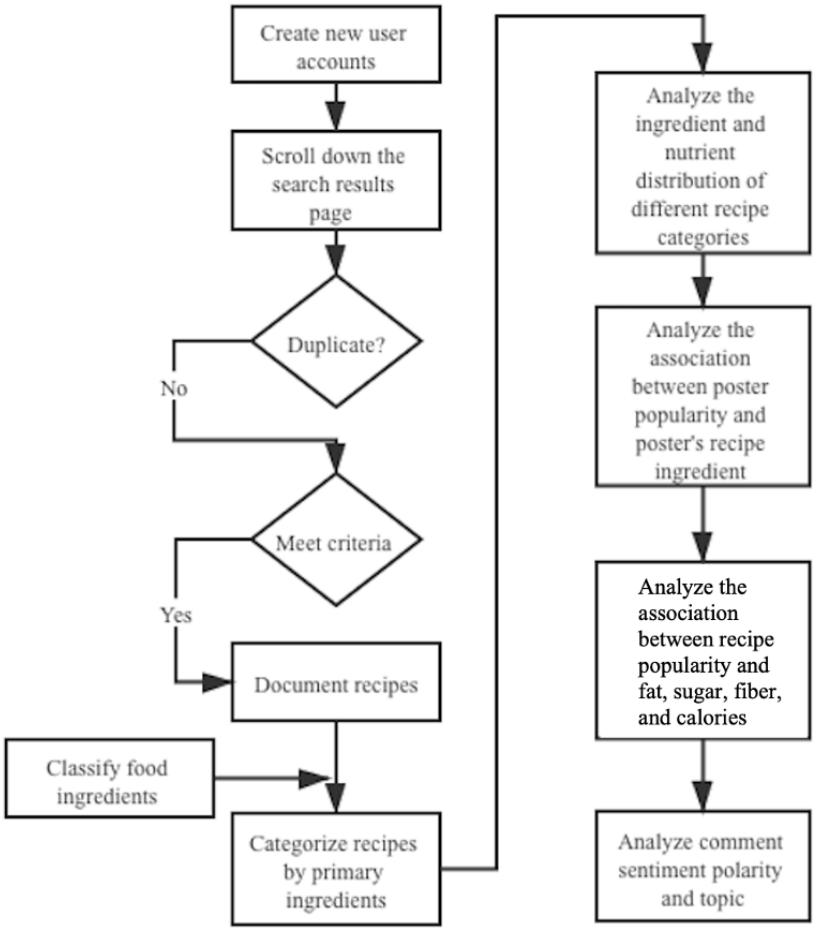
For the content analysis, following the 2015-2020 United States Department of Agriculture Dietary Guidelines, food ingredients were classified as dark green vegetables, red and orange vegetables, legumes (beans and peas), starchy vegetables, other vegetables, fruits, seafood, meats, poultry, eggs, nuts or seeds or soy products, dairy, oil, and butter [23]. The recipes were then categorized into 4 types based on their primary ingredients: meat, poultry, seafood, and vegetable recipes. Recipes that only contained meat, poultry, or seafood were further categorized according to whether they contained any vegetables. All measurement units were converted into grams per serving. The variables measured at the nutrient level included total energy (calories per serving); sodium (mg per serving); and, in grams per serving, fat, protein, carbohydrate, fiber, sugar, and cholesterol. Each pinner's number of followers was classified based on a tertile distribution. Overall, 2 coders independently analyzed all study samples and performed cross-checks to ensure intercoder reliability. The raw agreement rate was 0.97 for all variables.

For the comment analysis, 3 keyword dictionaries were created with keywords related to health (eg, *health*, *healthy*, *calorie*, and *cholesterol*), taste (eg, *yummy*, *delicious*, *tasty*, and *creamy*), and the complexity of the recipe (eg, *quick*, *easy*, *sample*, and *difficult*). Keyword searching was applied to assess sentiment of comments posted by Pinterest users.

Descriptive analyses were performed for each type of food ingredient and their corresponding nutrient content. In addition, the popularity of recipe ingredients was assessed by the number

of recoded followers (presented in tertiles). The level of engagement for each recipe was also evaluated by categorizing comments and shared photos or videos into tertiles, with regard to the fat, sugar, and fiber content of the recipes. Comments and shared photos or videos were chosen as indicators of engagement based on prior literature that suggested that, in the context of Pinterest, the number of *likes* on each pin indicates relatively low engagement (users simply acknowledge or agree with content), whereas the number of comments indicates medium engagement (users created and shared such content) [24]. High engagement is indicated by actual offline participation and can be captured by users' shared photo or videos (images of what they made based on the same recipes) [24]. To process and analyze natural language data from the comments, both keyword search and sentiment analysis technology were applied. The sentiment analysis method we used was VADER (valence aware dictionary and sentiment reasoner). VADER performs sentiment analysis on textual data to determine whether the sentiment of a text is positive, negative, or neutral. VADER is a lexicon- and rule-based sentiment analysis method. It was developed specifically to analyze the sentiment of English text in microblog-like social media [25]. VADER requires no training data and provides high-speed analysis [26]. VADER helped us categorize all comments as being positive, neutral, or negative by polarity, that is, the representation of sentiment. All statistical analyses were conducted using STATA 15.1 (StataCorp LP) and Python 3.6 (Python Software Foundation). Figure 1 shows the entire data collection and analysis process.

Figure 1. Process of data collection and analysis.



Results

Summary Statistics

Table 1 describes the quantity of food ingredients and nutrients served by meat, poultry, seafood, and vegetable recipes. Although not statistically significant, 7 patterns were identified in the data. First, we found that meat, poultry, and seafood weighed more in recipes that only served these main dishes than those that also contained vegetables. For example, 124.79 g meat was served in meat-only recipes compared with 100.65 g in meat recipes that also contained vegetables. Second, recipes containing vegetables provided more total energy than recipes served without vegetables. For instance, poultry recipes that contained vegetables provided an average of 442.79 calories per serving compared with 433.31 calories provided by poultry-only recipes. Third, seafood recipes (329.38 calories per serving) or vegetable recipes (293.67 calories per serving) provided lower total energy content compared with meat- or poultry-based recipes (430-490 calories per serving). Fourth, 48.6% (230/473) and 56.5% (277/490) of the total energy in meat-only and meat-with-vegetable recipes came from fat, higher than the other types of recipes. Fifth, meat recipes, in general, contained higher fat (25.62-30.13 g per serving) than poultry- and seafood-based recipes (16.79-22.35 g per serving).

Sixth, seafood recipes tended to contain less sodium and sugar than meat and poultry recipes. Seventh, 38.2% (112/293) of total energy in vegetable-based recipes came from carbohydrate, with less sodium (483.63 mg per serving) and cholesterol (47.65 g per serving), compared with other types of recipes.

Table 1. The ingredient and nutrient distributions of meat, poultry, seafood, and vegetable recipes.

Recipes	Meat[a]		Poultry[a]		Seafood[a]		Vegetable[a,b] with eggs[c] (n=59), mean (SD)
	Meat only (n=25), mean (SD)	Meat with vegetable[c] (n=45), mean (SD)	Poultry only (n=35), mean (SD)	Poultry with vegetable[c] (n=76), mean (SD)	Seafood only (n=6), mean (SD)	Seafood with vegetable[c] (n=13), mean (SD)	
Food ingredients (g per serving)[d]							
Dark vegetable	N/A[e]	75.6 (0)	N/A	49.6 (14.1)	N/A	29.5 (15.4)	113.3 (0)
Red and other vegetable	N/A	90.6 (189.8)	N/A	29.5 (21.3)	N/A	N/A	72.5 (88.6)
Legumes and beans	N/A	42.5 (0)	N/A	81.5 (35.1)	N/A	N/A	70.8 (0)
Starchy vegetable	N/A	132.3 (26.8)	N/A	81.5 (35.1)	N/A	N/A	144.7 (129.2)
Meat	124.8 (128.5)	100.7 (98.2)	N/A	N/A	N/A	N/A	N/A
Poultry	N/A	N/A	106.7 (94.5)	100.4 (88.9)	N/A	N/A	N/A
Seafood	N/A	N/A	N/A	N/A	122.8 (66.2)	118.4 (73.0)	N/A
Eggs	51.4 (111.0)	51.4 (111.0)	14.5 (5.7)	16.6 (6.3)	N/A	N/A	28.3 (23.4)
Total energy (calories per serving)	490.9 (280.1)	473.8 (246.8)	433.3 (225.5)	442.7 (185.7)	320.6 (126.5)	329.3 (178.5)	293.6 (131.6)
Nutrients (g per serving)							
Fat	30.1 (17.9)	25.6 (17.4)	22.3 (13.8)	20.9 (13.0)	18.5 (9.6)	16.7 (13.7)	14.8 (9.7)
Percentage of energy from fat (%)[f]	56.6	48.5	46.7	41.8	50.5	40.4	44.8
Protein	26.4 (11.9)	28.5 (19.2)	32.6 (19.5)	36.8 (50.2)	23.5 (14.9)	27.1 (16.6)	12.3 (10.0)
Percentage of energy from protein (%)[f]	24.2	25.2	30.9	34.9	30.7	33.1	16.4
Carbohydrates	24.7 (17.7)	27.3 (20.0)	20.9 (24.2)	27.1 (24.6)	12.6 (17.1)	16.3 (17.1)	29.1 (24.1)
Percentage of energy from carbohydrates (%)[f]	22.7	25.6	17.9	23.8	14.5	24.9	38.3
Fiber	2.2 (2.)	3.8 (7.6)	1.3 (1.2)	2.9 (3.6)	0.8 (0.7)	1.7 (2.1)	4.0 (5.1)
Sodium (mg per serving)	861.2 (655.9)	775.6 (683.3)	911.0 (510.6)	767.0 (559.5)	489.8 (431.6)	777.3 (659.1)	483.6 (389.1)
Sugar	6.5 (12.9)	6.7 (11.6)	5.7 (8.0)	6.6 (7.9)	1.3 (1.0)	3.5 (4.2)	4.73 (7.81)
Cholesterol	114.6 (135.5)	97.2 (113.1)	126.5 (71.4)	109.3 (69.8)	98.5 (100.6)	152.9 (177.2)	47.6 (66.5)

[a]Recipes that included a main dish only and those that included a main dish served with vegetables were mutually exclusive. For example, a meat-only recipe was defined as a recipe that only included meat, whereas recipes that included both meat and vegetables were listed in the meat with vegetable category.

[b]No recipes were purely vegan; therefore, we reported ovo-lacto recipes.

[c]The calculation of sample average used only complete data, that is, some of the denominators were smaller than 45 and did not have standard errors.

[d]The food ingredients were categorized based on the guidelines of the United States Department of Agriculture.

[e]N/A: not applicable.

[f]The sum of column percentages of each recipe class may exceed 100% because each value was calculated separately as the percentage of energy from specific nutrients divided by the total energy provided in a recipe class.

Relationship Between Popularity and Ingredients of Shared Recipes

The bar charts in Multimedia Appendix 1 describe the relationship between the pinner's popularity and the ingredients of their shared recipes. Popularity was measured as the number of followers per pinner, stratified by tertile. Pinners in third tertile had more followers. Although not statistically significant, 3 patterns emerged. First, as the number of followers increased, the amount of meat served in the recipe decreased. Second, recipes with greater followings contained more poultry and seafood but fewer red and other vegetables. Third, starchy vegetables were distributed similarly, regardless of the number of followers.

Relationship Between Photo or Video and Comments Sharing and Recipe Features

Multimedia Appendix 2 presents the mean number of shared photos or videos and comments under the recipes according to the absolute amount of fat per serving in the recipes, classified in tertiles. Recipes in the third tertile had the highest fat content. Data showed that more photos or videos and comments were shared as the absolute amount of fat per serving increased. Although not statistically significant, among the 172 recipes with complete information, the average number of shared photos or videos and comments increased from 25 to 35 between the first and third tertiles, suggesting that recipes with higher fat content, such as creamy garlic butter chicken, were more popular among Pinterest users.

Multimedia Appendix 2 also presents the mean number of shared photos or videos and comments under recipes based on the recipes' sugar content. Among 166 recipes that had complete sugar information, the mean number of photos or videos and comments were distributed as an inverse *U* shape. Although not statistically significant, this pattern indicated an upward trend in the number of shared photos or videos and comments received (between the first and second tertiles) when a recipe contained high sugar content, but the trend turned down when the recipe's sugar level reached the third tertile. In terms of sugar, our data suggested that the number of comments was similar regardless of the fiber content. However, although not statistically significant, recipes containing the highest amount of fiber (third tertile) were less popular, having fewer shared photos and videos compared with those in the first or second tertile.

Multimedia Appendix 2 further presents the mean number of shared photos or videos and comments based on the recipes' number of calories per serving. Data suggested that the number of shared photos and videos and the number of comments were positively correlated with the number of calories per serving. Recipes that provided more calories per serving (in the second and third tertiles) were more popular than recipes that provided fewer calories per serving (in the first tertile), with fewest shared photos or videos and comments.

Comment Analysis

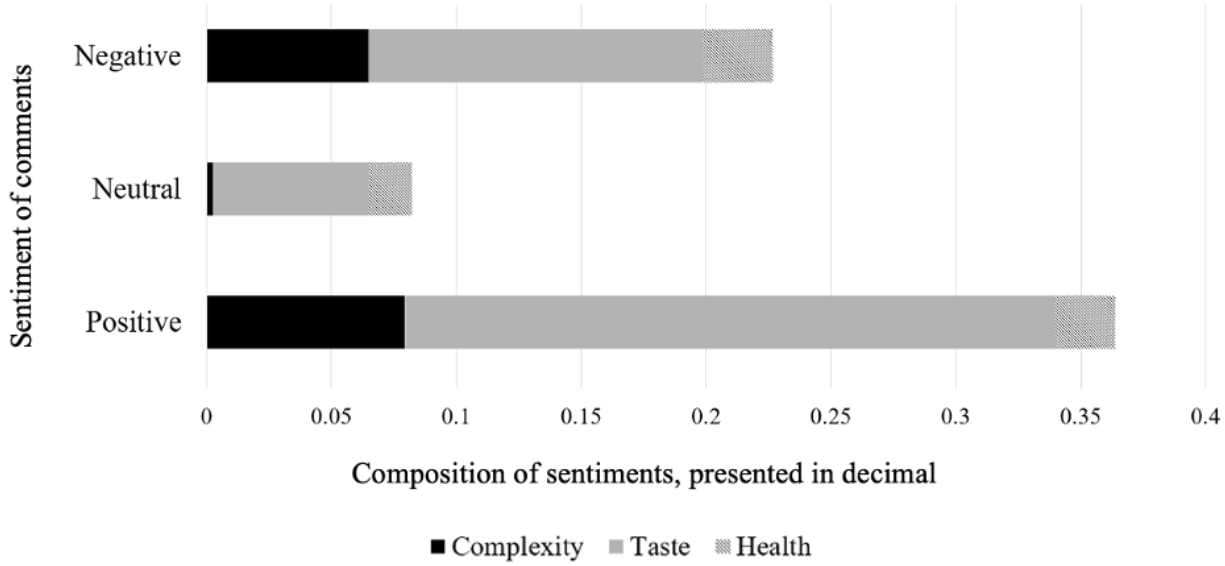
Table 2 shows sample comments, corresponding sentiment polarity, and related topic. Figure 2 shows the results of text mining from all 2818 comments analyzed. Out of 544 comments deemed as *taste* related, 25.9% (141/544) were positive, significantly higher than negative (72/544, 13.2%) and neutral (33/544, 6.1%; $P<.05$). The complexity of a recipe and its health attributes were commented on less frequently: less than 8% (225/2818) of the comments contained text related to complexity, and less than 3% (84/2828) of the comments contained text related to health. We found that *taste* and *complexity* were the most important factors in shaping Pinterest users' sentiments.

Table 2. Comment samples and polarity.

Comment	Sentiment polarity	Keyword	Topic
"How long do I leave them in the oven?"	Neutral	N/A[a]	N/A
"How many calories is this?"	Neutral	Calories	Health
"I do not like brown sugar in my meatloaf...ugh"	Negative	N/A	N/A
"Definitely way too salty and too greasy for me."	Negative	Salty and greasy	Taste
"It's easy! I did this again and LOVED it!"	Positive	Easy	Complexity
"It turned out amazing!! Very delicious."	Positive	Delicious	Taste

[a]N/A: not applicable.

Figure 2. Pinterest users' attitudes toward different aspects of recipes.



Discussion

Principal Findings

In this study, recipes posted on Pinterest were collected, analyzed, and compared for ingredients and nutrients. We found that, in most cases, recipes using seafood or vegetables as the main ingredient, for example, tuna salad (main ingredient: tuna; other ingredients: celery, onion, flat-leaf parsley, mayonnaise, mustard, and black pepper), had, on average, fewer calories and less sodium, sugar, and cholesterol than meat- or poultry-based recipes, for example, crispy chicken wraps (main ingredient: popcorn chicken; other ingredients: tomatoes, cheddar cheese, buffalo wing sauce, and flour tortillas) and Mongolian beef (main ingredient: flank steak; other ingredients: cornstarch, canola oil, ginger, garlic, soy sauce, dark brown sugar, and scallions). Recipes using meat as the main ingredient, for example, creamy herbed pork chops (main ingredient: pork chops; other ingredients: milk, Montreal steak sauce, butter, flour, basil, black pepper, and instant beef bouillon granules), provided more energy by fat. Although the most followed pinners tended to post recipes containing more poultry or seafood and less meat, recipes serving higher fat or providing more calories per serving were more popular, having more shared photos or videos and comments. Sentiment analysis based on text mining showed that Pinterest users, in general, valued taste more than health qualities when making comments or sharing photos or videos.

With the sharp increase in the number of social media users, platforms such as Pinterest have become influential mechanisms to transform knowledge sharing and acquisition, including dietary choice [27]. According to a survey conducted by the Pew Research Center in 2018, about 28% of US adults used Pinterest [18]. Although evidence has shown that intervention tactics through tailored web-based platforms can help promote evidence-based nutrition education to the public [28,29], members of the academic community have urged for more research evaluating the mutual influence between social media users and information or content providers [30]. Our study provides the first glimpse into how recipe information is disseminated and viewed from 2 distinct perspectives: Pinterest posters (pinners) and users.

From the perspective of content providers, we found that the most popular pinners, by sharing recipes containing more seafood and poultry (Multimedia Appendix 1) with less sodium, sugar, and cholesterol, are overall more health conscious. This finding can be understood in terms of the social cognitive theory. The social cognitive theory states that a human is an agent that has not only been a forethinker but also a motivator and self-regulator. In a sense, humans learn by observing others' actions and their consequences [31]. In our case, pinners are more likely to imitate posts that are socially rewarded. Early research on popular food blogs (webpages that can be pinned to Pinterest) aligned with our findings, suggesting that vegetarian and seafood recipes had significantly lower nutrition risks and more health benefits compared with red meat and poultry recipes [32]. Many health-conscious *social elites* do not eat meat at all; they only eat vegetables or seafood. It is not surprising that celebrities or social influencers, such as those with many followers on Pinterest, embrace this *elite social norm*, considering that most Pinterest users are from high-income households [18], which reinforced the role of self-reactiveness portrayed in the social cognitive theory [31]. In theory, an agent acts intentionally [31]. Popular pinners, such as celebrities, might be more health orientated and have implicitly or explicitly engaged in education, inspiration, and activism—the 3 stages of celebrity narratives—when they post recipes on Pinterest [33]. Red meat, for example, is classified as probably carcinogenic to humans (group 2A) by the Working Group of the International Agency for Research on Cancer [34] and is associated with type 2 diabetes, cardiovascular diseases, malignancies, and other diseases [35,36]. Such joint activity requires commitment to a shared intention [31,37]. Pinners may or may not be aware of this fact, but by posting recipes containing less red meat, they may have contributed to shifting public dietary choices to a healthier direction.

From the users' perspective, they are often learners in pursuit of inspirational recipes [38]. We found that recipes with higher

fat and sugar content tended to generate higher user engagement and greater numbers of shared photos or videos and comments. High engagement refers to actual offline use of the recipe rather than simply clicking *like* on a particular pin [24]. The social cognitive theory postulates that people are more motivated when they consider a subject worthwhile [37]. Our sentiment analysis corroborated the theory and literature by showing that users attached more importance to the taste of a recipe than its healthfulness and complexity. The social cognitive theory of mass communication also shows that behavior changes of an agent can be directly affected by media and indirectly influenced by connections to social systems that are diffused by media. The diffusion process relies heavily on the social-prompting power of the modeling [31]. Previous qualitative research showed that learners often quickly assumed the role of an expert or teacher when sharing nutrition information from social media [38]; therefore, it is concerning that ordinary people paid more attention to taste and were motivated by fat- and sugar-heavy pins. A likely downstream effect of these preferences of ordinary people is that *tasty* recipes will be disseminated quickly on social media through users' social networks (a part of the social system) and make their way into regular recipe rotations for more people [39]. Our findings suggest a high priority area for future social media–based nutrition interventions.

There appears to be a discrepancy between what pinners posted and how users consumed information, leading to an opportunity for future health interventions via Pinterest. Previous studies have shown that social media interventions can have a positive effect on nutritional outcomes [14]. Strategies to increase users' health consciousness can include, but are not limited to, (1) encouraging pinners to provide healthier (low sugar and cholesterol) alternative ingredients, (2) promoting recipes provided by health professionals and supported by evidence-based research [40,41], and (3) designing and promoting healthy recipes that are tasty and easy to prepare.

Limitations

This study had some limitations. First, because of the restrictions imposed by Pinterest, the content scrolling process is not automated. The manual data collection resulted in a relatively small sample size and a large margin of error. To address the issues related to the small sample size, we applied a machine learning technique to mine text from the comments. A total of 100 comments were randomly selected to assess sentiment error rates. We found that the error rate was 18%, which is better than the acceptable level used in previous studies by convention [42]. Second, only recipes posted in English were included. Thus, the sample was not representative of non–English-speaking cultures or users. Third, the measurement of healthfulness was assessed based on the types of food ingredients and amount of fat, sugar, and fiber; other aspects of health, such as cooking methods, were not included. Future research should incorporate these aspects. Finally, demographic information such as gender or race and ethnicity of the Pinterest users was unavailable from Pinterest. Our sample was restricted to those who could adopt the food culture embedded in Pinterest. As seafood- or vegetable-only recipes are often more expensive or beyond the reach of low-income populations, more research is needed to address the potential socioeconomic disparities inherent in popular social media platforms.

Conclusions

In this study, we used both content analysis and NLP techniques to analyze recipes posted on Pinterest. Seafood-based recipes and vegetarian recipes had fewer calories and less sodium, sugar, and cholesterol than meat-based recipes. Although the most popular pinners tended to exhibit more health consciousness by posting recipes with more seafood, poultry, and vegetables and less meat, recipes with higher fat and sugar content had higher user engagement, as demonstrated by the higher numbers of photo or video shares and comments. Population health could be improved with targeted interventions to address this disparity through efforts to enhance interest in and adoption of healthy recipes by Pinterest users.

Authors' Contributions

HX, XC, and SL made significant contributions to the conception and study design. HX, XC, and SL conducted data analyses. HX, KW, AH, DG, XZ, JW, and LC provided significant support for the interpretation of results. HX, XC, SL, KW, AH, DG, XZ, JW, and LC drafted the paper. All authors approved the final manuscript submitted.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Relationship between pinners' popularity and their recipe ingredients.
[PNG File , 141 KB - jmir_v23i4e25757_app1.png]

Multimedia Appendix 2
Association between popularity and recipes.
[PNG File , 106 KB - jmir_v23i4e25757_app2.png]

References

1. Zarrinpar A, Chaix A, Panda S. Daily eating patterns and their impact on health and disease. Trends Endocrinol Metab 2016 Mar;27(2):69-83 [FREE Full text] [doi: 10.1016/j.tem.2015.11.007] [Medline: 26706567]
2. St-Onge M, Ard J, Baskin ML, Chiuve SE, Johnson HM, Kris-Etherton P, American Heart Association Obesity Committee of the Council on LifestyleCardiometabolic Health; Council on Cardiovascular Disease in the Young; Council on Clinical Cardiology;Stroke Council. Meal timing and frequency: implications for cardiovascular disease prevention: a scientific statement from the American Heart Association. Circulation 2017 Mar 28;135(9):e96-e121. [doi: 10.1161/CIR.0000000000000476] [Medline: 28137935]
3. Hutto CJ, Gilbert E. VADER: a parsimonious rule-based model for sentiment analysis of social media text. In: Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media. 2015 Presented at: Eighth International AAAI Conference on Weblogs and Social Media; January 2015; Ann Arbor, MI.
4. Kåhrström CT, Pariente N, Weiss U. Intestinal microbiota in health and disease. Nature 2016 Jul 07;535(7610):47. [doi: 10.1038/535047a] [Medline: 27383978]
5. Owen L, Corfe B. The role of diet and nutrition on mental health and wellbeing. Proc Nutr Soc 2017 Jul 14;76(4):425-426. [doi: 10.1017/s0029665117001057]
6. Zhao Z, Yin Z, Zhao Q. Red and processed meat consumption and gastric cancer risk: a systematic review and meta-analysis. Oncotarget 2017 May 02;8(18):30563-30575 [FREE Full text] [doi: 10.18632/oncotarget.15699] [Medline: 28430644]
7. Damjanovska G, Severova G, Cakalaroski K, Antovska-Knight V, Danilovska I, Simovska V, et al. Beneficial short term effect of low protein diet on chronic kidney disease pro-gression in patients with chronic kidney disease stage G3a. A pilot study. Hippokratia 2018;22(4):178-182 [FREE Full text] [Medline: 31695305]
8. Hussain K, Murdin L, Schilder AG. Restriction of salt, caffeine and alcohol intake for the treatment of Ménière's disease or syndrome. Cochrane Database Syst Rev 2018 Dec 31;12 [FREE Full text] [doi: 10.1002/14651858.CD012173.pub2] [Medline: 30596397]
9. Yu E, Malik VS, Hu FB. Reprint of: cardiovascular disease prevention by diet modification: JACC health promotion series. J Am Coll Cardiol 2018 Dec 11;72(23 Pt B):2951-2963 [FREE Full text] [doi: 10.1016/j.jacc.2018.10.019] [Medline: 30522630]
10. Bowman SA, Clemens JC, Friday JE, Lynch KL, LaComb RP, Moshfegh AJ. Food patterns equivalents intakes by Americans: what we eat in America, NHANES 2003-2004 and 2013-2014. Food Surveys Research Group. 2017. URL: https://www.ars.usda.gov/ARSUserFiles/80400530/pdf/DBrief/17_Food_Patterns_Equivalents_0304_1314.pdf [accessed 2021-03-29]
11. Bowman SA, Clemens JC, Martin CL, Anand J, Steinfeldt LC, Moshfegh AJ. Added sugars intake of Americans: what we eat in America, NHANES 2013-2014. Food Surveys Research Group. 2017. URL: https://www.ars.usda.gov/ARSUserFiles/80400530/pdf/DBrief/18_Added_Sugars_Intake_of_Americans_2013-2014.pdf [accessed 2021-03-25]
12. Lois S, Carrie LM, Joseph DG, Moshfegh AJ. Meeting dietary guidelines recommendations: older adults what we eat in America, NHANES 2013-2016. Food Surveys Research Group. 2020. URL: https://www.ars.usda.gov/ARSUserFiles/80400530/pdf/DBrief/28_Meeting_Dietary_Guidelines_Recommendations_Older_Adults_1316.pdf [accessed 2021-03-25]
13. Coates AE, Hardman CA, Halford JC, Christiansen P, Boyland EJ. Social media influencer marketing and children's food intake: a randomized trial. Pediatrics 2019;143(4) [FREE Full text] [doi: 10.1542/peds.2018-2554] [Medline: 30833297]
14. Klassen KM, Douglass CH, Brennan L, Truby H, Lim MS. Social media use for nutrition outcomes in young adults: a mixed-methods systematic review. Int J Behav Nutr Phys Act 2018 Jul 24;15(1):70 [FREE Full text] [doi: 10.1186/s12966-018-0696-y] [Medline: 30041699]
15. Reicks M, Kocher M, Reeder J. Impact of cooking and home food preparation interventions among adults: a systematic review (2011-2016). J Nutr Educ Behav 2018 Feb;50(2):148-172. [doi: 10.1016/j.jneb.2017.08.004] [Medline: 28958671]
16. Adam M, Young-Wolff KC, Konar E, Winkleby M. Massive open online nutrition and cooking course for improved eating behaviors and meal composition. Int J Behav Nutr Phys Act 2015 Dec 03;12:143 [FREE Full text] [doi: 10.1186/s12966-015-0305-2] [Medline: 26630879]
17. Carlson N. Inside Pinterest: an overnight success four years in the making. 2012. URL: https://www.businessinsider.com/inside-pinterest-an-overnight-success-four-years-in-the-making-2012-4 [accessed 2021-03-25]
18. Perrin A, Anderson M. Share of U.S. adults using social media, including Facebook, is mostly unchanged since 2018. Pew Research Center. URL: https://www.pewresearch.org/fact-tank/2019/04/10/share-of-u-s-adults-using-social-media-including-facebook-is-mostly-unchanged-since-2018/ [accessed 2021-03-29]
19. Pinterest media consumption study. 2016. URL: https://www.businesswire.com/news/home/20160908005848/en/Ahalogy-Releases-2016-Pinterest-Media-Consumption-Study [accessed 2021-03-29]
20. Chau MM, Burgermaster M, Mamykina L. The use of social media in nutrition interventions for adolescents and young adults-a systematic review. Int J Med Inform 2018 Dec;120:77-91 [FREE Full text] [doi: 10.1016/j.ijmedinf.2018.10.001] [Medline: 30409348]
21. Guidry JP, Carlyle K, Messner M, Jin Y. On pins and needles: how vaccines are portrayed on Pinterest. Vaccine 2015 Sep 22;33(39):5051-5056 [FREE Full text] [doi: 10.1016/j.vaccine.2015.08.064] [Medline: 26319742]
22. Simpson CC, Mazzeo SE. Skinny is not enough: a content analysis of fitspiration on Pinterest. Health Commun 2017 May 21;32(5):560-567. [doi: 10.1080/10410236.2016.1140273] [Medline: 27326747]

23. U.S. Department of Health and Human Services and U.S. Department of Agriculture. 2015–2020 Dietary Guidelines for Americans. 8th Edition. Washington, D.C.; Dec 2015. URL: https://www.worldscientific.com/doi/abs/10.1142/9789811215094_0005
24. Paige SR, Stellefson M, Chaney BH, Alber JM. Pinterest as a resource for health information on chronic obstructive pulmonary disease (COPD): a social media content analysis. Am J Health Educ 2015 Jul 06;46(4):241-251. [doi: 10.1080/19325037.2015.1044586]
25. Hutto C, Gilbert E. VADER: a parsimonious rule-based model for sentiment analysis of social media text. In: Proceedings of the International AAAI Conference on Web and Social Media.: A parsimonious rule-based model for sentiment analysis of social media text. Proceedings of the International AAAI Conference on Web and Social Media; 2014 Presented at: Eighth International AAAI Conference on Weblogs and Social Media; June 1-4, 2014; University of Michigan.
26. Elbagir S, Yang J. Sentiment analysis on Twitter with Python's natural language toolkit and VADER sentiment analyzer. IAENG Transactions Eng Sci 2020:63-80 [FREE Full text] [doi: 10.1142/9789811215094_0005]
27. Dumitrescu D. Nonverbal communication in politics. Am Behav Sci 2016 Dec 19;60(14):1656-1675. [doi: 10.1177/0002764216678280]
28. Gibson S, Adamski M, Blumfield M, Dart J, Murgia C, Volders E, et al. Promoting evidence based nutrition education across the world in a competitive space: delivering a massive open online course. Nutrients 2020 Jan 28;12(2) [FREE Full text] [doi: 10.3390/nu12020344] [Medline: 32012993]
29. Blum ER, Stenfors T, Palmgren PJ. Benefits of massive open online course participation: deductive thematic analysis. J Med Internet Res 2020 Jul 08;22(7) [FREE Full text] [doi: 10.2196/17318] [Medline: 32672680]
30. Fung IC, Blankenship EB, Ahweyevu JO, Cooper LK, Duke CH, Carswell SL, et al. Public health implications of image-based social media: a systematic review of Instagram, Pinterest, Tumblr, and Flickr. Perm J 2020;24 [FREE Full text] [doi: 10.7812/TPP/18.307] [Medline: 31852039]
31. Bandura A. Social cognitive theory of mass communication. Media Psychol 2001 Aug;3(3):265-299. [doi: 10.1207/S1532785XMEP0303_03]
32. Schneider EP, McGovern EE, Lynch CL, Brown LS. Do food blogs serve as a source of nutritionally balanced recipes? An analysis of 6 popular food blogs. J Nutr Educ Behav 2013;45(6):696-700. [doi: 10.1016/j.jneb.2013.07.002] [Medline: 24206585]
33. Beck CS, Aubuchon SM, McKenna TP, Ruhl S, Simmons N. Blurring personal health and public priorities: an analysis of celebrity health narratives in the public sphere. Health Commun 2014;29(3):244-256. [doi: 10.1080/10410236.2012.741668] [Medline: 23548050]
34. Turesky RJ. Mechanistic evidence for red meat and processed meat intake and cancer risk: a follow-up on the International Agency for Research on Cancer Evaluation of 2015. Chimia (Aarau) 2018 Oct 31;72(10):718-724 [FREE Full text] [doi: 10.2533/chimia.2018.718] [Medline: 30376922]
35. Wang X, Lin X, Ouyang YY, Liu J, Zhao G, Pan A, et al. Red and processed meat consumption and mortality: dose-response meta-analysis of prospective cohort studies. Public Health Nutr 2016 Apr;19(5):893-905. [doi: 10.1017/S1368980015002062] [Medline: 26143683]
36. Misra R, Balagopal P, Raj S, Patel TG. Red meat consumption (Heme Iron Intake) and risk for diabetes and comorbidities? Curr Diab Rep 2018 Sep 18;18(11):100. [doi: 10.1007/s11892-018-1071-8] [Medline: 30229313]
37. Bandura A. Social cognitive theory: an agentic perspective. Annu Rev Psychol 2001;52:1-26. [doi: 10.1146/annurev.psych.52.1.1] [Medline: 11148297]
38. Adamski M, Truby H, M Klassen K, Cowan S, Gibson S. Using the internet: nutrition information-seeking behaviours of lay people enrolled in a massive online nutrition course. Nutrients 2020 Mar 12;12(3) [FREE Full text] [doi: 10.3390/nu12030750] [Medline: 32178291]
39. Tobey LN, Mouzong C, Angulo JS, Bowman S, Manore MM. How low-income mothers select and adapt recipes and implications for promoting healthy recipes online. Nutrients 2019 Mar 05;11(2) [FREE Full text] [doi: 10.3390/nu11020339] [Medline: 30764537]
40. Akhlaghi M, Ghasemi-Nasab M, Riasatian M. Mediterranean diet for patients with non-alcoholic fatty liver disease, a systematic review and meta-analysis of observational and clinical investigations. J Diabetes Metab Disord 2020 Jun;19(1):575-584 [FREE Full text] [doi: 10.1007/s40200-019-00475-2] [Medline: 32550210]
41. Abbasi J. Interest in the ketogenic diet grows for weight loss and Type 2 diabetes. J Am Med Assoc 2018 Jan 16;319(3):215-217. [doi: 10.1001/jama.2017.20639] [Medline: 29340675]
42. Roebuck K. Sentiment Analysis: High-impact Strategies - What You Need to Know: Definitions, Adoptions, Impact, Benefits, Maturity, Vendors. Queensland, Australia: Emereo Publishing; 2011:1-247.

Abbreviations

NLP: natural language processing
VADER: valence aware dictionary and sentiment reasoner

Edited by R Kukafka; submitted 13.11.20; peer-reviewed by W Pan, S Doan; comments to author 29.01.21; revised version received 01.03.21; accepted 18.03.21; published 20.04.21.

Please cite as:
Cheng X, Lin SY, Wang K, Hong YA, Zhao X, Gress D, Wojtusiak J, Cheskin LJ, Xue H
Healthfulness Assessment of Recipes Shared on Pinterest: Natural Language Processing and Content Analysis
J Med Internet Res 2021;23(4):e25757
URL: https://www.jmir.org/2021/4/e25757
doi:10.2196/25757
PMID:33877052

Original Paper

Determinants of Knowledge About Dietary Supplements Among Polish Internet Users: Nationwide Cross-sectional Study

Michał Seweryn Karbownik[1], MPharm, PhD; Robert Horne[2], MPharm, PhD; Ewelina Paul[3], MA; Edward Kowalczyk[1], MD, PhD; Janusz Szemraj[4], MSc, PhD

[1]Department of Pharmacology and Toxicology, Medical University of Lodz, Lodz, Poland

[2]Centre for Behavioural Medicine, The School of Pharmacy, University College London, London, United Kingdom

[3]Osom Studio, Lodz, Poland

[4]Department of Medical Biochemistry, Medical University of Lodz, Lodz, Poland

Corresponding Author:
Michał Seweryn Karbownik, MPharm, PhD
Department of Pharmacology and Toxicology
Medical University of Lodz
Zeligowskiego 7/9
Lodz, 90-752
Poland
Phone: 48 42 272 52 91
Email: michal.karbownik@umed.lodz.pl

Abstract

Background: An accurate understanding of dietary supplements (DS) is a prerequisite for informed decisions regarding their intake. However, there is a need for studies on this understanding among the public based on validated research tools.

Objective: This study aims to assess the knowledge about DS among Polish internet users with no medical education and to identify its determinants and design an appropriate predictive model.

Methods: The study protocol was prospectively registered with a statistical analysis plan. Polish users of a web-based health service and a social networking service were administered a survey consisting of the recently developed questionnaire on knowledge about DS, the questionnaire on trust in advertising DS, the beliefs about medicines questionnaire, and several other health-related single-item measures and sociodemographic questions. The results were subjected to general linear modeling.

Results: A total of 6273 participants were included. Of the 17 yes or no questions in the questionnaire of knowledge about DS, the mean number of correct responses was 9.0 (95% CI 8.9-9.1). Health service users performed worse than social networking users by 2.3 points (95% CI 2.1-2.5) in an analysis adjusted for potential confounders. Internet users had fewer true beliefs about DS if they presented higher trust in their advertising (adjusted β=−.37; 95% CI −.39 to −.34), used DS (adjusted β=−.14; 95% CI −.17 to −.12), experienced their positive effect (adjusted β=−.16; 95% CI −.18 to −.13), were older or younger than 35 years (adjusted β=−.14; 95% CI −.17 to −.12), expressed interest in the topic of DS (adjusted β=−.10; 95% CI −.13 to −.08), reported getting information about the products from friends (adjusted β=−.13; 95% CI −.15 to −.11), and believed that medicines are harmful (adjusted β=−.12; 95% CI −.15 to −.10). The proposed 5-predictor model could explain 31.2% of the variance in knowledge about DS. The model appeared resistant to overfitting and was able to forecast most of the observed associations.

Conclusions: Polish internet users with no medical education exhibit some false beliefs regarding DS. Trusting the advertising of DS appears to conflict with knowledge about them. There is an urgent need for effective web-based educational campaigns on DS and the promotion of advertising literacy. After the proposed predictive model is externally validated, it may help identify the least informed target audience.

(J Med Internet Res 2021;23(4):e25228) doi:10.2196/25228

KEYWORDS

dietary supplements; knowledge; beliefs; advertising; education; statistical model; health services; Poland; online social networking

Introduction

Background

Dietary supplements (DS) are consumed by approximately half of the adult population in developed countries [1-3]. They are used not only to fill potential nutrient gaps, as they are intended to, but also ostensibly to improve and maintain overall health and well-being, prevent or treat diseases, enhance cognitive and sports performance, and extend life expectancy [3-8]. The recommendations of health care providers seem to play only a minor role in the decision to use DS [4]. Instead, media, including the internet, appear to shape DS consumption patterns [5,9]. The internet is able to build a highly positive picture of DS [10-15]. In contrast, recent findings of large-scale, high-quality research studies highlight overall negligible benefits [2,16] and potential threats [2,17,18] related to DS use. It is important to evaluate the actual knowledge about DS held by the public and its determinants, not only for cognitive reasons but also for tailoring educational campaigns capable of fostering informed decisions regarding DS consumption [3,19-22].

Knowledge about DS has already been examined in numerous research studies worldwide, in multiple populations [3,23-29], including the general public [30-33]. Most participants were likely internet users [34]. Although the findings consistently reported an inadequate knowledge level, a systematic review could not draw any conclusions regarding knowledge about DS because of the heterogeneity of the data [35]. Moreover, most of the reports had multiple methodological limitations: the sample sizes were fairly low and selected with highly nonprobabilistic techniques, and the methods applied to examine knowledge about DS presented modest or vaguely described validity proofs. Apart from a general nutrition knowledge questionnaire [36] and similar measures for health care workers [28,37], no satisfactorily validated tool was available to test knowledge about DS in the general population until such a test was developed in 2019 [38]. The questionnaire exhibited acceptable and well-documented validity and was designed to screen commonly identifiable beliefs that are important from the public health perspective. This questionnaire was developed in the same country and the same language as this research was performed, thus allowing a valid assessment of knowledge about DS and its determinants on a large scale.

Objectives

This study aims to evaluate the level of knowledge about DS and identify its determinants among adult Polish internet users without medical education. The following research questions were addressed: (1) what is the level of knowledge about DS among Polish internet users? (2) what are the characteristics of the population members who are unknowledgeable about DS? and (3) how can the level of knowledge about DS be modeled in this population? These research questions were additionally addressed to some subpopulations with a high impact on public health: older people, residents of rural areas, and people with low earnings.

Methods

Ethical Considerations

The study was approved by the Bioethics Committee of the Medical University of Lodz, Poland (KE/1382/19, received on October 15, 2019). Expressing informed consent in an electronic manner was mandatory to participate in the study. The study protocol with the data analysis plan was prospectively registered on November 26, 2019, in the public repository Open Science Framework (Center for Open Science; Open Science Framework) [39]. A few changes to the preregistered study protocol were made after the study commencement, which are presented and discussed in Multimedia Appendix 1. DOZ.pl (Pelion), a web-based health service entity, was the research partner in this study. The only role of DOZ.pl was to prepare promotional materials (each one approved by the researchers) to help recruit participants and to enter the survey content (authored by the researchers with no influence of DOZ.pl) into an external web-based survey system (see section *Research Instruments*). This paper was outlined according to the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) guidelines for cross-sectional studies [40] and, to some extent, the TRIPOD (Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis) guidelines [41].

Study Design

This was a nationwide cross-sectional study conducted among Polish internet users. A self-administered survey was accessed on the web. Each participant was asked to complete the survey once.

Research Instruments

The survey was created using the Survio web-based survey system (Survio). It was pretested in a qualitative manner by 5 nonmedically educated people (3 women; mean age 52.0 years, SD 19.1 years) for readability, understanding, appearance, face validity, and time to survey completion according to the procedure described in a study by Hilton [42]. The survey was modified according to the feedback received to construct the final version.

The survey comprised 5 consecutive parts:

1. *Introduction*: the first page provided the volunteers with basic information about the purpose of the study, the structure of the survey, the rights of participants, and contact person data. This page also included a statement of informed consent, which was to be expressed electronically by ticking a Start the survey now button.
2. *Knowledge about DS*: this part included a recently developed and validated questionnaire on knowledge about DS in its original Polish language version [38]. The questionnaire consisted of 17 true-or-false statements; the respondent received a point for each correct answer with a maximum of 17 points. The questionnaire was formed of 2 subscales to examine general and specific knowledge. Knowledge about DS–general assessed “familiarity with the useful facts concerning the legal status of DS in general” [38]. It included 7 items related to DS definition, DS quality

standards, and DS package labeling. Knowledge about DS–specific assessed "familiarity with common, scientifically proven and useful facts about popular dietary supplements" [38]. It included 10 items related to the efficacy of vitamin C, vitamin D, multivitamin, calcium, magnesium, and antioxidants and their adverse effects and oral absorption. The content of each questionnaire item is presented in Multimedia Appendix 2. The sum of correct answers in knowledge about DS–general and knowledge about DS–specific formed knowledge about DS–total, which operationalized the construct of knowledge about DS.

3. *Dietary supplement advertising*: this part included a single-item measure of having contact with DS advertisements within the past week and a recently developed and validated questionnaire on trust in advertising DS in its original Polish version [38]. The tool consisted of 8 pairs of opposing expressions characterizing DS advertisements and information conveyed by them to be assessed on a 5-point semantic differential scale. The questionnaire was formed from the reliability, intelligibility, and affect subscales. The sum of scores on the subscales operationalized the construct of the trust in advertising DS.
4. *Beliefs about medicines*: this part included the Polish version [43] of the beliefs about medicines questionnaire (BMQ) [44]. The tool included in this study involved only the BMQ-General, as it may be used separately from BMQ-Specific to assess ideas about medicines in general among people who may take no medicines [44]. The BMQ-General consisted of 8 statements for a respondent to express their opinions using a 5-point Likert scale. The BMQ-General was composed of 2 separate 4-item subscales to operationalize the construct of beliefs that medicines are overused by doctors (BMQ Overuse) and the construct of beliefs that medicines are harmful, addictive, poisonous, and should not be taken continuously (BMQ Harm).
5. *Other medical and sociodemographic data*: this part included a set of single-item measures of health (4-point Likert scale), diet and physical activity (5-point semantic differential scale), use of any DS within the past 30 days (further called use of DS; 2-point Likert scale), personal experience of positive (or negative) effect of DS (further called positive [or negative] effect of DS, for DS users only; both 2-point Likert scale), interest in DS (5-point semantic differential scale), sources of getting knowledge about DS (assessed in 5 categories: medical doctors, pharmacists, dieticians, friends with no medical education, and media such as magazines, television, radio, and internet; each of the categories assessed in 4-point Likert scale), and conventional cigarette smoking and electronic cigarette use (both as 3-point Likert scales: never; no, but I smoked/used in the past; and yes). A measure of self-rated diet used in this study was found in a pretest study (a convenience sample of 117 healthy adults from the general population and medical students) to significantly correlate (Pearson r=0.48; P<.001) with the Polish version (own translation with no full validation) of the Starting the Conversation scale, a brief dietary assessment tool [45]. In the same pretest study, a measure of self-rated physical activity was found to significantly correlate with the Polish version of the International Physical Activity Questionnaire Short Form [46] (Pearson r=0.49; P<.001). Demographic data in the survey used in this study included age, sex, educational level (with 5 options to choose from), having medical education (*no* or *yes*), number of inhabitants in a place of residence, and monthly net household earnings per family member (both with 4 options to choose from).

None of the survey questions employed forced answering. In all semantic differential scales, the central value was set as the default answer. It was predicted that the survey would take 5 to 10 minutes to complete. After completing the survey, participants were provided with correct answers to the knowledge about DS questionnaire with expert comments for educational purposes. A survey could be completed only once from a single internet protocol address to avoid duplicate records.

The detailed characteristics of the survey questions and their method of operationalization are described in the study protocol [39]. The questions asked in the survey and its layout are presented in Polish (original version) and English in Multimedia Appendix 3.

Participants

The desired sample size was set to 10,000 participants. It was determined not according to the analysis of statistical power but based on the estimated ability of DOZ.pl to reach the audience. A nonprobability convenience sampling technique was used to recruit internet users. They were accessed through a web-based health service and a general social networking website between November 26, 2019, and March 11, 2020. DOZ.pl was used as the health service, whereas Wykop.pl (Wykop; modeled after the American Digg service) was the social networking service. In the 3 months from December 2019 to February 2020, when the study was conducted, DOZ.pl was the sixth most popular web-based health service and the first most popular web-based pharmacy in Poland, with a mean of 37.91 million page views and 4.39 million unique users (15.72% of internet users in Poland) per month. In the same period, Wykop.pl was the fifth-to-sixth most popular web-based social networking service in Poland, with 74.31 million page views and 4.28 million unique users (15.36% of internet users in Poland) each month [47].

Apart from the refusal of electronic informed consent, there were no specific exclusion criteria in the study. The survey could be completed by anybody who reached it on the web; however, a knowledge of Polish was needed to complete it.

Procedure

Throughout the period during which the study was conducted, promotional material encouraging participation with a link to the survey was temporarily placed in a slider on the main DOZ.pl website and some subpages, DOZ.pl social media (Facebook), and some other DOZ.pl channels. Moreover, the invitation to participate was emailed twice to the subscribers of DOZ.pl newsletter. Similarly, the promotional material with a link to the survey was posted on the Wykop.pl website to be entered and promoted by service users.

Data Analysis

Each survey record was assumed to be completed by a single respondent. Before performing the analysis, survey data were cleaned by removing records considered potentially meaningless [48] (as detailed in Multimedia Appendix 4). The number, frequency, and pattern of missing values were examined (Multimedia Appendix 5). Before any further analysis, the missing values were completed using a multiple imputation by chained equation procedure under a *missing at random* assumption about the unobserved data.

Details of the data analysis are depicted in a preregistered data analysis plan [39]. Briefly, descriptive statistics were presented, and participants' characteristics from the 2 web-based services were compared using the asymptotic Mann-Whitney *U* test, Pearson chi-square test, and general linear models (GLMs). Then, the associations between all the examined characteristics of participants and knowledge about DS–general, knowledge about DS–specific, and knowledge about DS–total were tested. The associations were examined in raw analyses and with adjustment for potential confounders (age, following its transformation; sex; education; the number of inhabitants; earnings; type of web service through which the survey was accessed; and calendar year in which a participant completed the survey). Adjustment for calendar year was performed to correct for the potential effect of "Broadcasting agreement about the rules and regulations for advertising dietary supplements," which was signed in Poland and became effective on January 01, 2020. Although the Likert and semantic differential scale data should be perceived as ordinal variables, parametric tests were used to allow for multivariate modeling with GLM. Sensitivity analysis of the associations was performed in 2 steps. First, univariate associations were tested using a corresponding nonparametric procedure (Spearman rho). Second, all parametric associations were repeated in a complete case database before data imputation. The Benjamini and Hochberg procedure was used to reduce the false discovery rate to 0.05, which was inflated by testing multiple hypotheses.

The predictive model of knowledge about DS was built using multivariate linear regression analysis. The selection of knowledge about DS predictors was based on the following criteria: first, the characteristics substantially associated with the knowledge about DS–total were preferred. Second, objective measures, which reflect the underlying constructs with strong proof of validity, were favored. Third, a set of predictors with negligible collinearity was retained. Collinearity was assessed using explanatory factor analysis and Pearson *r* correlation matrix. Data transformation was also considered. The final model was selected according to the best subset selection algorithm to promote model simplicity based on the Akaike and Bayesian information criteria. Its performance was illustrated with a calibration plot as well as mean absolute error (MAE) of prediction (the mean difference between predicted and observed knowledge about DS–total score) and root mean squared error (RMSE). The final model was internally validated using a 10-fold cross-validation procedure to correct for overfitting bias.

P lower than Benjamini-Hochberg corrected significance level or $P<.05$ were considered statistically significant. The analyses were performed using STATISTICA software version 13.3 (Statsoft) and R software version 4.0.0 (package *mice* version 3.8.0; R Foundation for Statistical Computing).

Results

Database

The survey was displayed 24,400 times and completed 7632 times (7632/24,400, 31.28% of the displayed surveys). A total of 6273 records (6273/7632, 82.19% of the completed records) were retained in the final database following the cleaning procedure, which is detailed in Multimedia Appendix 4. Missing data comprised 0.38% of the values in the database, and convincing evidence was found against the pattern of *missing completely at random*. Details of the data missingness analysis are presented in Multimedia Appendices 5 and 6. Participants completed the survey in a median time of 6 minutes and 25 seconds (first to third quartile: 5 minutes and 9 seconds to 8 minutes and 34 seconds).

Study Participants Characteristics

Following the missing data imputation, out of 6273 study participants, 3640 (58.03%) were female. The study participants' mean age was 38.4 years (SD 13.4 years; range 18-90 years). A total of 64.05% (4018/6273) of participants accessed the survey through the health service and 35.95% (2255/6273) through the social networking service.

Substantial differences were found between the participants from different web-based services. Those from the health service—according to their reports—were older, mostly women, living in a smaller place of residence, and a little better educated but earning less money. Although they reported having worse overall health, they tended to eat a healthier diet, engage in more physical activity, and were less likely to smoke cigarettes or use e-cigarettes. Health service visitors were more negative about medicines, more likely to use DS, and were more interested in DS issues. They also declared having more contact with DS advertisements and trusting them more. Media were reported to be the major source of knowledge about DS, irrespective of the type of web-based service used, followed by pharmacists, medical doctors, friends, and dieticians. Detailed characteristics of the study participants, with differences between the participants from 2 web-based services, are presented in Table 1.

Table 1. Sociodemographic characteristics, health-related characteristics, and dietary supplement–related characteristics of the study participants. Data for the total sample and the comparison of participant characteristics from different web-based services are provided.

Characteristics	Total sample (N=6273)	Differences between the type of web-based service				
		Health service users (n=4018)	Social networking service users (n=2255)	Test statistics and *P* values for the comparison[a]		
				Z value	χ^2 (*df*)	*P* value
Sociodemographic						
Age				32.86	N/A[b]	<.001
Values (years), mean (SD)	38.4 (13.4)	42.6 (14.3)	31.1 (7.2)			
Values (years), median (Q_1-Q_3)[c]	35 (28-46)	40 (31-53)	30 (26-35)			
Sex				N/A	2305.1 (1)	<.001
Female, n (%)	3640 (58.03)	3232 (80.44)	408 (18.09)			
Male, n (%)	2633 (41.97)	786 (19.56)	1847 (81.91)			
Education				3.05	N/A	.002
Primary, n (%)	42 (0.67)	30 (0.75)	12 (0.53)			
Secondary or vocational, n (%)	1874 (29.87)	1205 (29.99)	669 (29.67)			
Bachelor, n (%)	1222 (19.48)	686 (17.07)	536 (23.77)			
Master, n (%)	3000 (47.82)	1997 (49.70)	1003 (44.48)			
Doctorate, n (%)	135 (2.15)	100 (2.49)	35 (1.55)			
Number of inhabitants				−6.87	N/A	<.001
Below 5000, n (%)	783 (12.48)	532 (13.24)	251 (11.13)			
5000-50,000, n (%)	1250 (19.93)	865 (21.53)	385 (17.07)			
50,000-500,000, n (%)	2105 (33.56)	1377 (34.27)	728 (32.28)			
Over 500,000, n (%)	2135 (34.03)	1244 (30.96)	891 (39.51)			
Earnings, PLN[d] (US $)				−23.67	N/A	<.001
Below 1000 (256), n (%)	369 (5.88)	297 (7.39)	72 (3.19)			
1000-2000 (256-512), n (%)	1396 (22.25)	1139 (28.35)	257 (11.40)			
2000-3000 (512-768), n (%)	1856 (29.59)	1320 (32.85)	536 (23.77)			
Over 3000 (768), n (%)	2652 (42.28)	1262 (31.41)	1390 (61.64)			
Health-related						
Overall health						
Health status				−9.60	N/A	<.001
Values, mean (SD)	2.59 (0.79)	2.52 (0.78)	2.71 (0.80)			
Values, median (Q_1-Q_3)	3 (2-3)	3 (2-3)	3 (2-3)			
Diet				17.73	N/A	<.001
Values, mean (SD)	3.37 (0.92)	3.53 (0.85)	3.09 (0.98)			
Values, median (Q_1-Q_3)	3 (3-4)	4 (3-4)	3 (2-4)			
Physical activity				4.68	N/A	<.001
Values, mean (SD)	2.76 (1.10)	2.81 (1.05)	2.68 (1.16)			
Values, median (Q_1-Q_3)	3 (2-4)	3 (2-4)	3 (2-4)			
Nicotine status						
Current cigarette smoker, n (%)	785 (12.51)	401 (9.98)	384 (17.03)	N/A	65.5 (1)	<.001

Characteristics	Total sample (N=6273)	Differences between the type of web-based service				
		Health service users (n=4018)	Social networking service users (n=2255)	Test statistics and *P* values for the comparison[a]		
				Z value	χ^2 (*df*)	*P* value
Past but not current cigarette smoker, n (%)	1251 (19.94)	746 (18.57)	505 (22.39)	N/A	13.3 (1)	<.001
Current e-cigarette user, n (%)	343 (5.47)	94 (2.34)	249 (11.04)	N/A	211.6 (1)	<.001
Past but not current e-cigarette user, n (%)	260 (4.14)	100 (2.49)	160 (7.10)	N/A	77.1 (1)	<.001
Beliefs about medicines						
Overuse				14.98	N/A	<.001
Values, mean (SD)	12.9 (3.5)	13.4 (3.5)	12.0 (3.4)			
Values, median (Q_1-Q_3)	13 (10-16)	14 (11-16)	12 (9-15)			
Harm				15.28	N/A	<.001
Values, mean (SD)	9.5 (3.3)	10.0 (3.4)	8.6 (8.6)			
Values, median (Q_1-Q_3)	9 (7-12)	10 (7-12)	8 (6-11)			
DS[e]-related						
Use						
Use of DS[f], n (%)	4615 (73.57)	3311 (82.40)	1304 (57.83)	N/A	448.7 (1)	<.001
Positive effect[g] of DS, n (%)	3162 (68.52)	2330 (70.37)	832 (63.80)	N/A	18.7 (1)	<.001
Negative effect[g] of DS, n (%)	169 (3.66)	126 (3.81)	43 (3.30)	N/A	0.7 (1)	.41
Advertising						
Having contact with DS advertisements[h], n (%)	5418 (86.37)	3603 (89.67)	1815 (80.49)	N/A	103.5 (1)	<.001
Trust in DS advertisements				27.96	N/A	<.001
Values, mean (SD)	17.7 (5.8)	19.2 (5.9)	15.0 (4.5)			
Values, median (Q_1-Q_3)	17 (13-22)	19 (15-24)	15 (12-18)			
Interest in DS				26.45	N/A	<.001
Values, mean (SD)	2.9 (1.1)	3.2 (1.0)	2.4 (1.1)			
Values, median (Q_1-Q_3)	3 (2-4)	3 (3-4)	2 (1-3)			
Getting knowledge from						
Medical doctors				11.93	N/A	<.001
Values, mean (SD)	0.72 (0.81)	0.80 (0.81)	0.57 (0.77)			
Values, median (Q_1-Q_3)	1 (0-1)	1 (0-1)	0 (0-1)			
Pharmacists				18.63	N/A	<.001
Values, mean (SD)	0.88 (0.85)	1.02 (0.86)	0.62 (0.78)			
Values, median (Q_1-Q_3)	1 (0-1)	1 (0-2)	0 (0-1)			
Dieticians				6.77	N/A	<.001
Values, mean (SD)	0.47 (0.78)	0.51 (0.81)	0.38 (0.73)			
Values, median (Q_1-Q_3)	0 (0-1)	0 (0-1)	0 (0-1)			
Friends				9.98	N/A	<.001
Values, mean (SD)	0.68 (0.76)	0.75 (0.78)	0.55 (0.69)			

Characteristics	Total sample (N=6273)	Differences between the type of web-based service				
		Health service users (n=4018)	Social networking service users (n=2255)	Test statistics and *P* values for the comparison[a]		
				Z value	χ^2 (*df*)	*P* value
Values, median (Q_1-Q_3)	1 (0-1)	1 (0-1)	0 (0-1)			
Media				12.61	N/A	<.001
Values, mean (SD)	1.41 (1.04)	1.53 (1.00)	1.19 (1.06)			
Values, median (Q_1-Q_3)	1 (1-2)	2 (1-2)	1 (0-2)			

[a]Asysmptotic Mann-Whitney *U* test (*Z* statistic is provided) or chi-square test (χ^2_{df} is provided); Benjamini-Hochberg corrected significance level: 0.048.

[b]N/A: not applicable.

[c]Q_1-Q_3: 1st to 3rd quartile.

[d]PLN: Polish złoty. PLN was converted to US $ according to the average exchange rate on the study beginning date (source: Narodowy Bank Polski).

[e]DS: dietary supplements.

[f]Within the past 30 days.

[g]Frequency calculated in relation to the number of dietary supplements users.

[h]Within the past week.

Knowledge About DS

The knowledge about DS of web-based service users could be assessed as low, being not much better than a random guess, the success rate of which is 50% for binary questions. Knowledge about DS–general presented overall better results than knowledge about DS–specific. Knowledge about DS of health service users was lower than that of social networking users in both its subscales, approaching an effect size of 6%, as expressed by partial eta-squared, in an analysis adjusted for potential confounders. Details of the knowledge about DS analysis of web-based service users are presented in Table 2. The numbers and frequencies of correct answers to each item of the knowledge about DS questionnaire are presented in Multimedia Appendix 2.

Table 2. Knowledge about dietary supplements among web-based health service users and social networking service users.

Variable	Total sample (N=6273)		Differences between the type of web-based service							
			Type of analysis[a]	Health service users (n=4018), mean (95% CI)	Social networking service users (n=2255), mean (95% CI)	Effect size of the difference[b]		Test statistics and *P* values for the comparison		
	Mean	95% CI				$p\eta^2$ value (%)[c]	Mean (95% CI)	*F* test (*df*)	*F* test (*df*)	*P* value
Knowledge about dietary supplements–General[d]	4.4	4.4-4.5								
			Raw	4.0 (3.9-4.1)	5.3 (5.2-5.3)	8.7	−1.3 (−1.4 to −1.2)	595.4 (1,6271)	N/A[e]	<.001
			Adjusted	4.1 (4.1-4.2)	5.0 (4.9-5.1)	2.0	−0.9 (−1.0 to −0.7)	N/A	129.7 (1,6265)	<.001
Knowledge about dietary supplements–Specific[f]	4.5	4.5-4.6								
			Raw	4.0 (3.9-4.0)	5.5 (5.4-5.6)	13.8	−1.5 (−1.6 to −1.4)	1008.1 (1,6271)	N/A	<.001
			Adjusted	4.0 (4.0-4.1)	5.5 (5.4-5.5)	6.2	−1.4 (−1.5 to −1.3)	N/A	411.9 (1,6265)	<.001
Knowledge about dietary supplements–Total[g]	9.0	8.9-9.1								
			Raw	8.0 (7.9-8.1)	10.8 (10.7-10.9)	16.5	−2.8 (−3.0 to −2.7)	1237.9 (1,6271)	N/A	<.001
			Adjusted	8.2 (8.1-8.3)	10.5 (10.3-10.6)	5.9	−2.3 (−2.5 to −2.1)	N/A	393.4 (1,6265)	<.001

[a]Raw analyses: performed only with the variables reported; adjusted analyses: adjusted for |Age−35|, sex, education, number of inhabitants, earnings, and calendar year—all included as linear factors; estimates in adjusted analyses reported as estimated marginal means.

[b]Reported as a partial eta-squared ($p\eta^2$) and a difference between knowledge about dietary supplements of health service users and knowledge about dietary supplements of social networking service users (95% CI).

[c]$p\eta^2$: partial eta-squared.

[d]An expected result of random guess is 3.5.

[e]N/A: not applicable.

[f]An expected result of random guess is 5.

[g]An expected result of random guess is 8.5.

Determinants of Knowledge About DS

As depicted in Table 3, people less knowledgeable about DS were female, older, had a lower education status, lived in an area with fewer inhabitants, and earned less money per family member. Age appeared to be not linearly linked with knowledge about DS; people around 35 years were the most knowledgeable, and any increase or decrease in age from this point was linked to lower knowledge about DS in a nearly linear manner (such a curvilinear relationship was stable across the type of web-based service and after adjusting for potential confounders). Health-related habits were negligibly linked to knowledge about DS after adjusting for potential confounders. People who believed medicines were harmful or overprescribed by doctors had lower levels of knowledge about DS. Using a DS within the past 30 days was also an indicator of lower knowledge about DS. Among DS users, DS's perceived beneficial effect was a negative modulator of knowledge about DS, whereas the harmful

effect was a positive modulator. Trust in advertising DS appeared to be the strongest negative predictor of knowledge about DS, which was equal in both knowledge about DS domains. The effect of trust in advertising DS on knowledge about DS was particularly high among those who had contact with the respective advertisements. Interestingly, having contact with DS advertisements, without taking trust in advertising DS into account, had a somewhat positive effect on knowledge about DS. Being interested in DS was linked to a lower level of knowledge about DS. Although reported as the major source of knowledge about DS for web-based service users, internet and traditional media were negatively but weakly associated with knowledge about DS. Instead, the major source of false information regarding DS was found to be friends with no medical education. Among health care specialists, pharmacists appeared to be a misleading source of knowledge about DS.

High compatibility was found between nonparametric and parametric tests for the sensitivity analysis of the associations reported in Table 3 (the correlation of Spearman rho with Pearson *r* coefficients in univariate analyses was *r*=0.9985; 95% CI 0.9976-0.9990, and the median absolute difference between the corresponding coefficients was 0.004 of a maximum value of 0.034). Sensitivity analysis performed using GLM methods in a database of complete cases only (n=5633) was also highly consistent with the analysis performed in the original database following missing data imputation (the correlation of corresponding β regression coefficients was *r*=0.9991; 95% CI 0.9988-0.9993, and the median absolute difference between the corresponding coefficients was 0.003 of a maximum value of 0.018).

Table 3. Association between characteristics of the study participants and their knowledge about dietary supplements. Analyses were performed in the total sample of 6273 internet users. The results presented in italics are statistically significant at the Benjamini-Hochberg-corrected significance level of 0.036.

Characteristics	Raw analyses[a]						Adjusted analyses[b]					
	Knowledge about DS[c]–general		Knowledge about DS–specific		Knowledge about DS–total		Knowledge about DS–general		Knowledge about DS–specific		Knowledge about DS–total	
	Value, β coefficient (95% CI)	*P* value	Value, β coefficient (95% CI)	*P* value	Value, β coefficient (95% CI)	*P* value	Value, β coefficient (95% CI)	*P* value	Value, β coefficient (95% CI)	*P* value	Value, β coefficient (95% CI)	*P* value
Sociodemographic												
Age[d]	*−.16 (−.18 to −.13)*	*<.001*	*−.26 (−.28 to −.24)*	*<.001*	*−.25 (−.28 to −.23)*	*<.001*	*−.05 (−.08 to −.03)*	*<.001*	*−.12 (−.15 to −.10)*	*<.001*	*−.11 (−.13 to −.08)*	*<.001*
|Age−35|	*−.19 (−.22 to −.17)*	*<.001*	*−.23 (−.26 to −.21)*	*<.001*	*−.26 (−.28 to −.24)*	*<.001*	*−.10 (−.13 to −.08)*	*<.001*	*−.13 (−.16 to −.11)*	*<.001*	*−.14 (−.17 to −.12)*	*<.001*
Sex (0=female and 1=male)	*.22 (.20 to .25)*	*<.001*	*.22 (.19 to .24)*	*<.001*	*.27 (.24 to .29)*	*<.001*	*.09 (.06 to .12)*	*<.001*	−.00 (−.03 to .03)	<.001	*.06 (.03 to .08)*	*<.001*
Education	*.13 (.11 to .16)*	<.001	−.00 (−.03 to .02)	.89	*.08 (.06 to .11)*	*<.001*	*.11 (.08 to .13)*	<.001	−.02 (−.05 to .00)	.08	*.06 (.03 to .08)*	*<.001*
Number of inhabitants	*.11 (.09 to .14)*	*<.001*	*.07 (.05 to .09)*	*<.001*	*.11 (.09 to .14)*	*<.001*	*.06 (.03 to .08)*	*<.001*	*.04 (.01 to .06)*	*.002*	*.06 (.04 to .08)*	*<.001*
Earnings	*.19 (.16 to .21)*	*<.001*	*.14 (.12 to .16)*	*<.001*	*.20 (.18 to .23)*	*<.001*	*.05 (.02 to .07)*	*<.001*	.02 (−.01 to .05)	.13	*.04 (.02 to .07)*	*<.001*
Health-related												
Health status	*.06 (.03 to .08)*	*<.001*	*.08 (.05 to .10)*	*<.001*	*.08 (.06 to .11)*	*<.001*	*−.03 (−.05 to −.00)*	*.02*	.01 (−.02 to .03)	.67	−.01 (−.04 to .01)	.22
Diet	*−.06 (−.08 to −.03)*	*<.001*	*−.10 (−.12 to −.08)*	*<.001*	*−.09 (−.12 to −.07)*	*<.001*	.01 (−.01 to .04)	.36	.00 (−.02 to .02)	.96	.01 (−.02 to .03)	.53
Physical activity	*−.03 (−.05 to −.00)*	*.03*	*−.03 (−.06 to −.01)*	*.01*	*−.04 (−.06 to −.01)*	*.05*	*−0.03 (−.05 to −.00)*	*.03*	−.01 (−.03 to .02)	.54	−.02 (−.04 to .00)	.07
Current cigarette smoker	−.01 (−.03 to .02)	.44	.02 (−.01 to .04)	.13	.01 (−.02 to .03)	.68	*−.03 (−.05 to −.00)*	*.02*	−.02 (−.04 to .00)	.08	*−.03 (−.05 to −.01)*	*.01*
Past but not current cigarette smoker	.04 (.02 to .07)	<.001	−.01 (−.04 to .01)	.33	.02 (−.00 to .04)	.11	*0.03 (.01 to .06)*	.006	−.03 (−.05 to −.00)	.03	.01 (−.02 to .03)	.61
Current e-cigarette user	*.07 (.05 to .10)*	*<.001*	*.07 (.04 to .09)*	*<.001*	*.09 (.06 to .11)*	*<.001*	.02 (−.00 to .04)	.10	−.00 (−.03 to .02)	.74	.01 (−.01 to .03)	.37
Past but not current e-cigarette user	*.04 (.02 to .07)*	*.001*	*.06 (.03 to .08)*	*<.001*	*.06 (.04 to .08)*	*<.001*	0.01 (−.01 to .03)	.45	0.01 (−.01 to .03)	.31	.01 (−.01 to .04)	.26
Beliefs that medicines are overused	*−.08 (−.10 to −.05)*	*<.001*	*−.19 (−.21 to −.16)*	*<.001*	*−0.16 (−.19 to −.14)*	*<.001*	−.01 (−.03 to .02)	.54	*−.11 (−.13 to −.09)*	*<.001*	*−.07 (−.09 to −.05)*	*<.001*
Beliefs that medicines are harmful	*−.19 (−.21 to −.16)*	*<.001*	*−.17 (−.20 to −.15)*	*<.001*	*−.22 (−.24 to −.20)*	*<.001*	*−.11 (−.13 to −.09)*	*<.001*	*−0.09 (−.11 to −007)*	*<.001*	*−.12 (−.15 to −.10)*	*<.001*
DS-related												

Characteristics	Raw analyses[a]						Adjusted analyses[b]					
	Knowledge about DS[c]–general		Knowledge about DS–specific		Knowledge about DS–total		Knowledge about DS–general		Knowledge about DS–specific		Knowledge about DS–total	
	Value, β coefficient (95% CI)	*P* value	Value, β coefficient (95% CI)	*P* value	Value, β coefficient (95% CI)	*P* value	Value, β coefficient (95% CI)	*P* value	Value, β coefficient (95% CI)	*P* value	Value, β coefficient (95% CI)	*P* value
Use of DS	*−.11 (−.13 to −.08)*	*<.001*	*−.29 (−.32 to −.27)*	*<.001*	*−.24 (−.26 to −.22)*	*<.001*	*−.03 (−.06 to −.01)*	*.005*	*−.21 (−.23 to −.18)*	*<.001*	*−.14 (−.17 to −.12)*	*<.001*
Positive effect of DS	*−.12 (−.14 to −.09)*	*<.001*	*−.27 (−.29 to −.24)*	*<.001*	*−.23 (−.26 to −.21)*	*<.001*	*−.06 (−.09 to −.04)*	*<.001*	*−.20 (−.22 to −.17)*	*<.001*	*−.16 (−.18 to −.13)*	*<.001*
Positive effect of DS with adjustment for DS use	*−.09 (−.12 to −.06)*	*<.001*	*−.14 (−.17 to −.11)*	*<.001*	*−.14 (−.17 to −.11)*	*<.001*	*−.06 (−.09 to −.04)*	*<.001*	*−.12 (−.15 to −.09)*	*<.001*	*−.11 (−.14 to −.08)*	*<.001*
Negative effect of DS	.02 (−.01 to .04)	.22	.01 (−.01 to .04)	.36	.02 (−.01 to .04)	.18	*.03 (.01 to .05)*	*.01*	*.03 (.00 to .05)*	*.03*	*.03 (.01 to .06)*	*.003*
Negative effect of DS with adjustment for DS use	.03 (.00 to .05)	.04	*.04 (.02 to .07)*	*<.001*	*.04 (.02 to .07)*	*<.001*	*.03 (.01 to .06)*	*.006*	*.04 (.02 to .07)*	*<.001*	*.05 (.02 to .07)*	*<.001*
Having contact with DS advertisements	.00 (−.02 to .03)	.88	*−.04 (−.06 to −.01)*	*.004*	−.02 (−.05 to .00)	.11	*.04 (.02 to .07)*	*<.001*	.01 (−.01 to .04)	.26	*.03 (.01 to .06)*	*.003*
Trust in advertising DS	*−.39 (−.41 to −.37)*	*<.001*	*−.39 (−.41 to −.37)*	*<.001*	*−.48 (−.50 to −.46)*	*<.001*	*−.31 (−.33 to −.28)*	*<.001*	*−.29 (−.31 to −.27)*	*<.001*	*−.37 (−.39 to −.34)*	*<.001*
Having contact with DS advertisements×trust in advertising DS[e]	*−.14 (−.23 to −.05)*	*.004*	−.06 (−.16 to .03)	.21	*−.12 (−.21 to −.03)*	*.007*	*−.14 (−.23 to −.04)*	*.004*	−.06 (−.15 to .03)	.18	*−.12 (−.21 to −.04)*	*.005*
Interest in DS	*−.11 (−.13 to −.08)*	*<.001*	*−.28 (−.30 to −.26)*	*<.001*	*−.24 (−.26 to −.21)*	*<.001*	−.01 (−.03 to .02)	.61	*−.17 (−.19 to −.14)*	*<.001*	*−.10 (−.13 to −.08)*	*<.001*
Getting knowledge about DS from medical doctors[f]	−.01 (−.04 to .02)	.38	−.01 (−.04 to .01)	.31	−.02 (−.04 to .01)	.25	.00 (−.02 to .03)	.73	.01 (−.02 to .03)	.49	.01 (−.02 to .03)	.52
Getting knowledge about DS from pharmacists[f]	*−.08 (−.11 to −.05)*	*<.001*	*−.09 (−.11 to −.06)*	*<.001*	*−.10 (−.13 to −.07)*	*<.001*	−.02 (−.05 to .01)	.18	*−.03 (−.05 to −.00)*	*.03*	*−.03 (−.05 to −.00)*	*.03*
Getting knowledge about DS from dieticians[f]	.02 (−.01 to .04)	.23	−.01 (−.04 to .01)	.33	.00 (−.02 to −.03)	.84.	.00 (−.02 to .03)	.74	−.02 (−.04 to .01)	.21	−.01 (−.03 to .02)	.59
Getting knowledge about DS from friends[f]	*−.09 (−.12 to −.07)*	*<.001*	*−.14 (−.16 to −.11)*	*<.001*	*−.14 (−.16 to −.11)*	*<.001*	*−.09 (−.11 to −.06)*	*<.001*	*−.13 (−.15 to −.11)*	*<.001*	*−.13 (−.15 to −.11)*	*<.001*

Characteristics	Raw analyses[a]						Adjusted analyses[b]					
	Knowledge about DS[c]–general		Knowledge about DS–specific		Knowledge about DS–total		Knowledge about DS–general		Knowledge about DS–specific		Knowledge about DS–total	
	Value, β coefficient (95% CI)	*P* value	Value, β coefficient (95% CI)	*P* value	Value, β coefficient (95% CI)	*P* value	Value, β coefficient (95% CI)	*P* value	Value, β coefficient (95% CI)	*P* value	Value, β coefficient (95% CI)	*P* value
Getting knowledge about DS from media[f]	−.02 (−.04 to .01)	.20	*−.16 (−.18 to −.13)*	*<.001*	*−.10 (−.13 to −.08)*	*<.001*	*.03 (.01 to .06)*	*.006*	*−.09 (−.12 to −.07)*	*<.001*	*−.03 (−.06 to −.01)*	*.003*

[a]Performed only with the variables reported.

[b]Adjusted for |Age−35|, sex, education, number of inhabitants, earnings, type of web-based service, and calendar year—all included as linear factors.

[c]DS: dietary supplement.

[d]Adjusted analyses do not include |Age−35|; the association between Age and Knowledge about dietary supplements seemed to be an inverted-U shape with a maximum in knowledge for age of about 35 years; consequently, Age was transformed in further analyses to |Age−35| to approximate the association to linear; the assoction between Age and Trust in advertising dietary supplements similarly looked as a nonlinear U shape with a minimum of trust for age of about 30 years.

[e]Adjusted for contact with dietary supplements advertisements and trust in advertising dietary supplements; having contact with dietary supplements advertisements in this model associated with knowledge about dietary supplements total in adjusted analysis with β of .11 (95% CI .04 to .18), *P*=.001; Moreover, trust in advertising DS associated with knowledge about dietary supplements—total in adjusted analysis among those who reported having contact with dietary supplements advertisements with β of −.38 (95% CI −.41 to −.36), *P*<.001, and those who did not report having contact with β of −.24 (95% CI −.31 to −.18), *P*<.001.

[f]The associations of knowledge about dietary supplements with getting knowledge about dietary supplements from a particular source were adjusted for all other sources of knowledge about dietary supplements.

Prediction Model of Knowledge About DS

The variables selected for knowledge about DS modeling were trust in advertising DS, BMQ Harm, Age (following transformation to |Age−35|), Sex, and Getting knowledge about DS from friends. The model explained 31.2% of the variance in knowledge about DS. The predictors in the model were intercorrelated to a low extent, with a median absolute correlation coefficient of 0.114 and a maximum of 0.214 for trust in advertising DS and sex. The assumption of a linear contribution of predictors to the model was satisfactory. The model characteristics are listed in Table 4. According to this model, knowledge about DS can be predicted using the following formula:

$$\text{Knowledge about DS} = 19.397539 - 0.412775 \times \text{Trust in advertising DS} - 0.23657 \times \text{BMQ Harm} - 0.113879 \times |\text{Age} - 35| + 1.850937 \times \text{Sex} - 0.691575 \times \text{Getting knowledge about DS from friends}$$

In the formula, *Trust in advertising DS* represents trust in advertising DS operationalized with the questionnaire on trust in advertising DS, ranging from 8 to 40; *BMQ Harm* represents BMQ operationalized with the BMQ questionnaire, subscale Harm, ranging from 4 to 20; *|Age−35|* is the absolute value of the difference between age (years) and 35; *Sex* was operationalized as 0 for females and 1 for males; and *Getting knowledge about DSfrom friends* represents a self-reported declaration on the extent of getting knowledge about DS from friends with no medical education (operationalized as 0 for *not at all*, 1 for *to little extent*, 2 for *to medium extent*, and 3 for *to large extent*).

Table 4. Characteristics of a model to describe knowledge about dietary supplements. The total sample of 6273 internet users was included. Test statistics, *P* value and coefficient of determination of a full model are *F*(5,6267)=568.2; *P*<.001; R^2=0.312, respectively.

Predictor	Effect size			*F* test (*df*)	*P* value
	β coefficient	95% CI	$p\eta^2$ (%)[a]		
Trust in advertising DS[b]	−.39	−.41 to −.37	16.4	123.4 (1,6267)	<.001
Beliefs that medicines are harmful	−.13	−.15 to −.11	2.2	143.1 (1,6267)	<.001
|Age−35|	−.17	−.19 to −.15	3.9	251.2 (1,6267)	<.001
Sex (0=female and 1=male)	.15	.13 to .17	2.9	187.7 (1,6267)	<.001
Getting knowledge about DS from friends	−.08	−.11 to −.06	1.0	62.0 (1,6267)	<.001

[a]$p\eta^2$: partial eta-squared.

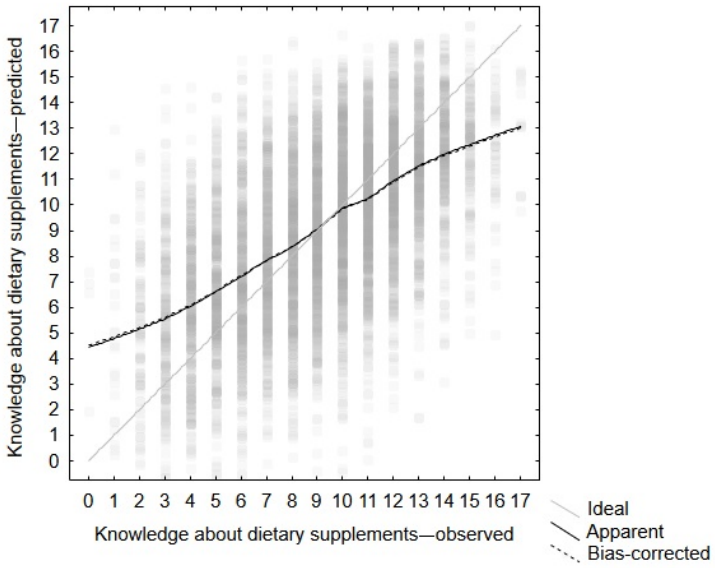
[b]DS: dietary supplement.

The model could predict knowledge about DS with an MAE of 2.52 and an RMSE of 3.19 points (median 2.09, 1st to 3rd quartile 1.00-3.60). Although this performance was far from ideal, it was significantly better than a random guess based

solely on knowledge about DS distribution that yielded MAE of 3.82 and RMSE of 4.76 points (median 3.23, 1st to 3rd quartile 1.59-5.49). Results of the model internal validation indicated that predictions of a bias-corrected model almost overlapped the original model with MAE of 2.51 and RMSE of 3.17 points (median 2.09, 1st to 3rd quartile 1.00-3.58), and the correlation between knowledge about DS values predicted by the original model and the bias-corrected one was very close (r=0.99986; 95% CI 0.99985-0.99986). Extreme values predicted by the bias-corrected model did not exceed an upper limit of 17 points; however, a lower limit was below zero due to the *Age* component not being restricted (36, ie, 0.57% of observations were below zero, with a minimum of −5.76). The model could predict relatively well the knowledge about DS values around its central value of 8.99 but exhibited worse performance with extreme knowledge about DS values, which was reflected by the deviation of a calibration plot from linearity. A calibration plot of the model is shown in Figure 1.

The 5 predictors could model not only knowledge about DS–total but also knowledge about DS–general and knowledge about DS–specific separately, with each of the predictors making a significant contribution to modeling both constructs, reaching determination coefficients of 20.4% and 22.0%, respectively. Importantly, although the model did not include as a predictor the type of web-based service through which a study participant accessed the survey, the model could predict (though overestimate) that the health service users had lower knowledge about DS than the users of the social networking service (7.6, 95% CI 7.5-7.7 vs 11.5, 95% CI 11.4-11.6, respectively; P<.001; a bias-corrected model). Moreover, the knowledge about DS score predicted by the bias-corrected model was found to be associated with the other participant characteristics examined in this study in a similar pattern as the associations of the observed knowledge about DS scores. The link between both associations, as expressed by Pearson r, was 0.95 (95% CI 0.91-0.97); the median absolute difference between the corresponding β coefficients was 0.03 (1st to 3rd quartile 0.01-0.09), and the maximum absolute difference was 0.38. Associations between knowledge about DS, predicted with the biased-corrected model and characteristics of the study participants, are reported in Multimedia Appendix 7.

Figure 1. Calibration plot of the model to describe knowledge about dietary supplements. Black lines represent lowess (locally weighted scatterplot smoothing) fitted curves with a smoothing parameter of 0.25. The solid black line denotes the original (apparent) model, whereas the dotted black resulted from cross-validation (bias-corrected). The gray diagonal line represents hypothetical ideal prediction. Gray-filled circles depict individual data points predicted by the bias-corrected model. The circles are partially transparent; thus, the intensity of gray shadings reflects the density of points in an area.



Knowledge About DS in Selected Subpopulations

Knowledge about DS was determined in some subpopulations thought to be critical to public health: older people aged 60 years or more, residents of rural areas (below 5000 inhabitants), and people with low earnings (below 1000 PLN [US $256] of monthly net household earnings per family member). All these subpopulations were found to have lower knowledge about DS than the remaining internet users. However, after adjusting for potential confounders, the effect persisted only for the older people and, to some extent, rural residents. For people with low earnings, the effect shrank to insignificant, as explained by confounders. The analysis of knowledge about DS in subpopulations is presented in Table 5. The proposed 5-predictor model to describe knowledge about DS worked well for rural residents and people with low earnings: all the predictors significantly contributed to prediction and presented a similar prediction error to other people. On the other hand, knowledge about DS among older people was not efficiently predicted by the 5-predictor model, with only trust in advertising DS and Age contributing significantly. The model performance according to subpopulations is presented in Table 5.

Table 5. Knowledge about dietary supplements and the model performance in subpopulations.

Type of analysis or model characteristics	Subpopulation		
	Older people (≥60 years; n=695)	Rural residents (<5000 inhabitants; n=783)	People with low earnings[a] (<1000 PLN[b] [US $256]; n=369)
Difference in knowledge about DS[c]–total between a subpopulation and the remaining internet users			
Raw difference[d]			
Values, mean	−2.1	−0.5	−1.0
95% CI	−2.4 to −1.8	−0.8 to −0.3	−1.3 to −0.6
P value	<.001	<.001	<.001
Adjusted difference[e]			
Values, mean	−1.1	−0.3	−0.2
95% CI	−1.3 to −0.9	−0.5 to −0.0	−0.5 to 0.1
P value	<.001	.02	.18
Performance of the proposed 5-predictor model in a subpopulation			
Predictors evaluation			
Trust in advertising DS			
Values, β coefficient	−.34	−.44	−.37
95% CI	−.42 to −.27	−.50 to −.38	−.46 to −.28
P value	<.001	<.001	<.001
Beliefs that medicines are harmful			
Values, β coefficient	−.06	−.18	−.15
95% CI	−.13 to .01	−.24 to −.12	−.24 to −.06
P value	.08	<.001	.002
|Age−35|			
Values, β coefficient	−.10	−.13	−.20
95% CI	−.18 to −.03	−.19 to −.08	−.29 to −.11
P value	<.001	<.001	<.001
Sex (0=female and 1=male)			
Values, β coefficient	.01	.11	.13
95% CI	−.06 to .08	.05 to .17	.04 to .22
P value	.76	.001	.005
Getting knowledge about DS from friends			
Values, β coefficient	−.04	−.07	−.14
95% CI	−.11 to −.03	−.12 to −.01	−.23 to −.05
P value	.31	.03	.002
Other indices			
Full model statistics			
F test (*df*)	2.4 (5,689)	N/A[f]	N/A
F test (*df*)	N/A	8.2 (5,777)	N/A
F test (*df*)	N/A	N/A	28.1 (5,363)
P value	<.001	<.001	<.001
R-squared	0.129	0.340	0.279

Type of analysis or model characteristics	Subpopulation		
	Older people (≥60 years; n=695)	Rural residents (<5000 inhabitants; n=783)	People with low earnings[a] (<1000 PLN[b] [US $256]; n=369)
MAE[g,h]	2.96	2.51	2.52
RMSE[g,i]	3.70	3.17	3.25

[a]Monthly net household earnings per family member.

[b]PLN: Polish złoty. PLN was converted to US$ according to the average exchange rate on the study beginning date (source: Narodowy Bank Polski).

[c]DS: dietary supplements.

[d]Performed only with the variables reported.

[e]Adjusted for |Age-35|, sex, education, number of inhabitants, earnings, type of web-based service, calendar year—all expressed as linear factors—with the exclusion of a variable differentiating a subpopulation.

[f]N/A: not applicable.

[g]Assessed in a biased-corrected model.

[h]MAE: mean absolute error of prediction.

[i]RMSE: root mean squared error.

Discussion

Principal Findings

There is an ongoing debate over DS [49,50]. Recently, in Poland, several influential high-quality reports suggested unsatisfactory control over the DS market, resulting in inadequate safety [21,22,51]. These reports highlight the need for adequate public education to make informed decisions regarding DS intake [21,22]. This study fits into this discourse by examining knowledge about DS among Polish internet users. The results indicate the level of knowledge to be low and not much different from the previous preliminary report, which used the same research tool [38]. Several characteristics were found to be associated with knowledge about DS, with trust in advertising DS being the most influential; this study outlines a predictive model that could explain almost a third of the variance in knowledge about DS.

Although the results indicate the level of knowledge to be low, such a pessimistic diagnosis appears incomplete. A low level of knowledge about DS is not only derived from the unavailability of knowledge or ignorance; it should be interpreted rather as holding a *false belief* [52] that DS are thoroughly controlled, well tested, effective, and harmless or an expression of the overall confidence in DS properties, as outlined by the other authors [3,25,53]. A qualitative study aimed at exploring beliefs about DS and their use confirmed such interpretation as it identified that beliefs about DS people hold are largely related to their definition, effectiveness in health enhancement and illness prevention, and risks [54], which are the areas covered by the knowledge about DS questionnaire used in this study.

Our results indicate that the users of a web-based health service presented false beliefs about DS to a greater extent than the users of a web-based social networking service. This can be explained in at least three ways. First, health services may attract people who believe DS are effective and safe, as confidence in a product coexists with seeking information about it [55]. Second, web-based health services may not be efficient in promoting evidence-based facts regarding DS, which is obscured by advertising [56]. Third, social networking services users may be influenced to a strong degree by the nature of this environment [57] and may hence not fully engage with the survey and provide biased responses.

In general, people with low knowledge about DS presented such characteristics as interest in the topic of DS, tendency to seek information about them and eagerness to know more, use of DS, and subjective experience of their positive effects without negative effects. The literature on the relationship between DS intake and knowledge about DS appears confusing, with some of the studies reporting positive links [30] and others negative links [38]. However, this is only an apparent discrepancy, as the result depends on the method used. DS users seem to present high *subjective* awareness and confidence in DS [25], which are, in fact, *objectively* false beliefs [33]. The major reason why subjective and objective knowledge are not in line was outlined by the Dunning-Kruger effect [58], stating that unknowledgeable people tend to overestimate their ability [53]. However, the reason why some people do not achieve high scores in objective knowledge about DS, despite the efforts made, should be sought in the quality of information sources they use to learn about DS. Popular beliefs about DS are in conflict with evidence-based facts. Media coverage, particularly DS advertisements, often reinforces popular beliefs rather than providing facts [10-15]. For example, Wierzejska [59], while analyzing the Polish DS market, found that most DS advertisements promise beneficial effects to the human body, going beyond the standards established for food. Trusting such messages may contribute to the accumulation of false beliefs regarding DS.

trust in advertising DS was found to be the strongest among the examined predictors of knowledge about DS and was expressed in a negative way. trust in advertising DS was particularly negatively linked with knowledge about DS among those who reported having contact with DS advertisements recently; however, it could associate with knowledge about DS even among those who did not. Interestingly, having contact with DS advertisements itself, irrespective of trust in advertising DS, was weakly associated with knowledge about DS, but in a

positive way. Developing trust, particularly in its cognitive form, is a long-lasting process that requires deeper commitment than just *having a contact* [60,61]. The DS consumer who trusts advertisements is *dependent* on a message and therefore exposed to a significant risk [62] of being misinformed by DS advertisements, which spread unverified and misleading claims regarding health benefits of DS [59]. On the other hand, having contact with DS advertisements appears incapable of deceiving consumers. Moreover, being chronically exposed to DS advertisements may result in the opposite effect of advertising fatigue [63], leading to advertising distrust, which may, in turn, promote prudence and true beliefs regarding DS.

Causal relationship between knowledge about DS and trust in advertising DS is not evident, although possible [38]. The idea of gaining false beliefs about DS in response to high trust in advertising DS is particularly attractive to opponents of advertising. This is to some extent supported by our findings, as people reporting to get information about DS from media presented lower knowledge about DS. On the other hand, the idea of trust in advertising DS being shaped as a result of holding false beliefs about DS appears similarly likely and may reflect the natural phenomenon of gaining confidence in messages that appear to be true [55].

This study found the media (magazines, television, radio, and the internet) to be the primary source of information about DS to internet users. This is not surprising regarding the nature of the population studied and is in line with some [4,31], but not all [28], previous studies. The top rank of media may be regarded as undermining the unwavering position of health care specialists in providing health-related information [3,64,65]. Considering that web-based resources about DS leave much to be desired [10-15] and are full of deceitful DS advertisements, one could suspect that using the internet as a source of information about DS blurs the knowledge about DS to a large degree. Strikingly, the link presents merely a small effect size, and for knowledge about DS–general, the association was even positive. Although some popular internet resources about DS suggest their therapeutic benefits (possibly contributing to lower knowledge about DS–specific), they present the true characteristics of the products, quoting the definition or introducing the regulatory framework (possibly contributing to an increase in knowledge about DS–general) [66-68]. Similarly, the association of *interest in DS* and *use of DS* was found to be negative only with knowledge about DS–specific but not (or almost not) with knowledge about DS–general. Finally, accessing web-based DS information should not be perceived as a source of criticism. Quite the opposite, it offers new opportunities to improve the patient-physician relationship and, in turn, verify and reinforce true beliefs about DS [69].

Concerns should be raised about the negative link between knowledge about DS and getting information about DS from friends with no medical education. Word-of-mouth advertising was suggested to be efficient in spreading ideas long ago [70]. It is considered more credible than a commercial and may be successfully applied to informal promotions of health-related products [71]. As internet users report experiencing positive, much more frequently than negative, DS effects (according to our results) and word-of-mouth is likely to transmit subjective—yet unproven—information, it may strengthen false beliefs about DS.

Among health care practitioners, only obtaining information about DS from pharmacists was associated with lower knowledge about DS. Although the effect size was very small, it may mean that pharmacist advice contributed to strengthening false beliefs about DS. This is particularly worrying, as pharmacists were reported as the second most important source of information about DS and the first personal one. Concerns related to the ethics of DS being sold in pharmacies have already been raised, and a conflict between dispensing and counseling on DS has been identified [72]. Yet this is the role of pharmacists to educate patients about these products [73].

Our results indicate a negative association between knowledge about DS and both subscales of the BMQ-General, implying a link between confidence in DS and overall negative views about medicines. This association reflects an existing conflict between alternative and conventional medicine [74,75]. This confirms previous findings that individuals who advocate alternative medicine distrust or are dissatisfied with conventional care [76-78]. Our results support the need for reconciliation and integration of both attitudes for the sake of patients [74,75].

Some sociodemographic determinants of knowledge about DS were identified in this study. Women exhibited lower knowledge about DS than men. However, the result was not replicated by other knowledge about DS studies, which found insignificant [23,25,38] or inconsistent [28] differences between sexes. The effect size of the difference in this study, although very small following adjustment for confounders, was significant. It could partially reflect the overall tendency for women to use DS more than men [1,6,28,79] or hold more positive attitudes toward advertising [80], associated with more confidence in DS and lower objective knowledge about DS, as discussed above. A detailed consideration of the reasons why women are apparently more willing to use DS is presented elsewhere [5]. This study also found that people who are generally better educated, living in larger places of residence, or having more money for living presented higher knowledge about DS, particularly in its general domain. A previous preliminary study using the same knowledge about DS measuring tool also reported such findings for the overall education status [38]. Better educated people tend to rely less on advertising [81,82], which is a factor linked to better knowledge about DS. This finding may support our results.

Age presented the most complicated pattern of a relationship with knowledge about DS, with its peak observed in people around 35 years. There is a scarcity of literature linking age to knowledge about DS. In the only study that could be mentioned, knowledge about DS was tested in the general population of young adults with a mean age of 25 years and an SD of 5 years. A subgroup of participants aged >31 years was able to provide the definition of supplementation (objective knowledge about DS) in a greater percentage in comparison with younger participants. Moreover, there was a trend for younger participants to express higher levels of subjective, possibly false knowledge [33]. Although the cited study is in line with this study concerning young adults, it did not include older people. Studies examining DS intake across various age groups

suggested that older people used more DS [31,83], possibly having more false-positive beliefs about DS, or no significant difference was found [26,30]. It is not surprising that there is no clear replication of these results in the literature, as the knowledge about DS–age pattern revealed is not linear and would require an explorative research approach and a large sample size. Obtaining such results was not possible until this study was performed. Research on advertising attitudes may help to explain this phenomenon. Consumers aged 40 years or more were found to have a more positive attitude toward advertisements [81]. Older adults were also more likely to be persuaded by advertising messages [84]. In another study, a more favorable attitude was expressed by younger people aged less than 30 years [85]. None of these studies found a curvilinear relationship between age and susceptibility to advertising; however, their synthesis may approximate such an association. Interestingly, in this study, trust in advertising DS, which is a similar construct to the one considered above, also achieved its minimum in young adults around 30 years old and increased in both younger and older people. As trust in advertising DS is closely associated with knowledge about DS, it may contribute to the explanation of this phenomenon. On the other hand, not all the reports replicate the above findings [80], and such interpretation should not be regarded as straightforward.

In the culminating part of this study, a predictive model of knowledge about DS was built using a multivariate linear regression strategy. The model was able to explain almost one-third of the knowledge about DS variability and included 5 predictors. Although the choice of the predictors was based to some extent on statistical significance, the most objective characteristics outlined a priori in the study protocol [39] were reflected in the final model, which makes it partially based on subject matter knowledge as generally advised [86]. The proposed model was simple enough not to be prone to overfitting, as tested by the internal validation procedure, and versatile enough to predict differences between the 2 web-based services tested and forecast most of the associations tested. The model worked well in subpopulations of rural citizens and people of low income, but worse in older people, whose beliefs related to DS require special attention [26]. This is likely the model will predict the outcome not much worse in an external validation study and in practice. It also has the potential for electronic application. Although the model still made significant mistakes, leaving more than two-thirds of knowledge about DS variance unallocated, it may help scientists recognize further paths of research: health care workers identify people endangered with illusory beliefs about DS, and educators tailor campaigns to the public to induce behavioral change [20,87,88].

The issue of generalizability requires some debate before any conclusions are drawn. First, a convenience sample of internet users from a restricted number of services may present a barrier to these findings' generalizability to the entire web-based community. Health and social networking services presented extreme demographics, likely covering the scope of internet users in general. Considering the relatively large number of possibly not overlapping users of both services [47], this study could capture a broad range of internet users in Poland. Labeling the sample as representative, however, should be avoided. Similarly, the sample should not be regarded as reflecting the overall Polish population. Although most young people access the internet, more than three-fourths of Poles aged 65 years or more remain offline [89], and the group may represent different attitudes toward DS [26].

Apart from generalizability issues, this study has some limitations that warrant mention while interpreting the results. First, the results come from self-reported declarative data, which may be shifted according to the social desirability bias theory [90]. Such results may only approximate the true beliefs and behaviors of study participants and require cautious interpretation [91]. Second, the study did not include external validation of the findings, particularly the predictive model of knowledge about DS. External validation is essential before implementing the model in practice [92]. Third, although single-item measures have the potential to accurately reflect the construct assessed [93], some of the measures in this study were not rigorously validated, although they were used in the same form elsewhere [25]. Fourth, the study did not examine other potential knowledge about DS predictors such as general medical knowledge [94], experience with particular DS products, or personality traits. The survey did not include these variables to keep it short and to minimize the burden of responding [95]. This study, however, substantially reduced possible bias through prospective registration of the study protocol, using some valid research instruments, reporting data transparently, and performing sensitivity analyses.

This report requires further testing to establish the results' external validity among different groups of internet users to account for their diversity. Further investigation into the mechanisms explaining the relationships between the analyzed constructs is needed; this should be supplemented with more qualitative research to confirm the obtained results and reveal a deeper understanding of the problem. Finally, this and any follow-up studies should aim to develop and implement an effective strategy for public education regarding DS, ideally in a web-based form, to ensure rational use of the products for the public's greatest benefit.

Conclusions

Internet users in Poland with no medical education, particularly those attending a health service, exhibit some false beliefs regarding the quality requirements, efficacy, and safety of DS. Holding such false beliefs is positively associated with trusting DS advertising; their intake, particularly when positive effects are subjectively experienced; being interested in the issues of DS; getting information about them through word-of-mouth; beliefs that medicines are harmful or overused; and some demographic characteristics. The proposed 5-predictor model to forecast knowledge about DS can explain almost a third of its variability and appears resistant to overfitting. The results call for implementing efficient education about DS, including the promotion of advertising literacy, to correct consumers' attitudes toward DS. The results also underline the need to introduce more conservative regulatory frameworks for DS marketing. Further research is needed for external validation of the results and to obtain a more comprehensive understanding of the revealed phenomena.

Acknowledgments

This work was supported by DOZ.pl Sp. z o.o, with the grant received by MSK (contract date October 31, 2019), and by the Medical University of Lodz, Poland, with the grant received by JS (grant 500/6-086-01/500-41). The authors thank DOZ.pl Sp. z o.o. for the partnership in this study.

Authors' Contributions

MSK was a manager of the project, conceptualized the study, obtained part of the funding, developed the methods, designed and approved the survey, preregistered the study protocol, performed the analysis, interpreted the results, and drafted the paper as the first author. RH helped interpret the study results, contributed to the study validity, and reviewed and edited the drafted paper. EP helped develop the methods, helped interpret the study results, and reviewed and edited the drafted paper. EK supervised the project, and reviewed the drafted paper. JS obtained part of the funding, supervised the project, and reviewed the drafted paper.

Conflicts of Interest

EP, a coauthor of this study, has been employed in Osom Studio, an e-marketing agency. The role of DOZ.pl Sp. z o.o. in this study was to prepare promotional materials to help recruit participants and to enter the survey content into an external web-based survey system. DOZ.pl Sp. z o.o, Medical University of Lodz, and Osom Studio had no role in the study design; data collection; analysis and interpretation; decision to publish; or preparation, review, and approval of the manuscript.

Multimedia Appendix 1
Modifications in the study protocol made after the study commencement.
[PDF File (Adobe PDF File), 100 KB - jmir_v23i4e25228_app1.pdf]

Multimedia Appendix 2
Statements in the questionnaire on knowledge about dietary supplements and analysis of the items.
[PDF File (Adobe PDF File), 21 KB - jmir_v23i4e25228_app2.pdf]

Multimedia Appendix 3
The survey.
[PDF File (Adobe PDF File), 946 KB - jmir_v23i4e25228_app3.pdf]

Multimedia Appendix 4
Database cleaning procedure and results.
[PDF File (Adobe PDF File), 142 KB - jmir_v23i4e25228_app4.pdf]

Multimedia Appendix 5
Data missingness analysis.
[PDF File (Adobe PDF File), 125 KB - jmir_v23i4e25228_app5.pdf]

Multimedia Appendix 6
Data missingness analysis—continued.
[XLSX File (Microsoft Excel File), 51 KB - jmir_v23i4e25228_app6.xlsx]

Multimedia Appendix 7
Associations of knowledge about dietary supplements predicted with the biased-corrected model with characteristics of the study participants.
[PDF File (Adobe PDF File), 135 KB - jmir_v23i4e25228_app7.pdf]

References

1. Kantor ED, Rehm CD, Du M, White E, Giovannucci EL. Trends in dietary supplement use among US adults from 1999-2012. J Am Med Assoc 2016 Oct 11;316(14):1464-1474 [FREE Full text] [doi: 10.1001/jama.2016.14403] [Medline: 27727382]
2. Kamangar F, Emadi A. Vitamin and mineral supplements: do we really need them? Int J Prev Med 2012 Mar;3(3):221-226 [FREE Full text] [Medline: 22448315]

3. Sirico F, Miressi S, Castaldo C, Spera R, Montagnani S, Meglio FD, et al. Habits and beliefs related to food supplements: results of a survey among Italian students of different education fields and levels. PLoS One 2018 Jan 19;13(1):- [FREE Full text] [doi: 10.1371/journal.pone.0191424] [Medline: 29351568]
4. Bailey RL, Gahche JJ, Miller PE, Thomas PR, Dwyer JT. Why US adults use dietary supplements. JAMA Intern Med 2013 Mar 11;173(5):355-361. [doi: 10.1001/jamainternmed.2013.2299] [Medline: 23381623]
5. Conner M, Kirk SF, Cade JE, Barrett JH. Why do women use dietary supplements? The use of the theory of planned behaviour to explore beliefs about their use. Soc Sci Med 2001 Feb;52(4):621-633. [doi: 10.1016/s0277-9536(00)00165-9]
6. Dickinson A, Blatman J, El-Dash N, Franco JC. Consumer usage and reasons for using dietary supplements: report of a series of surveys. J Am Coll Nutr 2014 Apr 14;33(2):176-182. [doi: 10.1080/07315724.2013.875423] [Medline: 24724775]
7. Loiacono C, Palermi S, Massa B, Belviso I, Romano V, Gregorio AD, et al. Tendinopathy: pathophysiology, therapeutic options, and role of nutraceutics. A narrative literature review. Medicina (Kaunas) 2019 Aug 07;55(8):447 [FREE Full text] [doi: 10.3390/medicina55080447] [Medline: 31394838]
8. Meglio FD, Sacco A, Belviso I, Romano V, Camargo F, Loiacono C, et al. Influence of supplements and drugs used for the treatment of musculoskeletal disorders on adult human tendon-derived stem cells. Muscle Ligaments and Tendons J 2020 Sep;10(03):376. [doi: 10.32098/mltj.03.2020.04]
9. Gabriels G, Irhuma M. The potential impact of dietary supplement adulteration on patient assessment and treatment from a healthcare provider's perspective. S Afr Fam Pract 2019 May 31;61(6):260-263. [doi: 10.1080/20786190.2019.1609757]
10. Baudischova L, Straznicka J, Pokladnikova J, Jahodar L. The quality of information on the internet relating to top-selling dietary supplements in the Czech Republic. Int J Clin Pharm 2018 Feb 24;40(1):183-189. [doi: 10.1007/s11096-017-0564-x] [Medline: 29177614]
11. Jordan MA, Haywood T. Evaluation of internet websites marketing herbal weight-loss supplements to consumers. J Altern Complement Med 2007 Nov;13(9):1035-1043. [doi: 10.1089/acm.2007.7197] [Medline: 18047451]
12. Palmour N, Vanderbyl BL, Zimmerman E, Gauthier S, Racine E. Alzheimer's disease dietary supplements in websites. HEC Forum 2013 Dec 14;25(4):361-382. [doi: 10.1007/s10730-013-9217-1] [Medline: 23765585]
13. Zhao Y, Zhang J. Consumer health information seeking in social media: a literature review. Health Info Libr J 2017 Dec;34(4):268-283 [FREE Full text] [doi: 10.1111/hir.12192] [Medline: 29045011]
14. Schenker Y, Arnold RM, London AJ. The ethics of advertising for health care services. Am J Bioeth 2014 Mar 04;14(3):34-43. [doi: 10.1080/15265161.2013.879943] [Medline: 24592839]
15. Benigeri M, Pluye P. Shortcomings of health information on the internet. Health Promot Int 2003 Dec 01;18(4):381-386. [doi: 10.1093/heapro/dag409] [Medline: 14695369]
16. Cohen PA. The supplement paradox: negligible benefits, robust consumption. J Am Med Assoc 2016 Oct 11;316(14):1453-1454. [doi: 10.1001/jama.2016.14252] [Medline: 27727369]
17. Myung SK, Kim Y, Ju W, Choi HJ, Bae WK. Effects of antioxidant supplements on cancer prevention: meta-analysis of randomized controlled trials. Ann Oncol 2010 Jan;21(1):166-179 [FREE Full text] [doi: 10.1093/annonc/mdp286] [Medline: 19622597]
18. White CM. Dietary supplements pose real dangers to patients. Ann Pharmacother 2020 Aug 24;54(8):815-819. [doi: 10.1177/1060028019900504] [Medline: 31973570]
19. Sligo FX, Jameson AM. The knowledge?behavior gap in use of health information. J Am Soc Inf Sci 2000;51(9):858-869. [doi: 10.1002/(sici)1097-4571(2000)51:9<858::aid-asi80>3.0.co;2-q]
20. de Vries H, Pajor E, Curfs K, Eggers S, Oenema A. How to foster informed decision making about food supplements: results from an international Delphi study. Health Educ Res 2019 Aug 01;34(4):435-446 [FREE Full text] [doi: 10.1093/her/cyz017] [Medline: 31157369]
21. Poland and supplements: opportunities for European leadership. Polish Economic Institute. 2019. URL: https://web.archive.org/web/20200722201623/https://pie.net.pl/wp-content/uploads/2020/03/PIE-Poland_and_Supplements-EN.pdf [accessed 2021-03-10]
22. NIK o dopuszczaniu do obrotu suplementów diety - Najwyższa Izba Kontroli. Supreme Audit Office. 2017. URL: https://web.archive.org/web/20201020164937/https://www.nik.gov.pl/aktualnosci/nik-o-dopuszczaniu-do-obrotu-suplementow-diety.html [accessed 2020-07-20]
23. Aina BA, Ojedokun OA. Knowledge and use of dietary supplements by students of College of Medicine, University of Lagos, Idi-Araba, Lagos, Nigeria. J Basic Clin Pharm 2014 Mar;5(2):34-39 [FREE Full text] [doi: 10.4103/0976-0105.134952] [Medline: 25031497]
24. Axon DR, Vanova J, Edel C, Slack M. Dietary supplement use, knowledge, and perceptions among student pharmacists. Am J Pharm Educ 2017 Jun 01;81(5):92 [FREE Full text] [doi: 10.5688/ajpe81592] [Medline: 28720920]
25. Carvey CE, Farina EK, Lieberman HR. Confidence in the efficacy and safety of dietary supplements among United States active duty army personnel. BMC Complement Altern Med 2012 Oct 10;12(1):182 [FREE Full text] [doi: 10.1186/1472-6882-12-182] [Medline: 23051046]
26. Marinac JS, Buchinger CL, Godfrey LA, Wooten JM, Sun C, Willsie SK. Herbal products and dietary supplements: a survey of use, attitudes, and knowledge among older adults. J Am Osteopath Assoc 2007 Jan;107(1):13-21. [Medline: 17299031]

27. Molinero O, Márquez S. Use of nutritional supplements in sports: risks, knowledge, and behavioural-related factors. Nutr Hosp 2009;24(2):128-134. [Medline: 19593480]
28. Sharma A. Knowledge, attitude and practices related to dietary supplements and micronutrients in health sciences students. J Clin Diagn Res 2014:-. [doi: 10.7860/jcdr/2014/9329.4683]
29. Žeželj S, Tomljanović A, Jovanović G, Krešić G, Peloza O, Dragaš-Zubalj N, et al. Prevalence, knowledge and attitudes concerning dietary supplements among a student population in Croatia. Int J Environ Res Public Health 2018 May 23;15(6):1058 [FREE Full text] [doi: 10.3390/ijerph15061058] [Medline: 29882906]
30. Algaeed HA, AlJaber MI, Alwehaibi AI, AlJaber LI, Arafah AM, Aloyayri MA, et al. General public knowledge and use of dietary supplements in Riyadh, Saudi Arabia. J Family Med Prim Care 2019 Oct;8(10):3147-3154 [FREE Full text] [doi: 10.4103/jfmpc.jfmpc_511_19] [Medline: 31742134]
31. Alowais MA, Selim MAEH. Knowledge, attitude, and practices regarding dietary supplements in Saudi Arabia. J Family Med Prim Care 2019 Feb;8(2):365-372 [FREE Full text] [doi: 10.4103/jfmpc.jfmpc_430_18] [Medline: 30984640]
32. Dodge T, Litt D, Kaufman A. Influence of the dietary supplement health and education act on consumer beliefs about the safety and effectiveness of dietary supplements. J Health Commun 2011 Mar 28;16(3):230-244. [doi: 10.1080/10810730.2010.529493] [Medline: 21120738]
33. Kołodziej G, Cyran-Grzebyk B, Majewska J, Kołodziej K. Knowledge concerning dietary supplements among general public. Biomed Res Int 2019 Jun 24;2019:- [FREE Full text] [doi: 10.1155/2019/9629531] [Medline: 31341913]
34. Clement J. Global digital population as of July 2020. Statista. URL: https://web.archive.org/web/20201012105306/https://www.statista.com/statistics/617136/digital-population-worldwide/ [accessed 2020-10-13]
35. Kwan D, Hirschkorn K, Boon H. U.S. and Canadian pharmacists' attitudes, knowledge, and professional practice behaviors toward dietary supplements: a systematic review. BMC Complement Altern Med 2006 Sep 19;6(1):31 [FREE Full text] [doi: 10.1186/1472-6882-6-31] [Medline: 16984649]
36. Parmenter K, Wardle J. Development of a general nutrition knowledge questionnaire for adults. Eur J Clin Nutr 1999 Apr 6;53(4):298-308. [doi: 10.1038/sj.ejcn.1600726] [Medline: 10334656]
37. Steyn NP, Labadarios D, Nel JH, Heidi-Lee R. Development and validation of a questionnaire to test knowledge and practices of dietitians regarding dietary supplements. Nutrition 2005 Jan;21(1):51-58. [doi: 10.1016/j.nut.2004.09.008] [Medline: 15661478]
38. Karbownik MS, Paul E, Nowicka M, Nowicka Z, Kowalczyk RP, Kowalczyk E, et al. Knowledge about dietary supplements and trust in advertising them: development and validation of the questionnaires and preliminary results of the association between the constructs. PLoS One 2019 Jun 24;14(6):- [FREE Full text] [doi: 10.1371/journal.pone.0218398] [Medline: 31233516]
39. Karbownik MS. Determinants of knowledge about and use of dietary supplements. OSFHome. URL: https://osf.io/5e92d/ [accessed 2020-10-23]
40. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP, STROBE Initiative. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. J Clin Epidemiol 2008 Apr;61(4):344-349. [doi: 10.1016/j.jclinepi.2007.11.008] [Medline: 18313558]
41. Moons KGM, Altman DG, Reitsma JB, Ioannidis JPA, Macaskill P, Steyerberg EW, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. Ann Intern Med 2015 Jan 06;162(1):1-73. [doi: 10.7326/M14-0698] [Medline: 25560730]
42. Hilton CE. The importance of pretesting questionnaires: a field research example of cognitive pretesting the Exercise referral Quality of Life Scale (ER-QLS). Int J Soc Res Methodol 2015 Oct 07;20(1):21-34. [doi: 10.1080/13645579.2015.1091640]
43. Karbownik MS, Jankowska-Polańska B, Horne R, Górski KM, Kowalczyk E, Szemraj J. Adaptation and validation of the Polish version of the beliefs about medicines questionnaire among cardiovascular patients and medical students. PLoS One 2020 Apr 13;15(4):- [FREE Full text] [doi: 10.1371/journal.pone.0230131] [Medline: 32282826]
44. Horne R, Weinman J, Hankins M. The beliefs about medicines questionnaire: the development and evaluation of a new method for assessing the cognitive representation of medication. Psychol Health 1999 Jan;14(1):1-24. [doi: 10.1080/08870449908407311]
45. Paxton AE, Strycker LA, Toobert DJ, Ammerman AS, Glasgow RE. Starting the conversation performance of a brief dietary assessment and intervention tool for health professionals. Am J Prev Med 2011 Jan;40(1):67-71. [doi: 10.1016/j.amepre.2010.10.009] [Medline: 21146770]
46. Biernat E, Stupnicki R, Gajewski A. International Physical Activity Questionnaire (IPAQ) - Polish version. Physical Education and Sport. 2007. URL: https://www.researchgate.net/publication/234833002_International_Physical_Activity_Questionnaire_IPAQ_-_Polish_version [accessed 2021-03-10]
47. Badanie Gemius/PBI. SitePoint. URL: https://web.archive.org/web/20201012105904/http://pbi.org.pl/badania/ [accessed 2020-10-13]
48. Leiner DJ. Too fast, too straight, too weird: post hoc identification of meaningless data in internet surveys. SSRN Journal 2019:229-248. [doi: 10.2139/ssrn.2361661]
49. Dwyer JT, Coates PM, Smith MJ. Dietary supplements: regulatory challenges and research resources. Nutrients 2018 Jan 04;10(1):41 [FREE Full text] [doi: 10.3390/nu10010041] [Medline: 29300341]

50. Starr RR. Too little, too late: ineffective regulation of dietary supplements in the United States. Am J Public Health 2015 Mar;105(3):478-485. [doi: 10.2105/ajph.2014.302348]
51. Makowska M, Jasiński Ł. A discussion of the unresolved 2016/17 plans for regulating the Polish dietary supplements market. Health Policy 2019 Jun;123(6):544-549 [FREE Full text] [doi: 10.1016/j.healthpol.2019.04.001] [Medline: 30981554]
52. Parikh R, Renero A. Justified true belief: plato, gettier, turing. In: Floyd J, Bokulich A, editors. Philosophical Explorations of the Legacy of Alan Turing; Boston Studies in the Philosophy and History of Science. Switzerland: Springer; 2017:93-102.
53. Homer PM, Mukherjee S. Lay theories and consumer perceptions of dietary supplements. J Consumer Behav 2019 Jul 26;18(5):363-377. [doi: 10.1002/cb.1776]
54. Pajor EM, Oenema A, Eggers SM, de Vries H. Exploring beliefs about dietary supplement use: focus group discussions with Dutch adults. Public Health Nutr 2017 Aug 03;20(15):2694-2705. [doi: 10.1017/s1368980017001707]
55. Rong L, Kim J, Park J. The effects of internet shoppers' trust on their purchasing intention in China. J Inf Technol Manag 2007 Dec 01;4(3):269-286. [doi: 10.4301/s1807-17752007000300001]
56. Sun Y, Zhang Y, Gwizdka J, Trace CB. Consumer evaluation of the quality of online health information: systematic literature review of relevant criteria and indicators. J Med Internet Res 2019 May 02;21(5):- [FREE Full text] [doi: 10.2196/12522] [Medline: 31045507]
57. Waheed H, Anjum M, Rehman M, Khawaja A. Investigation of user behavior on social networking sites. PLoS One 2017 Feb 2;12(2):- [FREE Full text] [doi: 10.1371/journal.pone.0169693] [Medline: 28151963]
58. Kruger J, Dunning D. Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. J Pers Soc Psychol 1999;77(6):1121-1134. [doi: 10.1037/0022-3514.77.6.1121]
59. Wierzejska R. Whether the advertisement of dietary supplements is objective source of data about their impact on health? Analysis of broadcasting advertisements in the terms of the food law. Wiad Lek 2016;69(1):14-18. [Medline: 27162289]
60. Lane C, Bachmann R. Trust within and between organizations. Oxford, United Kingdom: Oxford University Press; 1998:1-352.
61. Zhang X, Zhang Q. Online trust forming mechanism: approaches and an integrated model. In: Proceedings of the 7th international conference on Electronic commerce. 2005 Presented at: 7th international conference on Electronic commerce; August 2005; Mainland, China. [doi: 10.1145/1089551.1089591]
62. Nooteboom B, Six F. The trust process in organizations : empirical studies of the determinants and the process of trust development. Cheltenham, United Kingdom: Edward Elgar Publishing; 2003:1-256.
63. Rao A, Miller P. Advertising/sales response functions. J Advert Res 1975;15(2):7-15.
64. Hesse BW, Nelson DE, Kreps GL, Croyle RT, Arora NK, Rimer BK, et al. Trust and sources of health information: the impact of the Internet and its implications for health care providers: findings from the first Health Information National Trends Survey. Arch Intern Med 2005;165(22):2618-2624. [doi: 10.1001/archinte.165.22.2618] [Medline: 16344419]
65. Alduraywish SA, Altamimi LA, Aldhuwayhi RA, AlZamil LR, Alzeghayer LY, Alsaleh FS, et al. Sources of health information and their impacts on medical knowledge perception among the Saudi Arabian population: cross-sectional study. J Med Internet Res 2020 Mar 19;22(3):- [FREE Full text] [doi: 10.2196/14414] [Medline: 32191208]
66. PoradnikZdrowie. PoradnikZdrowie.pl znów na czele - ABCzdrowie.pl, Medonet.pl i MP.pl w górę (TOP10 serwisów o zdrowiu). pl, Medonet. URL: https://web.archive.org/web/20201021053430/https://www.wirtualnemedia.pl/artykul/poradnikzdrowie-pl-liderem-serwisow-o-zdrowiu-ranking-abczdrowie-pl-medonet-pl-i-mp-pl-w-gore-top10
67. Suplementy diety - charakterystyka, składniki, odchudzanie. WP abcZdrowie. URL: https://web.archive.org/web/20201021054017/https://zywienie.abczdrowie.pl/suplementy-diety [accessed 2020-07-22]
68. Do czego potrzebne są suplementy?. URL: https://web.archive.org/web/20201021050313/https://www.doz.pl/czytelnia/a2142-Do_czego_potrzebne_sa_suplementy [accessed 2020-10-21]
69. Tan SSL, Goonawardene N. Internet health information seeking and the patient-physician relationship: a systematic review. J Med Internet Res 2017 Jan 19;19(1):e9 [FREE Full text] [doi: 10.2196/jmir.5729] [Medline: 28104579]
70. Lazarfeld P, Merton R. Mass communication, popular taste, and organized social action. Mass Communications. United States: University of Illinois Press; 1949. URL: https://pdfs.semanticscholar.org/50c6/e01e5e345853eab8bebec12b5d95b398f16a.pdf [accessed 2021-03-10]
71. Holdford D. Understanding the dynamics of the pharmaceutical market using a social marketing framework. J Consum Mark 2005 Dec;22(7):388-396. [doi: 10.1108/07363760510631129]
72. Boon H, Hirschkorn K, Griener G, Cali M. The ethics of dietary supplements and natural health products in pharmacy practice: a systematic documentary analysis. Int J Pharm Pract 2009 Feb 01;17(1):31-38 [FREE Full text] [doi: 10.1211/ijpp.17.1.0005] [Medline: 20218027]
73. Douglas PL, McCarthy H, McCotter LE, Gallen S, McClean S, Gallagher AM, et al. Nutrition education and community pharmacy: a first exploration of current attitudes and practices in Northern Ireland. Pharmacy (Basel) 2019 Mar 05;7(1):27. [doi: 10.3390/pharmacy7010027] [Medline: 30841590]
74. Baars EW, Hamre HJ. Whole medical systems versus the system of conventional biomedicine: a critical, narrative review of similarities, differences, and factors that promote the integration process. Evid Based Complement Alternat Med 2017;2017:- [FREE Full text] [doi: 10.1155/2017/4904930] [Medline: 28785290]

75. Truant TL, Balneaves LG, Fitch MI. Integrating complementary and alternative medicine into cancer care: Canadian oncology nurses' perspectives. Asia Pac J Oncol Nurs 2015;2(4):205-214. [doi: 10.4103/2347-5625.167233] [Medline: 27981116]
76. Nahin RL, Dahlhamer JM, Stussman BJ. Health need and the use of alternative medicine among adults who do not use conventional medicine. BMC Health Serv Res 2010 Jul 29;10(1):220. [doi: 10.1186/1472-6963-10-220] [Medline: 20670418]
77. Chao MT, Wade C, Kronenberg F. Disclosure of complementary and alternative medicine to conventional medical providers: variation by race/ethnicity and type of CAM. J Natl Med Assoc 2008 Nov;100(11):1341-1349. [doi: 10.1016/s0027-9684(15)31514-5]
78. Egan B, Hodgkins C, Shepherd R, Timotijevic L, Raats M. An overview of consumer attitudes and beliefs about plant food supplements. Food Funct 2011 Dec;2(12):747-752. [doi: 10.1039/c1fo10109a] [Medline: 22080158]
79. Foote JA, Murphy SP, Wilkens LR, Hankin JH, Henderson BE, Kolonel LN. Factors associated with dietary supplement use among healthy adults of five ethnicities: the Multiethnic Cohort Study. Am J Epidemiol 2003 May 15;157(10):888-897 [FREE Full text] [doi: 10.1093/aje/kwg072] [Medline: 12746241]
80. Ahlluwalia S, Singh S. Consumers' demographics as predictors of their perception of online advertising: are they still relevant in the e-landscape? Global Business Review 2020 Feb 18:-. [doi: 10.1177/0972150919897429]
81. Milakovic IK, Mihić M. Predictors and outcome of attitudes towards advertising: demographics, personal factors and wom. 2015. URL: https://www.semanticscholar.org/paper/PREDICTORS-AND-OUTCOME-OF-ATTITUDES-TOWARDS-FACTORS-Milakovic-Mihi%C4%87/f2f966d5c8ab9aa7188605dace9af91c6f15be55 [accessed 2021-03-10]
82. Shavitt S, Lowrey P, Haefner J. Public attitudes toward advertising: more favorable than you might think. J Advert Res. 1998. URL: https://experts.illinois.edu/en/publications/public-attitudes-toward-advertising-more-favorable-than-you-might [accessed 2021-03-10]
83. Gahche JJ, Bailey RL, Potischman N, Dwyer JT. Dietary supplement use was very high among older adults in the United States in 2011-2014. J Nutr 2017 Oct 30;147(10):1968-1976 [FREE Full text] [doi: 10.3945/jn.117.255984] [Medline: 28855421]
84. Phillips DM, Stanton JL. Age-related differences in advertising: recall and persuasion. J Target Meas Anal Mark 2004 Sep 1;13(1):7-20. [doi: 10.1057/palgrave.jt.5740128]
85. Diehl S, Mueller B, Terlutter R. The influence of demographic factors on the perception of humane-oriented (CSR) appeals in advertisements: a multi-country analysis. In: Rosengren S, Dahlén M, Okazaki S, editors. Advances in Advertising Research (Vol. IV), The Changing Roles of Advertising. Wiesbaden: Springer Gabler; 2013.
86. Wynants L, Collins GS, Van Calster B. Key steps and common pitfalls in developing and validating risk models. Brit J Obstet Gynec 2017 Feb;124(3):423-432. [doi: 10.1111/1471-0528.14170] [Medline: 27362778]
87. Chiba T, Kobayashi E, Okura T, Sekimoto M, Mizuno H, Saito M, et al. An educational intervention improved knowledge of dietary supplements in college students. BMC Public Health 2020 May 07;20(1):633 [FREE Full text] [doi: 10.1186/s12889-020-08786-3] [Medline: 32381078]
88. Begdache L, Kianmehr H, Heaney CV. College education on dietary supplements may promote responsible use in young adults. J Diet Suppl 2020 Sep 25;17(1):67-80. [doi: 10.1080/19390211.2018.1482983] [Medline: 30252551]
89. Internet use. Public Opinion Research Center. 2017. URL: https://web.archive.org/web/20201021053211/https://www.cbos.pl/SPISKOM.POL/2017/K_049_17.PDF [accessed 2021-03-10]
90. Latkin CA, Edwards C, Davey-Rothwell MA, Tobin KE. The relationship between social desirability bias and self-reports of health, substance use, and social network factors among urban substance users in Baltimore, Maryland. Addict Behav 2017 Oct;73:133-136 [FREE Full text] [doi: 10.1016/j.addbeh.2017.05.005] [Medline: 28511097]
91. Steene-Johannessen J, Anderssen SA, van der Ploeg HP, Hendriksen IJM, Donnelly AE, Brage S, et al. Are self-report measures able to define individuals as physically active or inactive? Med Sci Sports Exerc 2016 Feb;48(2):235-244 [FREE Full text] [doi: 10.1249/MSS.0000000000000760] [Medline: 26322556]
92. Bleeker SE, Moll HA, Steyerberg EW, Donders ART, Derksen-Lubsen G, Grobbee DE, et al. External validation is necessary in prediction research: a clinical example. J Clin Epidemiol 2003 Sep;56(9):826-832. [doi: 10.1016/s0895-4356(03)00207-5]
93. Loftfield E, Yi S, Immerwahr S, Eisenhower D. Construct validity of a single-item, self-rated question of diet quality. J Nutr Educ Behav 2015 Mar;47(2):181-187. [doi: 10.1016/j.jneb.2014.09.003] [Medline: 25449828]
94. Bachmann LM, Gutzwiller FS, Puhan MA, Steurer J, Steurer-Stey C, Gigerenzer G. Do citizens have minimum medical knowledge? A survey. BMC Med 2007 May 31;5(1):14 [FREE Full text] [doi: 10.1186/1741-7015-5-14] [Medline: 17540024]
95. Egleston BL, Miller SM, Meropol NJ. The impact of misclassification due to survey response fatigue on estimation and identifiability of treatment effects. Stat Med 2011 Dec 30;30(30):3560-3572 [FREE Full text] [doi: 10.1002/sim.4377] [Medline: 21953305]

Abbreviations

BMQ: beliefs about medicines questionnaire

DS: dietary supplement
GLM: general linear model
MAE: mean absolute error
RMSE: root mean squared error

Edited by R Kukafka; submitted 26.10.20; peer-reviewed by R Wierzejska, F Sirico, G Gabriels, I Shubina; comments to author 01.12.20; revised version received 10.12.20; accepted 04.02.21; published 21.04.21.

Please cite as:
Karbownik MS, Horne R, Paul E, Kowalczyk E, Szemraj J
Determinants of Knowledge About Dietary Supplements Among Polish Internet Users: Nationwide Cross-sectional Study
J Med Internet Res 2021;23(4):e25228
URL: https://www.jmir.org/2021/4/e25228
doi:10.2196/25228
PMID:33658173

Original Paper

User Perspectives of Diet-Tracking Apps: Reviews Content Analysis and Topic Modeling

Mila Zečević[1], MSc; Dejan Mijatović[2], MSc; Mateja Kos Koklič[1], PhD; Vesna Žabkar[1], PhD; Petar Gidaković[1], MSc

[1]School of Economics and Business, University of Ljubljana, Ljubljana, Slovenia

[2]Zuehlke Engineering, Schlieren, Switzerland

Corresponding Author:
Mila Zečević, MSc
School of Economics and Business
University of Ljubljana
Kardeljeva ploscad 17
Ljubljana, 1000
Slovenia
Phone: 386 15892429
Email: mila.zecevic@ef.uni-lj.si

Abstract

Background: The availability and use of mobile apps in health and nutrition management are increasing. Ease of access and user friendliness make diet-tracking apps an important ally in their users' efforts to lose and manage weight. To foster motivation for long-term use and to achieve goals, it is necessary to better understand users' opinions and needs for dietary self-monitoring.

Objective: The aim of this study was to identify the key topics and issues that users highlight in their reviews of diet-tracking apps on Google Play Store. Identifying the topics that users frequently mention in their reviews of these apps, along with the user ratings for each of these apps, allowed us to identify areas where further improvement of the apps could facilitate app use, and support users' weight loss and intake management efforts.

Methods: We collected 72,084 user reviews from Google Play Store for 15 diet-tracking apps that allow users to track and count calories. After a series of text processing operations, two text-mining techniques (topic modeling and topical n-grams) were applied to the corpus of user reviews of diet-tracking apps.

Results: Using the topic modeling technique, 11 separate topics were extracted from the pool of user reviews. Most of the users providing feedback were generally satisfied with the apps they use (average rating of 4.4 out of 5 for the 15 apps). Most topics referred to the positive evaluation of the apps and their functions. Negatively rated topics mostly referred to app charges and technical difficulties encountered. We identified the positive and negative topic trigrams (3-word combinations) among the most frequently mentioned topics. Usability and functionality (tracking options) of apps were rated positively on average. Negative ratings were associated with trigrams related to adding new foods, technical issues, and app charges.

Conclusions: Motivating users to use an app over time could help them better achieve their nutrition goals. Although user reviews generally showed positive opinions and ratings of the apps, developers should pay more attention to users' technical problems and inform users about expected payments, along with their refund and cancellation policies, to increase user loyalty.

(J Med Internet Res 2021;23(4):e25160) doi:10.2196/25160

KEYWORDS

diet-tracking apps; mobile apps; user reviews; topic modeling; n-grams; mHealth; nutrition; diet; well-being

Introduction

Obesity and overweight are the result of a plethora of environmental factors that are known to influence individuals' food intake and physical activity [1,2]. As obesity is a global public health challenge that leads to numerous health, social, and economic difficulties [3], it is important to help individuals make better and healthier dietary choices. Changing lifestyles (eg, the modern sedentary lifestyle) along with other features of an obesogenic environment (eg, physical design challenges; political, social, and cultural factors; the (in)ability to promote an active lifestyle; better food choices) have created a complex environment for consumers to manage their weight and diet [4,5]. Despite this dysfunctional environment, people are aware

of the negative impact that inappropriate diets can have on their health and well-being, and are generally interested in making healthier choices and educating themselves about the foods they consume and their nutritional value [6].

Technological advancements, including those related to mobile devices, are enabling developments of an increasing number of tools to help individuals take control of their health and nutrition. Mobile apps have become an important source of information for their users. A study conducted in the United States showed that more than half of the population using a mobile phone has downloaded at least one health-related app. Most of these were fitness and nutrition apps, which also have the highest reported usage rates [7,8].

Particularly in the area of nutrition, people use apps for a variety of purposes, including to learn about products, read or contribute recipes, interact in app community forums, evaluate their food choices, check product labels, and obtain an overview of the healthiness of products [9], as well as to track their food intake and diets [8]. Given the importance of self-monitoring and diet tracking in motivating behavior change and persistence in healthier eating [8,10], we have been focusing on the nutrition apps that provide a calorie-tracking option (ie, diet-tracking apps). These apps promote diet tracking and maintenance of healthier habits and do not necessarily include weight loss as a goal, which has often been the focus of previous studies (eg, [7]).

The low inclusion of behavior change strategies in diet tracking apps may hinder their ability to help users achieve their long-term diet and nutrition goals [11,12]. However, diet-tracking apps that successfully employ behavior change strategies can have a positive effect on their users' motivation, habits, and diet and nutrition outcomes [13-16]. These apps have also proven to be helpful in behavioral control and weight management [15]. Self-monitoring is considered the cornerstone of successful weight management [17], and can be particularly helpful when combined with tailored goals [16]. Goal setting, which diet-tracking apps enable, is one of the relevant factors influencing behavior change, along with motivation and self-efficacy [15]. Together with apps that offer personalized meal planning programs [18], there is evidence that diet-tracking apps can efficiently support behavior change and weight management [18,19].

Nevertheless, the usability of diet-tracking apps in the weight management process has not always been positively assessed. Intervention studies focusing on diet-tracking apps show that users sometimes dislike the apps due to their complexity, lack of personalization and long-term support for users, and the focus on calorie counting, which can easily become an obsession [20]. Maintaining high levels of motivation throughout the weight management journey is challenging for most people, and diet-tracking apps sometimes fail to provide proper motivational support for sustainable weight loss and weight management [20,21]. In addition, the long-term effects of diet-tracking apps on food intake and user behavior remain unexplored [16]. These issues indicate that additional research on diet-tracking apps and their features is needed [7].

As consumers increasingly rely on apps to support their daily activities, they also generate invaluable feedback for both developers and potential users through app reviews and ratings. These reviews typically contain information that is valuable for app evaluation, including user opinions about the app, information about their experiences with the app, and bug complaints or feature suggestions [22]. A previous study showed that almost a quarter (23.3%) of app reviews contain an app feature request or app assessment [23].

App developers and companies have recognized the value of user reviews and frequently examine these reviews to improve the user experience. Most mobile health apps are free to use, with the option to upgrade profiles to paid premium options (for more features or personalized advice). The initial free download and use make these apps easily disposable and users' decision to switch between them is widespread [24]. This drives app developers' interest in listening to users' comments, complaints, ideas, and suggestions [25]. In reviews, users are text producers for other potential consumers, businesses, and society at large. This text can then be used to predict and understand user preferences and behaviors [26].

To increase usage and enable better health and nutrition outcomes, it is necessary to better understand user needs. It is evident that users are not simply looking for pure information when using mobile health apps. The way the information is presented; the usability of the app; the degree to which it engages and connects users; and its effectiveness, timeliness, design, and functionality are also important considerations [27]. A deeper analysis of user reviews of apps can lead to better knowledge of desired features and user preferences, and ideally increase the usability, appeal, and effectiveness of the app in achieving users' health and nutrition goals [7]. These aspects require more academic research into user reviews [22], especially as functionality and appearance often have greater contributions to the popularity of dieting apps than the quality of the information provided by the app [28].

Owing to their presence and relevance in the dieting field, diet-tracking apps have attracted the interest of many researchers who have used app evaluation strategies in an attempt to better understand and evaluate app features [29-32]. The influence of diet-tracking apps on users' food choices and their opinions about these apps have been tested using experimental and survey data collection methods [10,15,18,33]. Previous studies also assessed the consistency of information provided by different apps, and recommended further collaboration and harmonization of information [34,35].

To improve the understanding of user opinions on diet-tracking apps, this study was performed based on the collection and analysis of the textual reviews and numerical ratings that users leave for these apps. We focused on identifying the main positive and negative aspects that users express about diet-tracking apps and openly share in app reviews. By identifying the features or functions that attract consumers' attention, this study suggests areas for app development and improvement that have potential to increase users' positive evaluation and motivation, leading to nutrition/health improvement.

Methods

Data Collection

To evaluate users' views on diet-tracking apps, we used a different methodology from the experimental and survey data collection methods previously reported [10,15,18,33]. We performed a thorough content analysis of the app user reviews available on Google Play Store. The text mining method was used to detect the word and topic combinations that are frequently mentioned in user reviews. These word combinations (n-grams) were evaluated and the most frequent combinations were selected for further analysis. The identified topics were named and are briefly discussed.

The search for diet-tracking apps in Google Play Store included the following keywords: "nutrition apps," "diet apps," "calorie counter apps," "food scanner," "calorie app," "calorie tracker," and "calorie scanner." We identified 131 unique apps that offer calorie information (eg, calorie tables, calorie tracking, diet diaries). These apps were further reviewed to determine if they offered a calorie counting option for food items and an intake diary for users, and if their operating language was English. Final app selection was based on the number of downloads and reviews (impact evaluation; apps with over 1 million downloads and at least 10,000 user ratings were selected), which is a standard procedure in the app assessment process (eg, [7,30]). A total of 15 apps were identified that met all of the above conditions (Table 1), and their reviews were scraped from Google Play Store using a custom-written Python script. In total, 81,660 of the latest user reviews for these apps were collected at this stage.

Table 1. Overview of apps included in the study.

App	Mean app rating	Number of downloads	Number of ratings
Health Pal-Fitness, Weight Loss Coach, Pedometer	4.1	>1 million	>20,000
iEatBetter: Food Diary	4.2	>1 million	>20,000
Health & Fitness Tracker with Calorie Counter	4.1	>1 million	>20,000
Calorie Counter by Fat Secret	4.7	>10 million	>300,000
Calorie Counter by Lose It!	4.6	>10 million	>100,000
Fooducate-Eat better. Lose weight. Get healthy.	4.4	>1 million	>15,000
Calorie Counter-MyNetDiary, Food Diary Tracker	4.6	>1 million	>40,000
Healthify me-Calorie Counter, Weight Loss Coach	4.5	>10 million	>100,000
MyPlate Calorie Tracker	4.6	>1 million	>30,000
Calorie Counter-My Fitness Pal	4.4	>50 million	>2 million
Lifesum-Diet Plan, Macro Calculator & Food Diary	4.4	>10 million	>200,000
Noom: Health & Weight	4.4	>10 million	>200,000
YAZIO Calorie Counter, Nutrition Diary & Diet Plan	4.6	>10 million	>300,000
Calorie, Carb & Fat Counter	4.5	>1 million	>50,000
Calorie Counter Calories!	4.5	>1 million	>10,000

Data Analysis

Text Mining Approaches

With the increase in publicly available user-generated content due to the proliferation of internet-assisted communication, researchers have developed several automated approaches to identify, summarize, and classify the available information [26,36]. The development of new tools allows researchers to obtain more information about users' opinions and sentiments in their writing. There is a trend to shift the focus of opinion mining from studying long texts to shorter user posts on various social media platforms and websites [22]. The rapid development and increase in the efficiency and capabilities of text mining and natural language processing (NLP) algorithms is evident, and they are becoming an important part of social science research in the study of user-generated content [37,38]. Speed, reproducibility, and reliability are considered some of the most important advantages of text mining when it comes to classifying and categorizing text [39].

In this study, we applied two text mining methods to our dataset: topical n-grams identification and topic modeling. Both methods work by identifying and grouping words that occur simultaneously in the text (user reviews in our case). Data analysis required preprocessing of the raw data, which was performed through several procedures commonly used in data preparation and preprocessing for text mining analysis [40].

Data Preprocessing

Data preprocessing is a data mining technique that transforms raw data into an understandable format. Real-world data are often incomplete, inconsistent, contain a substantial amount of redundant information, and are likely to include many errors [41]. This is a critical and time-consuming process as the output depends on the quality of the data.

Since both the topical n-grams identification and topic modeling approaches have the same preprocessing steps, the same preprocessed dataset was used in both methods. For these tasks, we used the Python programming language in combination with its data science–specific tools (ie, libraries) that made this process possible given the large amount of data. The Python libraries numpy and pandas, which are well known in the data science community, were used extensively throughout the process, along with several other libraries, each specialized for a particular task. The following data preprocessing steps were applied.

First, we removed all non-English reviews. We were only interested in English reviews at this point since our dataset contains reviews in different languages such as Portuguese, Spanish, and German. These reviews make up about 12% of our dataset (9576 reviews), and it was safer to remove them than to translate them into English. A total of 72,084 user reviews in English were identified in this step using the Python library langdetect. These reviews were used for all analyses performed in this study.

We then converted the text to lowercase, performed an extensive spell check of every review, and made necessary corrections using the Speller Python library. Words such as "I," "are," "and," and "the" were considered "stop words" and removed, as such common words tend to dominate the results. We further removed any special characters and numbers from the reviews.

Second, lemmatization was performed, which is a process of grouping the inflected forms of words so that they can be analyzed as a single item. These forms are identified by the lemma of the word (ie, both "tracking" and "tracks" share the same lemma and become "track"). The lemmatization algorithm considers the morphological analysis of the words during data preparation [42]. Finally, we applied part-of-speech tagging, which determines the category of the word (eg, noun, adjective, adverb, verb) based on both its definition and context [43].

After data preprocessing, a new dataset was obtained with cleaned data that could be used for both topic modeling and n-grams identification. This new dataset comprised a collection of arrays of words. For example, the item in the previous dataset ''I love this app and it's easy to use'' is converted into the word array "love app easy use," which was used as such for further analysis.

Topical N-Grams Identification

The use of topical n-grams is common in text and topic mining, as is the NLP approach when tracking word or phrase frequencies [44,45]. For the purpose of this study, we implemented the analysis of the most frequent trigrams (ie, groups of three words used together) occurring in user reviews (textual user evaluation of the app). For example, applying this to our previously mentioned word array from the cleaned dataset "love app easy use," we could extract exactly two trigrams: (love, app, easy) and (app, easy, use).

These two trigrams would then be added to the trigrams extracted from other reviews, resulting in a total of 744,808 trigrams from 72,084 reviews.

Evaluation of the word combination was then used in conjunction with the users' numerical app assessment (ie, rating), which is often used as a proxy for sentiment (eg, [46,47]). User ratings for apps in Google Play Store are scored on a scale from 1 to 5, where 1 is the worst and 5 is the best rating an app can receive. In our study, scores below 3 were associated with a negative app assessment, whereas scores of 4 and 5 were considered positive. This approach is common in studies in this field (eg, [46]). The ratings were averaged by grouping ratings from reviews containing given trigrams.

Topic Modeling

Topic modeling is another text-mining and NLP method that is commonly used to discover latent topics in a corpus of text. Topic modeling has been shown to be useful for clustering documents or text, and is considered a probabilistic statistical technique for semantic structures [48]. In this study, we used the Python Gensim library, which is commonly used in NLP, for topic modeling analysis. This library helped us to build a mathematical model that could classify each review by topic. The list of possible topics was determined during model training, and we predetermined the number of possible topics. To find the most appropriate number of topics, we used the coherence score. Topic coherence measures the degree of semantic similarity between the highly scored words in the topic, which can help to distinguish between topics that are semantically interpretable and topics that are artifacts of statistical inference [49]. This value is given after each model training process and helped us determine the performance of our trained model.

Finding the best number of topics that would give optimal results required several trials, starting with a randomly selected number of topics until we narrowed down to the model with the best score. We would simply select this model and apply it to our dataset. For example, if our model found 11 topics in the dataset, for each review in our dataset, the model would provide us with the probabilities of how likely the review is to belong to each of the 11 topics.

Results

Topic Modeling

Topic Selection Process

Since topic modeling is an unsupervised method, it was not constrained by certain predefined standards (ie, number of topics). Instead, for the first run, we programmed the script to start with only 2 topics, repeat and increase by 4 (since this is a computationally intensive and demanding process, we had to minimize the number of runs) until reaching 30 (ie, finding optimal number of topics anywhere between 2 and 30 topics). This high number was randomly chosen to find the optimal range for our topic number. This analysis revealed that the number of topics with the best coherence score was between 6 and 13 (Figure 1). For the second iteration, we again reprogrammed the script to repeatedly run in this range, this time increasing by 1 to more accurately determine the best score. The best coherence (0.646) was achieved with 11 topics (Figure 2, Table 2).

Figure 1. Topics coherence score (range between 2 and 30 topics). Num: number.

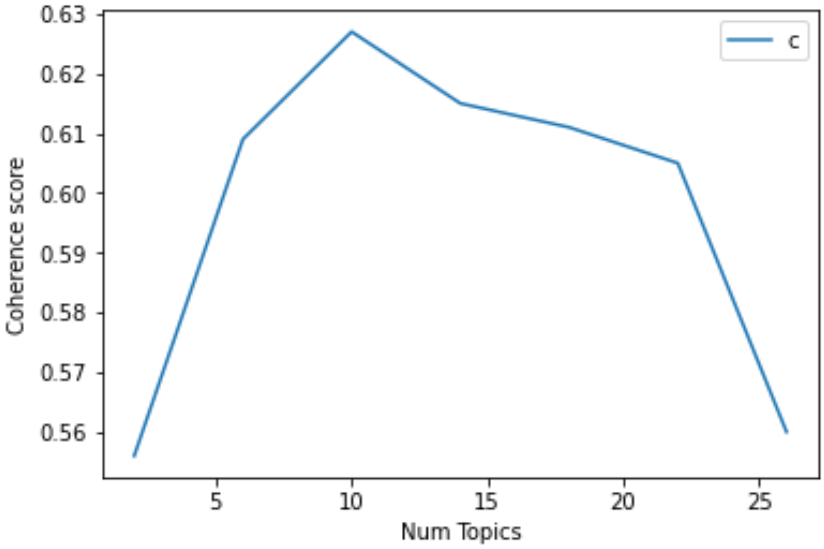


Figure 2. Topics coherence score (range between 6 and 13 topics). Num: number.

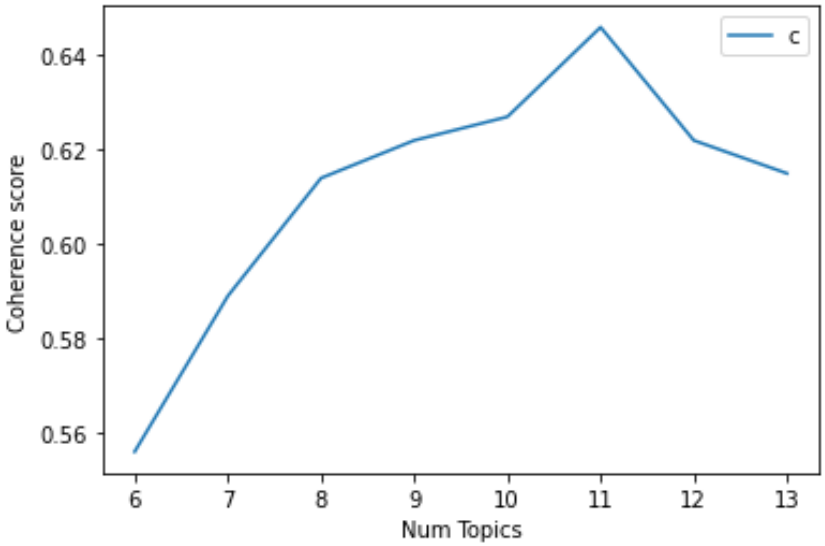


Table 2. Coherence scores for 6 to 13 topics.

Number of topics	Coherence score
6	0.59
7	0.618
8	0.617
9	0.614
10	0.614
11	0.646
12	0.618
13	0.622

Identified Topics

Most of the identified topics included the use of positive words when describing apps in the reviews. In their feedback, users often use words such as “love,” “nice,” “easy,” “good,” and “amaze” to describe the apps. Positively rated topics were more common than negatively rated topics. Users who leave feedback for diet-tracking apps positively rate the possibility to track their

food intake; use food scanners and create/access food databases in the apps; and consider the apps to be user-friendly, convenient, and easy to use overall. Weight loss was another important topic, appearing in 10% of user reviews (Table 3).

By pairing the topics with the average ratings of user reviews from the topic, we found that difficulties with cancellations, payment plans, and charges seem to bother users the most (topic average rating 2.32). One of the users described her experience as follows:

> *Personally I felt company was highly interested in my weight loss journey before I activated plan then no one cares about me, or I noticed they've charged me for another 3 months subscription which I did not authorize. I've made 3 attempts via email to make contact with [app] to end my membership and request a refund and I received an email back stating they'll be in contact in the next 48hrs but I have never heard back from them.*

In addition, technical difficulties appear to create issues in using the app (average topic rating 2.84):

> *Good app when it works. Otherwise, there's too many bugs. It lags too often and takes a long time to load…*
>
> *Stopped working. When I try to add food or search it's a blank screen. Please fix!*

Table 3. Modeling results for 11 selected topics.

Topic	Words	Mean rating	Proportion of reviews (%)
Health and fitness tracking	App, great, work, good, health, step, fitness, tracker, sync, google	4.413731	10.00
Macros tracking	App, great, love, track, awesome, carbs, fat, diet, macro, feature	4.590072	8.13
App praising	App, good, food, thing, lot, find, put, log, info, pretty	4.342717	7.34
App support	Give, room, program, coach, support, day, information, follow, people, plan	4.264332	7.21
App charges	Free, pay, version, plan, premium, cancel, money, charge, month, trial	2.329310	8.96
Weight loss	Weight, lose, week, goal, loss, start, pound, year, month, set	4.697312	10.94
Intake tracking	Calorie, track, exercise, intake, count, daily, day, water, great, burn	4.578164	10.06
Food adding and database	Food, add, meal, option, item, database, recipe, enter, list, search	3.872253	8.77
App “loving”	Easy, love, food, helpful, scan, user, simple, find, scanner, feature	4.709999	12.67
Diet change	Eat, make, change, diet, healthy, recommend, choice, learn, life, habit	4.780649	8.37
Technical issues	Time, work, log, update, day, star, back, phone, issue, problem	2.848619	7.53

Topical Trigrams

Overall Ratings

Similar to the topic modeling results, our overall trigram analysis suggested that, on average, users rate the diet-tracking apps positively in their reviews (Table 4). Trigrams indicating the apps’ ease of use and helpfulness (21/50 top trigrams by frequency of mention), the expression of liking/loving the app (13/50 top trigrams by frequency of mention), and the apps’ tracking option (19/50 top trigrams by frequency of mention) dominated the top 50 trigrams identified from the pool of user reviews. These topics were also frequently associated with high average app ratings; 49 of the top 50 trigrams by frequency attained ratings over 4.

Table 4. Top 50 most frequently mentioned trigrams.

Trigram	Count	Mean rating	Category
(app, easy, use)	948	4.772152	Easy Use/Help
(help, keep, track)	895	4.773184	Easy Use/Help; Tracking
(keep, track, calorie)	591	4.626058	Tracking
(help, lose, weight)	457	4.628009	Easy Use/Help; Weight Loss
(help, stay, track)	345	4.837681	Easy Use/Help; Tracking
(app, keep, track)	341	4.750733	Tracking
(really, like, app)	309	4.281553	App Liking
(app, really, help)	307	4.856678	Easy Use/Help
(track, calorie, intake)	303	4.669967	Tracking
(keep, track, eat)	296	4.733108	Tracking
(easy, use, love)	290	4.858621	Easy Use/Help; App Liking
(track, food, intake)	283	4.614841	Tracking
(keep, track, food)	263	4.653992	Tracking
(app, track, calorie)	258	4.701550	Tracking
(love, app, help)	246	4.857724	Easy Use/Help; App Liking
(love, app, easy)	244	4.831967	Easy Use/Help; App Liking
(great, app, track)	241	4.726141	App Liking; Tracking
(great, app, easy)	239	4.861925	App Liking; Easy Use/Help
(bar, code, scanner)	236	4.199153	N/A[a]
(app, help, keep)	235	4.804255	Easy Use/Help; Tracking
(great, app, help)	226	4.774336	App Liking; Easy Use/Help
(easy, use, great)	224	4.843750	Easy Use/Help
(way, keep, track)	222	4.729730	Tracking
(easy, use, helpful)	216	4.847222	Easy Use/Help
(really, help, keep)	204	4.803922	Easy Use/Help; Tracking
(make, good, choice)	204	4.779412	N/A
(weight, loss, journey)	201	4.731343	Weight Loss
(use, app, year)	201	4.129353	N/A
(easy, keep, track)	201	4.850746	Easy Use/Help; Tracking
(app, help, lose)	200	4.755000	Weight Loss; Easy Use/Help
(easy, use, help)	191	4.858639	Easy Use/Help
(really, easy, use)	188	4.781915	Easy Use/Help
(weight, loss, program)	186	4.688172	Weight Loss
(super, easy, use)	180	4.911111	Easy Use/Help
(easy, use, keep)	173	4.809249	Easy Use/Help; Tracking
(great, app, keep)	171	4.719298	App Liking; Tracking
(easy, use, app)	171	4.666667	Easy Use/Help
(keep, track, everything)	169	4.869822	Tracking
(really, good, app)	169	4.426036	App Liking
(use, free, version)	165	4.230303	NA
(good, app, track)	163	4.595092	App Liking; Tracking
(weight, loss, goal)	161	4.714286	Weight Loss

Trigram	Count	Mean rating	Category
(start, use, app)	160	4.387500	N/A
(scan, bar, code)	158	4.227848	N/A
(try, lose, weight)	157	4.414013	Weight Loss
(use, keep, track)	154	4.779221	Tracking
(absolutely, love, app)	153	4.790850	App Liking
(love, app, use)	144	4.472222	App Liking
(would, give, star)	144	3.638889	N/A
(best, app, ever)	142	4.929577	App Liking

[a]N/A: not applicable; no relevant category.

The number of mentions of positive and negative trigrams in user reviews also showed a trend of positive evaluation dominance among users leaving reviews. The top 50 most frequent positive trigrams appeared 12,723 times, while the top 50 most frequent negative trigrams were mentioned 1270 times in our dataset of 72,084 user reviews.

Positively Evaluated App Characteristics

With respect to positively valenced user ratings, we identified a significant presence of reviews praising the apps in general. Most of the top 50 positively rated user reviews refer to the apps' ease of use and helpfulness (21/50 top positive trigrams by frequency of mention), intake and calorie tracking (20/50 top positive trigrams by frequency of mention), and weight loss (6/50 top positive trigrams by frequency of mention) (Table 5). Users generally characterized the apps as "very easy to use and intuitive" and "easy to use and pretty straightforward." The ability to track their intake is described as "keeps me accountable" and "very informative about the diets you choose." Some users also viewed the apps as contributors to their better and healthier food choices:

> *Helps get a better understanding of the different foods calorie loads so I can make better choices.*
>
> *Easy to check total carb, fat and protein content and individual food values so I can make better choices next day.*

Similar comments were found for weight loss:

> *Teaches you how and why you need to change your eating habits;*
>
> *Mind changing weight loss program;*
>
> *This is about sustainable weight loss...*

Table 5. Top 50 “positive” trigrams (most frequently mentioned trigrams with ratings over 4).

Trigrams	Count	Mean rating	Category
(app, easy, use)	948	4.772152	Easy Use/Help
(help, keep, track)	895	4.773184	Easy Use/Help; Tracking
(keep, track, calorie)	591	4.626058	Tracking
(help, lose, weight)	457	4.628009	Easy Use/Help; Weight Loss
(help, stay, track)	345	4.837681	Tracking; Easy Use/Help
(app, keep, track)	341	4.750733	Tracking
(really, like, app)	309	4.281553	App Liking
(app, really, help)	307	4.856678	Easy Use/Help
(track, calorie, intake)	303	4.669967	Tracking
(keep, track, eat)	296	4.733108	Tracking
(easy, use, love)	290	4.858621	Easy Use/Help; App Liking
(track, food, intake)	283	4.614841	Tracking
(keep, track, food)	263	4.653992	Tracking
(app, track, calorie)	258	4.701550	Tracking
(love, app, help)	246	4.857724	Easy Use/Help; App Liking
(love, app, easy)	244	4.831967	Easy Use/Help; App Liking
(great, app, track)	241	4.726141	App Liking; Tracking
(great, app, easy)	239	4.861925	App Liking; Easy Use/Help
(bar, code, scanner)	236	4.199153	N/A[a]
(app, help, keep)	235	4.804255	Easy Use/Help; Tracking
(great, app, help)	226	4.774336	Easy Use/Help; App Liking
(easy, use, great)	224	4.843750	Easy Use/Help
(way, keep, track)	222	4.729730	Tracking
(easy, use, helpful)	216	4.847222	Easy Use/Help
(really, help, keep)	204	4.803922	Easy Use/Help; Tracking
(make, good, choice)	204	4.779412	N/A
(weight, loss, journey)	201	4.731343	Weight Loss
(use, app, year)	201	4.129353	N/A
(easy, keep, track)	201	4.850746	Easy Use/Help; Tracking
(app, help, lose)	200	4.755000	Weight Loss; Easy Use/Help
(easy, use, help)	191	4.858639	Easy Use/Help
(really, easy, use)	188	4.781915	Easy Use/Help
(weight, loss, program)	186	4.688172	Weight Loss
(super, easy, use)	180	4.911111	Easy Use/Help
(easy, use, keep)	173	4.809249	Easy Use/Help; Tracking
(great, app, keep)	171	4.719298	App Liking; Tracking
(easy, use, app)	171	4.666667	Easy Use/Help
(keep, track, everything)	169	4.869822	Tracking
(really, good, app)	169	4.426036	App Liking
(use, free, version)	165	4.230303	N/A
(good, app, track)	163	4.595092	App Liking; Tracking
(weight, loss, goal)	161	4.714286	Weight Loss

Trigrams	Count	Mean rating	Category
(start, use, app)	160	4.387500	N/A
(scan, bar, code)	158	4.227848	N/A
(try, lose, weight)	157	4.414013	Weight Loss
(use, keep, track)	154	4.779221	Tracking
(absolutely, love, app)	153	4.790850	App Liking
(love, app, use)	144	4.472222	App Liking
(best, app, ever)	142	4.929577	App Liking
(app, track, food)	142	4.549296	Tracking

[a]N/A: not applicable; no relevant category.

Negatively Evaluated App Characteristics

Based on our results, an aggressive approach to advertising premium app options, and unclear policies of subscription charges and cancellations seem to be particularly problematic (15/50 top negative trigrams by frequency of mentions). The reviews also indicate that constant display of ads and reminders about premium app options often lead users to delete the app, as illustrated by the following excerpts from the reviews:

> *Every time I open the app it pushes its premium service at me. I get that they are here to make money, but seriously, just throw a nonintrusive ad window in somewhere and don't pester me.*
>
> *Go for alternatives until these guys stop giving you ads for paid plans.*
>
> *NO MONEY BACK NO MATTER WHAT. This app didn't work for me since I'm not overweight. I tried it to give it a chance thinking it was what I was looking for, but it wasn't. and then found out that NO MATTER what, you can't get your money back once they have charged the subscription even if it's on THE SAME DAY.*

Given the size of the apps in question and their numbers of users, it is easy to see how the ability to upload content could be difficult to manage from a technical perspective. Nevertheless, 14 of the 50 most frequently mentioned negative trigrams refer to technical issues experienced by users (Table 6), which include app crashing, inability to use or open the app, and similar:

> *It freezes every time I try to add food or exercise.*
>
> *Not sure what's going on however ever since paid ads keep popping up the app has just gone down hill. Today in particular has been awful. Screen goes black, freezes. Constant crashing. I've used this app for 3 years now and am seriously looking at using another app.*
>
> *I really wanted to give this app 5 stars, especially since it's helped me to lose more than 20 pounds in the last 6 weeks. Unfortunately, the app itself is so laggy & buggy that I can't give it more than 1-star. Every time I shift between apps, [app] needs 15-60 seconds to start up. The whole app crashes on me at least 10 times a day.*

Some content-related complaints could also be found in reviews, such as "need to be able to add new food with more than the 100 grams" or "needs an option to let users easily add new foods and correct scanned foods that have incorrect nutritional data."

Adding new foods creates additional issues for users (5/50 top frequent negative trigrams). Namely, users' complaints in this area usually refer to the inability to add a product to the database due to technical challenges:

> *There's an option to add a new food item, but no way to save it.*
>
> *Would have been a perfect app but becomes utterly useless when trying to input my own foods. Everyone I enter the nutritional info it changes everything I put in to insane numbers like 2800 calories for cottage cheese.*
>
> *The reason I gave this a 4 is because the app doesn't always keep my information I add about new foods. It constantly says it's downloading the database for days on end.*

Table 6. Top 50 negative trigrams (most frequently mentioned trigrams with ratings lower than 3).

Trigrams	Count	Mean rating	Category
(every, time, try)	75	2.080000	Technical issues
(use, love, app)	71	2.478873	N/A[a]
(get, money, back)	50	1.240000	Charges/Ads
(day, free, trial)	43	1.674419	Charges/Ads
(can, not, get)	40	1.650000	NA
(try, add, food)	39	2.435897	Adding Food
(since, last, update)	35	2.685714	Technical issues
(sign, free, trial)	33	1.303030	Charges/Ads
(every, time, open)	33	2.484848	Technical issues
(time, open, app)	30	2.366667	Technical issues
(can, not, use)	30	1.633333	N/A
(app, keep, crash)	28	2.107143	Technical issues
(even, use, app)	26	1.653846	N/A
(add, food, meal)	26	2.884615	Adding Food
(bad, app, ever)	26	1.000000	N/A
(get, new, phone)	26	2.692308	N/A
(heart, rate, monitor)	26	2.884615	N/A
(pay, monthly, fee)	25	2.560000	Charges/Ads
(want, money, back)	25	1.160000	Charges/Ads
(every, time, go)	25	2.880000	Technical issues
(try, cancel, subscription)	24	1.166667	Charges/Ads
(app, stop, work)	24	1.916667	Technical issues
(waste, time, money)	24	1.166667	N/A
(can, not, add)	24	2.250000	Adding Food
(charge, credit, card)	24	1.041667	Charges/Ads
(never, use, app)	21	1.952381	Adding Food
(app, can, not)	21	2.142857	Technical issues
(every, single, time)	20	2.400000	Technical issues
(wish, could, give)	20	2.250000	N/A
(get, error, message)	19	1.736842	Technical issues
(something, go, wrong)	18	1.388889	Technical issues
(change, serve, size)	18	2.666667	N/A
(try, use, app)	18	1.888889	N/A
(would, great, app)	18	2.944444	N/A
(free, trial, end)	17	1.470588	Charges/Ads
(use, different, app)	17	2.941176	N/A
(use, app, without)	17	2.705882	N/A
(able, use, app)	17	2.470588	N/A
(try, get, refund)	17	1.470588	Charges/Ads
(give, money, back)	17	1.470588	Charges/Ads
(can, not, log)	17	2.058824	Technical issues
(create, new, account)	17	1.294118	Technical issues

Trigrams	Count	Mean rating	Category
(can, not, enter)	17	2.470588	Technical issues
(cancel, free, trial)	17	1.235294	Charges/Ads
(want, cancel, subscription)	16	1.437500	Charges/Ads
(message, goal, specialist)	16	1.875000	N/A
(create, new, food)	16	2.937500	Adding Food
(two, week, trial)	16	2.875000	Charges/Ads
(buy, pro, version)	16	2.812500	Charges/Ads
(ad, pop, every)	15	1.733333	Charges/Ads

[a]N/A: not applicable; no relevant category.

Discussion

Principal Findings

Despite extensive literature on nutrition apps and individual usage patterns in health and nutrition research, the investigation of these apps from the perspective of user-generated content (ie, publicly available user reviews) is in its infancy. Previous research has mainly focused on app development issues and feature evaluation to make apps more accessible and user-friendly (eg, [7,31]). Numerous studies have examined consumer opinions of existing apps through surveys, interviews, and qualitative content analysis. This existing research captured an overall positive assessment/evaluation of the available apps (ie, app liking), as well as users' perceptions of these apps as helpful and easy to use (eg, [8,15,50]). These findings were supported by our results, as was the importance of the diet-tracking feature to users [8] and the deterrent effect of ads [8,15]. In contrast to other studies that have incorporated user reviews to assess user perspectives on diet and nutrition apps using simple research methods to manually evaluate a smaller set of user reviews (eg, [8,50]), this study extends this knowledge, and includes over 70,000 user reviews that were evaluated and classified using text mining and NLP techniques.

In our study, we focused on the user perspective, and aimed to evaluate the diet-tracking apps and their features that are most frequently commented on by users in app reviews. Although users rated the apps they use very highly on average (the overall rating for all apps was 4.4 out of 5, with individual app ratings ranging from 4.1 to 4.7), some features could still be improved to enhance the user experience.

The predominant positivity in comments and reviews left by users online has already been noted in a recent study including 25 online platforms [51]. We can also support these findings for diet-tracking app reviews, as our results suggest that the users leaving reviews for diet-tracking apps generally tend to give positive ratings. The extracted topics show the prevalence of positive words used in reviews to describe apps, as well as the average ratings of user reviews. To provide diet-tracking app users with an even better experience, we recommend paying more attention to the food databases and the features for adding new foods to the app's database. Users are generally satisfied with the apps' functionality, while a richer database would make it easier for them to make choices, and they would likely feel more motivated than if they had to search for substitutes or go through long procedures to add a product. This is especially valid for apps that do not offer the option of adding products by taking a picture. As this is also one of the features associated with several complaints about technical issues (ie, difficulty in adding new, personalized food items or food scanning), working on advanced technological solutions, enriching the database, or working on a stronger network of products within food categories could reduce the negative evaluation of user experience.

Hidden costs and inadequate communication about the cost of using the app are some of the main reasons reported for users to stop using an app [8]. We also identified these issues as a source of numerous complaints from users leaving app reviews, along with the users' inability to get their money refunded. Therefore, to retain users and keep them satisfied and loyal, an app should provide adequate, clear, and respectful communication regarding the costs that may be incurred through app use (eg, paid premium versions, in-app purchases, refund policies).

Limitations

Although this study provides valuable insight into user opinions, it is not without limitations. Owing to feasibility constraints, we focused on available reviews and introduced a set of constraints that allowed us to structure and summarize the otherwise diverse user-generated content in the form of app reviews. Future research could apply other text-mining approaches for data collection, cleaning, and analysis. In performing similar studies, it may be beneficial to differentiate users and their motivations for using the diet-tracking app. This can be done (to a certain extent) by a deeper investigation of the review content and its sentiment.

The use of additional methods (eg, surveys, focus groups, or interviews) would be necessary to include and understand the opinions of users who do not leave feedback in the form of a review and to generalize the findings to the entire population of diet-tracking app users.

In addition, users from different cultures may have different app needs (eg, product availability, serving size differences, religious and other food restrictions). To ensure the generalizability and applicability of such findings to a specific market, the results should also include analysis of additional

apps and reviews (both global and local apps) in the local language.

This study focused on apps that offer their users the ability to count their calories and track their diets (ie, diet-tracking apps). Although these features are present in apps that are widely applicable (ie, nutrition apps), the results obtained in this study cannot be generalized to the entire segment of nutrition apps. The inclusion of additional selection criteria and apps would be necessary to claim broader applicability of the results. Similarly, reviewers (consumers who write reviews) have been shown to differ from other customers in terms of income, education, and purchasing behavior [52]. Since only a small proportion of users provide openly worded text, an analyst should be aware of nonresponse bias and how a reviewer's sentiments can be directly correlated with the sentiments of previous reviews, which affects the corpus in which posts are public [53]. Hence, our results cannot be generalized to the entire population of app users.

In addition, only apps that had the highest download numbers in the market were selected for this study. This selection was made due to their greater impact and ability to influence more individuals. Moreover, these apps make frequent updates to provide better service to their users, support the growing network of users, and avoid technical issues that are usually the subject of user complaints, including those in reviews. Because of these efforts, the apps used in this study were also quite homogenous in terms of their ratings (all 15 apps were rated above 4.1 out of 5). Including a broader range of apps (both in terms of greater variation in ratings and in terms of number of users and downloads) may reveal additional challenges users face when using diet-tracking apps.

Conclusions

Assessment of 72,084 user reviews for diet-tracking apps revealed an overall positive user evaluation. Users highly value the ability to track their food intake and manage their weight. Nonetheless, there is significant room for improvement, particularly in the area of charges associated with app use and features that enable adding food to the apps' databases. The findings of this study provide relevant insights into user opinions and evaluations of diet-tracking apps.

The implications of this study go beyond those for app developers as stakeholders; for example, in cases concerning health and nutrition, public policy and official institutions should be involved. Digital participation of current and future generations is increasing; there is also evidence that mobile apps are a potentially useful tool for shaping and tracking users' diets [8,14]. By exploring users' experiences with apps, along with their suggestions and comments, it is possible to better support the apps they need and improve their eating habits, health and diet management, and nutrition-related well-being.

Acknowledgments

The authors acknowledge financial support from the Slovenian Research Agency (research program P5-0128 and research project N5-0084).

Conflicts of Interest

None declared.

References

1. Popkin BM, Duffey K, Gordon-Larsen P. Environmental influences on food choice, physical activity and energy balance. Physiol Behav 2005 Dec 15;86(5):603-613. [doi: 10.1016/j.physbeh.2005.08.051] [Medline: 16246381]
2. Swinburn BA, Sacks G, Hall KD, McPherson K, Finegood DT, Moodie ML, et al. The global obesity pandemic: shaped by global drivers and local environments. Lancet 2011 Aug 27;378(9793):804-814. [doi: 10.1016/S0140-6736(11)60813-1] [Medline: 21872749]
3. Withrow DA, Alter D. The economic burden of obesity worldwide: a systematic review of the direct costs of obesity. Obes Rev 2011 Feb;12(2):131-141. [doi: 10.1111/j.1467-789X.2009.00712.x] [Medline: 20122135]
4. Lake A, Townshend T. Obesogenic environments: exploring the built and food environments. J R Soc Promot Health 2006 Nov;126(6):262-267. [doi: 10.1177/1466424006070487] [Medline: 17152319]
5. Kirk SFL, Penney TL, McHugh TLF. Characterizing the obesogenic environment: the state of the evidence with directions for future research. Obes Rev 2010 Feb;11(2):109-117. [doi: 10.1111/j.1467-789X.2009.00611.x] [Medline: 19493302]
6. Kervenoael R, Schwob A, Hasan R, Ting YS. Consumers' perceived value of healthier eating: A SEM analysis of the internalisation of dietary norms considering perceived usefulness, subjective norms, and intrinsic motivations in Singapore. J Consumer Behav 2020 Oct 08:1-14. [doi: 10.1002/cb.1884]
7. Frie K, Hartmann-Boyce J, Jebb S, Albury C, Nourse R, Aveyard P. Insights from Google Play Store user reviews for the development of weight loss apps: mixed-method analysis. JMIR Mhealth Uhealth 2017 Dec 22;5(12):e203 [FREE Full text] [doi: 10.2196/mhealth.8791] [Medline: 29273575]
8. Krebs P, Duncan DT. Health app use among US mobile phone owners: a national survey. JMIR Mhealth Uhealth 2015 Nov 04;3(4):e101 [FREE Full text] [doi: 10.2196/mhealth.4924] [Medline: 26537656]

9. Ni Mhurchu C, Eyles H, Jiang Y, Blakely T. Do nutrition labels influence healthier food choices? Analysis of label viewing behaviour and subsequent food purchases in a labelling intervention trial. Appetite 2018 Feb 01;121:360-365 [FREE Full text] [doi: 10.1016/j.appet.2017.11.105] [Medline: 29191745]
10. De Cock N, Vangeel J, Lachat C, Beullens K, Vervoort L, Goossens L, et al. Use of fitness and nutrition apps: associations with body mass index, snacking, and drinking habits in adolescents. JMIR Mhealth Uhealth 2017 Apr 25;5(4):e58 [FREE Full text] [doi: 10.2196/mhealth.6005] [Medline: 28442455]
11. Pagoto S, Schneider K, Jojic M, DeBiasse M, Mann D. Evidence-based strategies in weight-loss mobile apps. Am J Prev Med 2013 Nov;45(5):576-582. [doi: 10.1016/j.amepre.2013.04.025] [Medline: 24139770]
12. Webb TL, Joseph J, Yardley L, Michie S. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. J Med Internet Res 2010 Feb 17;12(1):e4 [FREE Full text] [doi: 10.2196/jmir.1376] [Medline: 20164043]
13. Flaherty SJ, McCarthy M, Collins A, McAuliffe F. Can existing mobile apps support healthier food purchasing behaviour? Content analysis of nutrition content, behaviour change theory and user quality integration. Public Health Nutr 2017 Oct 30;21(2):288-298. [doi: 10.1017/s1368980017002889]
14. Villinger K, Wahl DR, Boeing H, Schupp HT, Renner B. The effectiveness of app-based mobile interventions on nutrition behaviours and nutrition-related health outcomes: A systematic review and meta-analysis. Obes Rev 2019 Oct;20(10):1465-1484 [FREE Full text] [doi: 10.1111/obr.12903] [Medline: 31353783]
15. West JH, Belvedere LM, Andreasen R, Frandsen C, Hall PC, Crookston BT. Controlling your "app"etite: how diet and nutrition-related mobile apps lead to behavior change. JMIR Mhealth Uhealth 2017 Jul 10;5(7):e95 [FREE Full text] [doi: 10.2196/mhealth.7410] [Medline: 28694241]
16. Patel ML, Hopkins CM, Brooks TL, Bennett GG. Comparing self-monitoring strategies for weight loss in a smartphone app: randomized controlled trial. JMIR Mhealth Uhealth 2019 Feb 28;7(2):e12209 [FREE Full text] [doi: 10.2196/12209] [Medline: 30816851]
17. Hartmann-Boyce J, Johns DJ, Jebb SA, Aveyard P, Behavioural Weight Management Review Group. Effect of behavioural techniques and delivery mode on effectiveness of weight management: systematic review, meta-analysis and meta-regression. Obes Rev 2014 Jul;15(7):598-609. [doi: 10.1111/obr.12165] [Medline: 24636238]
18. Hu EA, Nguyen V, Langheier J, Shurney D. Weight reduction through a digital nutrition and food purchasing platform among users with obesity: longitudinal study. J Med Internet Res 2020 Sep 02;22(9):e19634 [FREE Full text] [doi: 10.2196/19634] [Medline: 32792332]
19. Flores Mateo G, Granado-Font E, Ferré-Grau C, Montaña-Carreras X. Mobile phone apps to promote weight loss and increase physical activity: a systematic review and meta-analysis. J Med Internet Res 2015 Nov 10;17(11):e253 [FREE Full text] [doi: 10.2196/jmir.4836] [Medline: 26554314]
20. Solbrig L, Jones R, Kavanagh D, May J, Parkin T, Andrade J. People trying to lose weight dislike calorie counting apps and want motivational support to help them achieve their goals. Internet Interv 2017 Mar;7:23-31 [FREE Full text] [doi: 10.1016/j.invent.2016.12.003] [Medline: 28286739]
21. Armstrong MJ, Mottershead TA, Ronksley PE, Sigal RJ, Campbell TS, Hemmelgarn BR. Motivational interviewing to improve weight loss in overweight and/or obese patients: a systematic review and meta-analysis of randomized controlled trials. Obes Rev 2011 Sep;12(9):709-723. [doi: 10.1111/j.1467-789X.2011.00892.x] [Medline: 21692966]
22. Genc-Nayebi N, Abran A. A systematic literature review: Opinion mining studies from mobile app store user reviews. J Syst Softw 2017 Mar;125:207-219. [doi: 10.1016/j.jss.2016.11.027]
23. Iacob C, Harrison R. Retrieving and analyzing mobile apps feature requests from online reviews. : IEEE; 2013 May Presented at: 10th Working Conference on Mining Software Repositories (MSR); May 2013; San Francisco p. 207-219. [doi: 10.1109/msr.2013.6624001]
24. Boudreaux ED, Waring ME, Hayes RB, Sadasivam RS, Mullen S, Pagoto S. Evaluating and selecting mobile health apps: strategies for healthcare providers and healthcare organizations. Transl Behav Med 2014 Dec;4(4):363-371 [FREE Full text] [doi: 10.1007/s13142-014-0293-9] [Medline: 25584085]
25. Pagano D, Maalej W. User feedback in the appstore: An empirical study. 2013 Jul Presented at: 21st IEEE International Requirements Engineering Conference (RE); 2013; Rio De Janeiro. [doi: 10.1109/re.2013.6636712]
26. Berger J, Humphreys A, Ludwig S, Moe WW, Netzer O, Schweidel DA. Uniting the tribes: using text for marketing insight. J Market 2019 Aug 29;84(1):1-25. [doi: 10.1177/0022242919873106]
27. Lee HE, Cho J. What motivates users to continue using diet and fitness apps? application of the uses and gratifications approach. Health Commun 2017 Dec;32(12):1445-1453. [doi: 10.1080/10410236.2016.1167998] [Medline: 27356103]
28. Bardus M, van Beurden SB, Smith JR, Abraham C. A review and content analysis of engagement, functionality, aesthetics, information quality, and change techniques in the most popular commercial apps for weight management. Int J Behav Nutr Phys Act 2016 Mar 10;13:35 [FREE Full text] [doi: 10.1186/s12966-016-0359-9] [Medline: 26964880]
29. Ferrara G, Kim J, Lin S, Hua J, Seto E. A focused review of smartphone diet-tracking apps: usability, functionality, coherence with behavior change theory, and comparative validity of nutrient intake and energy estimates. JMIR Mhealth Uhealth 2019 May 17;7(5):e9232 [FREE Full text] [doi: 10.2196/mhealth.9232] [Medline: 31102369]

30. Franco RZ, Fallaize R, Lovegrove JA, Hwang F. Popular nutrition-related mobile apps: a feature assessment. JMIR Mhealth Uhealth 2016 Aug 01;4(3):e85 [FREE Full text] [doi: 10.2196/mhealth.5846] [Medline: 27480144]
31. West JH, Hall PC, Hanson CL, Barnes MD, Giraud-Carrier C, Barrett J. There's an app for that: content analysis of paid health and fitness apps. J Med Internet Res 2012 May 14;14(3):e72 [FREE Full text] [doi: 10.2196/jmir.1977] [Medline: 22584372]
32. Li Y, Ding J, Wang Y, Tang C, Zhang P. Nutrition-related mobile apps in the China App Store: assessment of functionality and quality. JMIR Mhealth Uhealth 2019 Jul 30;7(7):e13261 [FREE Full text] [doi: 10.2196/13261] [Medline: 31364606]
33. König LM, Sproesser G, Schupp HT, Renner B. Describing the process of adopting nutrition and fitness apps: behavior stage model approach. JMIR Mhealth Uhealth 2018 Mar 13;6(3):e55 [FREE Full text] [doi: 10.2196/mhealth.8261] [Medline: 29535078]
34. Chen J, Cade JE, Allman-Farinelli M. The most popular smartphone apps for weight loss: a quality assessment. JMIR Mhealth Uhealth 2015 Dec 16;3(4):e104 [FREE Full text] [doi: 10.2196/mhealth.4334] [Medline: 26678569]
35. Fallaize R, Zenun Franco R, Pasang J, Hwang F, Lovegrove JA. Popular nutrition-related mobile apps: an agreement assessment against a UK reference method. JMIR Mhealth Uhealth 2019 Feb 20;7(2):e9838 [FREE Full text] [doi: 10.2196/mhealth.9838] [Medline: 30785409]
36. Kwartler T. Text mining in practice with R. Hoboken, NJ: John Wiley & Sons; 2017.
37. Tirunillai S, Tellis GJ. Mining marketing meaning from online chatter: strategic brand analysis of big data using latent Dirichlet allocation. J Market Res 2014 Aug;51(4):463-479. [doi: 10.1509/jmr.12.0106]
38. O'Connor B, Bamman D, Smith NA. Computational text analysis for social science: Model assumptions and complexity. 2011 Dec Presented at: Second Workshop on Comptuational Social Science and the Wisdom of Crowds (NIPS); 2011; Sierra, Nevada.
39. Bao Y, Datta A. Simultaneously discovering and quantifying risk types from textual risk disclosures. Manag Sci 2014 Jun;60(6):1371-1391. [doi: 10.1287/mnsc.2014.1930]
40. Gharatkar S, Ingle A, Naik T, Save A. Review preprocessing using data cleaning and stemming technique. 2017 Mar Presented at: International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS); 2017; Coimbatore p. 2017. [doi: 10.1109/iciiecs.2017.8276011]
41. Malley B, Ramazzotti D, Wu JT. Data pre-processing. In: MIT Critical Data , editor. Secondary Analysis of Electronic Health Records. Cham: Springer; 2016:115-141.
42. Plisson J, Lavrac N, Mladenic D. A rule based approach to word lemmatization. 2004 May Presented at: SIAM Conference on Imaging Science (IS04); May 2004; Salt Lake City, UT p. 1-4. [doi: 10.1093/gmo/9781561592630.article.41666]
43. Toutanova K, Klein D, Manning CD, Singer Y. Feature-rich part-of-speech tagging with a cyclic dependency network. 2003 Jun Presented at: Human Language Technology Conference (HLT-NAACL) 2003; May 27-June 1, 2003; Edmonton, AB p. 173-180. [doi: 10.3115/1073445.1073478]
44. Blei DM, Ng AY, Jordan MI. Latent dirichlet allocation. J Machine Learn Res 2003;3:993-1022 [FREE Full text]
45. Schonlau M, Guenther N. Text mining using n-grams. SSRN J 2017 Dec;17(4):866-881. [doi: 10.2139/ssrn.2759033]
46. Wang W, Feng Y, Dai W. Topic analysis of online reviews for two competitive products using latent Dirichlet allocation. Electronic Commer R A 2018 May;29:142-156. [doi: 10.1016/j.elerap.2018.04.003]
47. Pang B, Lee L. Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales. 2005 Jun Presented at: 43rd Annual Meeting of the Association for Computational Linguistics; 2005; Ann Arbor, MI p. 115-124. [doi: 10.3115/1219840.1219855]
48. Curiskis SA, Drake B, Osborn TR, Kennedy PJ. An evaluation of document clustering and topic modelling in two online social networks: Twitter and Reddit. Inf Process Manag 2020 Mar;57(2):102034. [doi: 10.1016/j.ipm.2019.04.002]
49. Mimno D, Wallach H, Talley E, Leenders M, McCallum A. Optimizing semantic coherence in topic models. 2011 Jul Presented at: Conference on Empirical Methods in Natural Language Processing; July 2011; Edinburgh, UK p. 262-272.
50. Zaidan S, Roehrer E. Popular mobile phone apps for diet and weight loss: a content analysis. JMIR Mhealth Uhealth 2016 Jul 11;4(3):e80 [FREE Full text] [doi: 10.2196/mhealth.5406] [Medline: 27400806]
51. Schoenmueller V, Netzer O, Stahl F. The polarity of online reviews: prevalence, drivers and implications. J Market Res 2020 Jul 30;57(5):853-877. [doi: 10.1177/0022243720941832]
52. Anderson ET, Simester DI. Reviews without a purchase: low ratings, loyal customers, and deception. J Market Res 2014 Jun;51(3):249-269. [doi: 10.1509/jmr.13.0209]
53. Chapman C. Commentary: Mind your text in marketing practice. J Market 2019 Nov 11;84(1):26-31. [doi: 10.1177/0022242919886882]

Abbreviations

NLP: natural language processing

Edited by R Kukafka; submitted 20.10.20; peer-reviewed by L Tully, W Van Lippevelde; comments to author 19.11.20; revised version received 06.01.21; accepted 16.03.21; published 22.04.21.

Please cite as:
Zečević M, Mijatović D, Kos Koklič M, Žabkar V, Gidaković P
User Perspectives of Diet-Tracking Apps: Reviews Content Analysis and Topic Modeling
J Med Internet Res 2021;23(4):e25160
URL: https://www.jmir.org/2021/4/e25160
doi:10.2196/25160
PMID:33885375

Original Paper

Patients' and Clinicians' Visions of a Future Internet-of-Things System to Support Asthma Self-Management: Mixed Methods Study

Chi Yan Hui[1], BEng, MSc, PhD, CEng; Brian McKinstry[1], MD; Olivia Fulton[1], BN; Mark Buchner[2], BSc; Hilary Pinnock[1], MD

[1]Asthma UK Centre for Applied Research, Usher Institute, University of Edinburgh, Edinburgh, United Kingdom

[2]Tactuum Ltd, Glasgow, United Kingdom

Corresponding Author:
Hilary Pinnock, MD
Asthma UK Centre for Applied Research
Usher Institute
University of Edinburgh
Old Medical School, Teviot Place
Edinburgh
United Kingdom
Phone: 44 0131 650 8102
Email: hilary.pinnock@ed.ac.uk

Abstract

Background: Supported self-management for asthma reduces acute attacks and improves control. The internet of things could connect patients to health care providers, community services, and their living environments to provide overarching support for self-management.

Objective: We aimed to identify patients' and clinicians' preferences for a future internet-of-things system and explore their visions of its potential to support holistic self-management.

Methods: In an exploratory sequential mixed methods study, we recruited patients from volunteer databases and charities' social media. We purposively sampled participants to interview them about their vision of the design and utility of the internet of things as a future strategy for supporting self-management. Respondents who were not invited to participate in the interviews were invited to complete a web-based questionnaire to prioritize the features suggested by the interviewees. Clinicians were recruited from professional networks. Interviews were transcribed and analyzed thematically using PRISMS self-management taxonomy.

Results: We interviewed 12 patients and 12 clinicians in the United Kingdom, and 140 patients completed the web-based questionnaires. Patients expressed mostly wanting a system to log their asthma control status automatically; provide real-time advice to help them learn about their asthma, identify and avoid triggers, and adjust their treatment. Peak flow (33/140, 23.6%), environmental (pollen, humidity, air temperature) (33/140, 23.6%), and asthma symptoms (25/140, 17.9%) were the specific data types that patient most wanted. Information about asthma and text or email access to clinical advice provided a feeling of safety for patients. Clinicians wanted automated objective data about the patients' condition that they could access during consultations. The potential reduction in face-to-face consultations was appreciated by clinicians which they perceived could potentially save patients' travel time and health service resources. Lifestyle logs of fitness regimes or weight control were valued by some patients but were of less interest to clinicians.

Conclusions: An automated internet-of-things system that requires minimal input from the user and provides timely advice in line with an asthma action plan agreed by the patient with their clinician was preferred by most respondents. Links to asthma information and the ability to connect with clinicians by text or email were perceived by patients as features that would provide a sense of safety. Further studies are needed to evaluate the usability and effectiveness of internet-of-things systems in routine clinical practice.

(J Med Internet Res 2021;23(4):e22432) doi:10.2196/22432

KEYWORDS

asthma; supported self-management; telehealth; mobile application; internet-of-things

Introduction

Asthma is a chronic disease, affecting 235 million people worldwide [1]. Supported self-management reduces emergency use of health care resources, improves asthma outcomes, and reduces morbidity [2-4]. The PRISMS (Practical Systematic Review of Self-Management Support for Long-term Conditions [5]) taxonomy consists of a 14-item list of strategies that have been used to support self-management in long-term conditions and which readers can select according to applicability to their context. In the context of asthma, these strategies include information about the condition, an action plan agreed upon between the patient and their clinician, self-monitoring of asthma status with feedback, recording of physiological measures, use of equipment, lifestyle, and social support. Technology can help support self-management [6], and many patients are interested in a broad range of self-management support strategies, with a seamless link to their clinician if needed [7]. The internet of things (IoT)[8] and wireless networks such as The Things Network (TTN; [9]), long-range wide area networks (LoRaWAN), Wi-Fi, and mobile networks, enable asthma status to be logged automatically by smart devices (eg, smart inhalers, smart peak flow meters, smart watches). Long-life invisible environmental sensors with embedded intelligence supported by Raspberry Pi [10] and Arduino [11] can measure indoor and outdoor environmental triggers which can be correlated to asthma logs that alert patients to changes status, give real-time advice on self-management, and can share data with health care advisors if necessary.

The IoT is a giant system comprising networks linking web-based services and clouds with online sensors and actuators. Sensors are able to communicate with each other to make distributed intelligent decisions in real time [12-15] An app or webpage can be provided to give users a *door* to interact with the system (eg, to view data dashboards; to receive advice; to receive reminders to collect data, reorder medication, or make appointments with their clinician). The IoT health care network has defined an IoT topology [16], which is an architecture and platform that can be used for diagnosis, personalized medication, emergency service, home rehabilitation, remote surgery remote monitoring, and self-management of conditions such as diabetes [17-20], hypertension [18,21], and asthma [18,22,23]. The implementation of 5G IoT will increase data transmission speed and reduce the transmission latency, allowing a faster and seamless service [16]. IoT interconnectivity between patients and their health care providers, the community services, and their living environment can be used to provide overarching and personalized patient self-management support.

There are many asthma-related smart devices in the market (eg, smart inhalers [24], smart spirometers and peak flow meters [25], respiratory rate sensors, wearable sensors that detect wheezing and sleeping patterns [26-28], and digital fraction of exhaled nitric oxide meters [29]), some of which have laboratory-proven accuracy [24,29,30]. Few, however, are able to be personalized to a patient's clinical profile, social preferences, and environmental context or integrated with a patient's electronic health records. Mobile systems that connect asthma smart devices and pull data from electronic health records [22,23] and apps that support asthma self-management [31] have been developed, but often, technology researchers focus on novelty and ensuring acceptability of their technology [32], rather than exploring the breadth of functionality that could support patients' everyday life of living with a condition and meet the demands of clinicians providing routine clinical care [33]. In contrast, clinical research typically focuses on evaluating older, established digital health technologies and their impact on patient health outcomes [34,35]. To our knowledge, there is no research that explores which IoT features are desired by patients and clinicians in the context of asthma self-management. We, therefore, aimed to explore the perspective of patients and clinicians on which self-management features (as defined by the PRISMS taxonomy) they would want in a future IoT system.

Methods

Ethical Approval

This mixed methods study was conducted between May 2019 and January 2020, with the approval of the National Health Service (NHS) London Fulham Research Ethics committee (ref 19/LO/0703). The study was sponsored by the University of Edinburgh and the NHS Lothian (Academic and Clinical Central Office for Research and Development). All participants were fully informed about the study and provided their consent.

Design

We used qualitative interviews and a web-based questionnaire to explore patients' and professionals' vision for future IoT features. We adopted an exploratory sequential design [36], using qualitative interviews with purposively selected patients and clinicians to identify the preferred IoT features, which in turn were used to inform a web-based questionnaire in a wider asthma patient community enabling triangulation of interview findings.

Patient and Clinician Recruitment

Patient Recruitment

We included UK-based adult patients who were actively using treatment for asthma [37] and excluded anyone who needed carers' support to manage their asthma. We recruited patients from volunteer databases and asthma charities social media. We identified patients in the Scottish Health Research Register (SHARE) [38], Register for Asthma Research (REACH), and Asthma UK volunteer database [39]. People registered in these databases have given consent to be contacted about asthma research. Eligible patients were invited with emails sent by SHARE, REACH, and Asthma UK, which included a recruitment link on behalf of the research team. We also posted advertisements on Asthma UK and Asthma UK Centre for Applied Research (AUKCAR) Facebook and Twitter that included a recruitment link. People who were interested in the

study used the link to register their interest. They were asked to read the information leaflet, confirm their eligibility, provide basic demographics, and give consent to be contacted (via the contact details they provided) in order complete the registration.

Sampling for Qualitative Interviews

We purposively recruited a maximum variation sample of 12 patients to take part in individual interviews. Sampling was based on age (16-25 years, 26-45 years, 16-65 years, 65 years or older); ownership of an action plan (or not); duration of asthma (diagnosed <6 months, 6-12 months, 1-10 years, >10 years); hospital admission in the previous 12 months (or not); confidence with using technology (ability to download apps by themselves, need help, never tried).

Clinician Recruitment

We recruited health care professionals from primary and secondary care through newsletters and social media (the NHS Research Scotland Primary Care Network and professional bodies such as the Primary Care Respiratory Society and the NHS Lothian Respiratory Managed Clinical Network). We also approached individual professionals who were actively involved in clinical care or research through personal networks.

Data Collection

Think-Aloud Qualitative Interviews

We adopted a think-aloud approach [40] to explore patients' and professionals' preferences on the future design and utility of IoT systems to support asthma self-management.

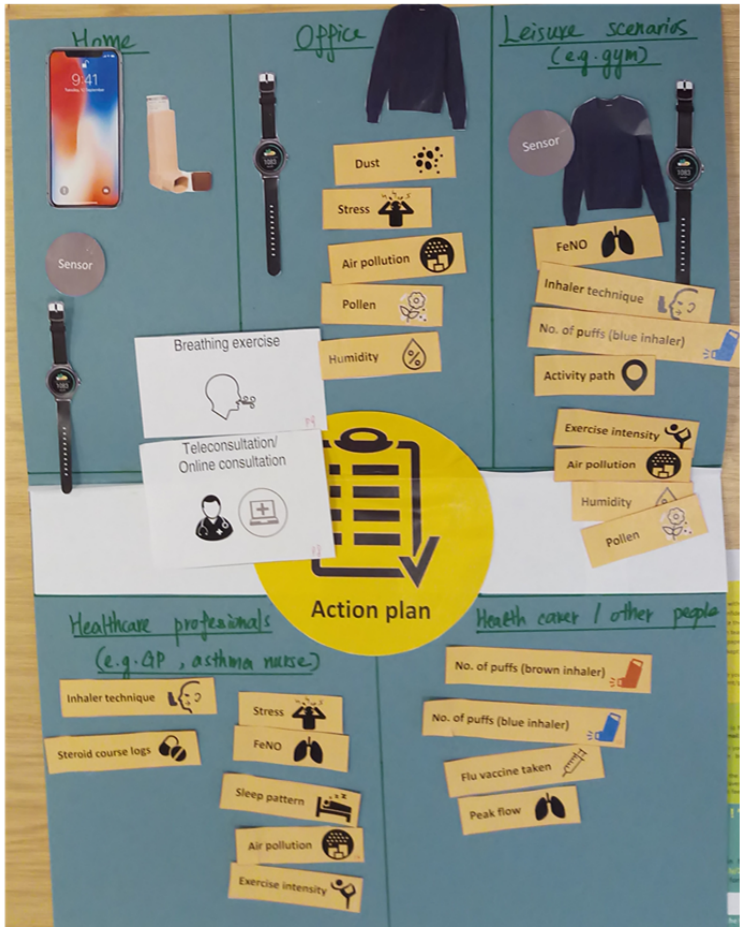
We asked patients to complete 2 tasks in the interview. For task 1, we provided a list of the most wanted app features (Multimedia Appendix 1) identified in our previous study [7] and asked patients to think about their previous usage of asthma apps and their previous asthma self-management experience to decide the features that they would want to be included in a future IoT system, and how often they think they would use them. For task 2, we provided images of current and future potential smart devices in the market (smart inhaler, smart peak flow meter, smart fabric, smart watch, voice assistant, smart purifier, human robot, and robot pet) and available data and asked patients to discuss their future potential. Specifically we asked them to use the images to create a IoT system that they would want to support their daily self-management, and if, how, and when they would use each device for self-monitoring and what data (if any) they would want to be able to send to their health care professionals and their health carers, for example, parents or spouse. (Figure 1 is an example of a task completed by a participant.)

At the end of each task, participants were asked to add any features, data, or smart devices that they thought would be useful but that were not included in the list or images (or were not yet available).

Professionals were asked to think about their previous experiences providing care for people with asthma and select the features from the list [7] that could support self-management and data that they would like to receive in order to assist their consultations with patients.

Finally, we showed participants a prototype app (an interface app which included many wanted features [7]) to stimulate their thoughts on how they would like to interact with a self-management support IoT system (Multimedia Appendix 2).

Figure 1. An example of a task completed by a participant.



Development of the Web-Based Questionnaire

The web-based questionnaire aimed to quantify the features wanted by interviewees and their perceived usefulness enabling us to triangulate the findings. Participants' feedback in the qualitative interview informed the features listed.

The questionnaire (Multimedia Appendix 3) collected basic demographic information (age group and gender) and asked questions regarding what data that they would want an IoT system to collect and where they would want the data to be collected. The flow of questions reflected the sequence that we used in the interviews. To reduce instances of missing data, each question had to be completed before progressing to the next. First, we offered a list of features based on our previous research [7] and the interview responses. Survey participants were asked to select the top 5 types of data that they would like collected by the IoT system and to prioritize them. A free-text option was provided. Second, we asked participants to choose where it would be most useful for the technology to collect these data. The choices were informed by the interviews—"home," "at work/office/school," "the place where they are at leisure activities that they do regularly (eg, gym, running, etc)," and "others." If participants chose the leisure option, we asked the participants to specify which leisure scenario they thought was relevant to that data type. We offered a variety of activities informed by the interviews and a free text option.

Administration of the Web-Based Questionnaire

We used Bristol Online Survey, a secure web-based survey platform that complies with ISO27001 information security standards and the General Data Protection Regulation, to build the web-based questionnaire and collect data. To test content validity and readability, we invited a patient volunteer and an independent researcher to try out the questionnaire independently. Their feedback was incorporated in the questionnaire before it was sent to participants. Potential participants who had not been selected for an interview and who had given permission to be contacted regarding the survey were emailed a participant information sheet and a link to the questionnaire.

Data Analysis and Synthesis

Qualitative Analysis

Interviews were digitally recorded, transcribed, and coded in NVivo (version 12; QSR International). We used PRISMS taxonomy [3] as a framework to categorize the IoT features (themes) that emerged from the interviews. In addition, our thematic analysis explored issues of particular importance to the participants. One reviewer (CYH) coded 2 interviews (patient, professional) independently, and another reviewer (HP) reviewed the coded transcriptions to standardize the coding, which was then applied (by CYH) to all the transcriptions. CYH (reviewed by HP) coded features suggested by patients and professionals separately and extracted the features, which were then combined in tables for comparison. We categorized the

individual assessments of the features as would "always use," "often use," "use when needed," or "less likely to use" according to the words used in the interviews (Table 1). In our previous study [7] on self-management features wanted by patients' and clinicians', no new app features were generated after 15 interviews. We estimated that a sample size of 12 would be likely to achieve data saturation and was within our resources for conducting individual interviews.

Table 1. Defining patients' perceptions of potential future usage.

Usage	Examples of words used by patient interviewees
Always use	"always use," "would always use," "very helpful," "would be good," "more important," "definitely a must"
Often use	"would often use," "would be good," "really/extremely useful," "it might be useful," "something that I'd probably use often, if I could, or if I can"; "once a month," "quite/might be/ useful," "might interest me"
Use it when needed	"safety net," "absolutely need it," but like "rare occurrences," "once a year," "very important just to remind yours occasionally," "it wouldn't be like every day," "a bit of a waste of time if you have that popping up every day," "check in on regularly"
Less likely to use	Unlikely to use because "I know what to do" "the factors do not affect me", "the feature is already provided by the current practice", "no need to have this feature as an extra"; or they "hope not to use this feature very often" or "would not rely on it"; or they "haven't used it before".
Not applicable	"Does not apply to them"; "no point in it"

Quantitative Analysis and Triangulation of Findings

We used descriptive statistics to analyze patients' preference ratings for the data they would like a future IoT system to collect and in which circumstances the system would be most useful. We used data from the survey to quantify the features wanted by interviewees and their perceived usefulness.

Interpretation

The findings, data synthesis, and interpretation were discussed regularly within the multidisciplinary study team, which included a patient representative, technology developer, health care professionals, and researcher. The researcher had an engineering and technology background and clinical research experience with patients and clinicians in asthma app development.

Results

Participants

Patients

Invitations were sent by email to 572 patients (SHARE n=211; Asthma UK n=220; REACH n=141). Advertisements were posted to the Asthma UK's Facebook and Twitter (185,800 followers) and AUKCAR Twitter (1224 followers). Of 362 patients who expressed interest in the study, 297 (82.0%) were from Asthma UK's social media, email, or website; 42 (11.6%) were from the AUKCAR Twitter; 10 (2.8%) were from SHARE; 3 (0.8%) were from REACH, 2 (0.6%) were recommended by their hospital or a relative, 8 (2.2%) received an e-mail invitation from Asthma UK and others responded to our advertising. Of the 362 patients, 12 were selected for an interview, and the remaining 350 were invited to complete the web-based questionnaire: 3 rejected the invitation (2 were no longer available to take part and 1 declined when they realized there was no payment), and 29 were undeliverable emails. Thus, 318 were sent the web-based questionnaire link, of whom 140 (44.0%) completed the questionnaire. Of the 140 participants, 139 participants (99.3%) completed the questionnaire within 15 minutes, and 1 person (0.7%) appeared to take 3 hours and 42 minutes to complete the questionnaire (most likely because the *thank you* page was not closed after completion). There were no instances of missing data.

Of 152 participants (12 interviewees and 140 web-based questionnaire respondents), most (74/152, 49%) were 46 years to 65 years of age, 69.1% (105/152) were female, 54.6% (83/152) had an action plan, 75.0% (114/152) had been diagnosed with asthma for more than 10 years, 12.5% (19/152) had been admitted to hospital in the previous 12 months, and 86.2% (131/152) were confident that they could download apps themselves (Table 2).

Table 2. Characteristics of patient participants.

Patient characteristics	Interviewees (n=12), n (%)	Questionnaire respondents (n=140), n (%)
Age (years)		
16-25	3 (25.0)	2 (1.4)
26-45	2 (16.7)	39 (27.9)
46-65	3 (25.0)	71 (50.7)
>65	4 (33.3)	28 (20.0)
Gender		
Female	8 (66.7)	97 (69.3)
Male	4 (33.3)	42 (30.0)
Prefer not to say	0 (0.0)	1 (0.7)
Action plan ownership		
Yes	4 (33.3)	79 (56.4)
No	3 (25.0)	16 (11.4)
No asthma action plan but I have been told what to do	5 (41.7)	45 (32.1)
Diagnosed with asthma		
Less than 6 months	0 (0.0)	1 (0.7)
Between 6 months and 1 year	0 (0.0)	1 (0.7)
Between 1 year and 10 years	4 (33.3)	32 (22.9)
More than 10 years	8 (66.7)	106 (75.7)
Admission to the hospital because of asthma in the last 12 months		
No	8 (66.7)	125 (89.3)
Yes and now my asthma care is provided by general practitioner/asthma nurse	1 (8.3)	5 (3.6)
Yes and I am still attending the hospital (specialist) clinic	3 (25.0)	10 (7.1)
App download experience		
I download apps by myself	12 (100)	119 (85.0)
I usually ask someone to download apps for me	0 (0.0)	5 (3.6)
I have never downloaded an app	0 (0.0)	16 (11.4)

Professionals

Twelve professionals were recruited from primary, secondary, and tertiary care in the United Kingdom. All provided care for people with asthma, and some had research experience using digital technology to monitor patients' medication use and symptoms for respiratory patients (Table 3).

Table 3. Characteristics of professional participants.

Professional role	n	Experience[a]	Additional description
General practitioner	2	>8 years	Primary care: respiratory lead n=1; accident & emergency experience n=1
Asthma nurse	2	>20 years	General practice asthma-trained nurse n=2
Pharmacist	4	1-20 years	Respiratory pharmacists n=3; community pharmacist n=1 (reviewing patients during asthma admissions or when referred by general practitioner or asthma nurse; checking inhaler technique, choosing devices, addressing medication adherence)
Consultant chest physician	1	__[b]	Secondary and tertiary care (severe asthma center and community lead)
Asthma pediatrician	3	__[b]	Lead consultant in a pediatric asthma service n=1; Using smart inhaler n=1; 30 years of pediatric asthma research experience n=1 (looking after children with a range of asthma severities, conducting face-to-face consultation, determining patients' symptoms, making management plans, offering advice to general practitioners, and reviewing test results)

[a]Experience seeing patients with asthma on a regular basis.

[b]Information not available.

Mixed Methods Assessment: Features and Their Perceived Usefulness

Interview Themes

Perceptions of the 12 patients about the potential usefulness of features were mapped to the PRISMS taxonomy (Table 4; the full version can be found in Multimedia Appendix 4). Features wanted by patients reached saturation within 10 interviews, and we stopped sampling at 12 interviews.

Patients decided the usefulness of potential IoT features based on their own past asthma self-management experiences (eg, ownership or use of action plan and smart peak flow meter, their relationship with their clinicians, medication usage, use of emergency services), their asthma triggers, their curiosity about what affected their asthma, the severity and control of their asthma condition, and what they considered (or had been told by professionals) was best practice for asthma.

Monitoring, supported by feedback advice, was the feature that most patients wanted to see in an IoT system. Information about asthma and an action plan were also priorities. Flexible access to follow-up advice with a general practitioner or asthma nurse by text or email service were "safety net" features that most patients thought would create a sense of "staying connected" with clinicians, although they stated they would only use it when needed. One patient with hearing problems who struggled to communicate during an exacerbation wanted a panic button to automatically text for emergency help. Perceptions on monitoring of control, and feedback are presented in more detail below. Other features were wanted by patients but were lower priority (Multimedia Appendix 4).

Table 4. Summary of the potential usefulness of features mapped to the PRISMS taxonomy [5].

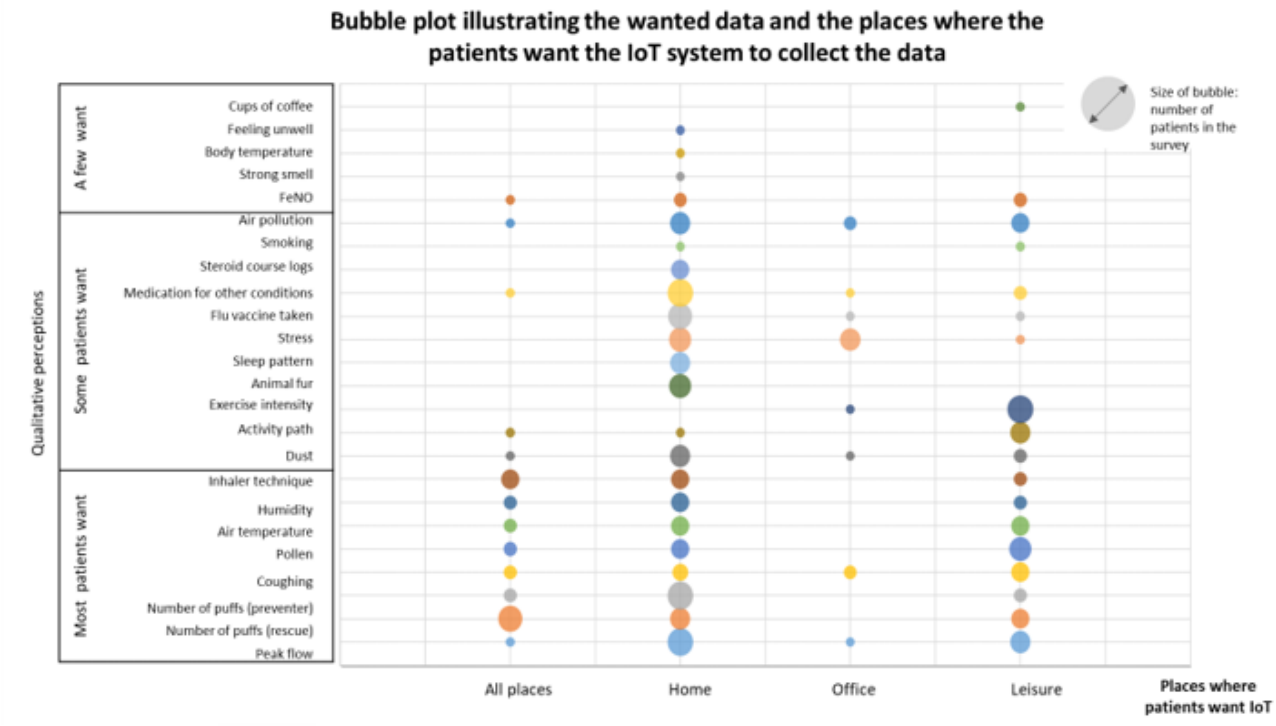
Theoretically based support	Features	Patient	Clinicians
Information about asthma and available resources	Information about asthma management	Online information was of interest, ideally personalized to clinical context and individual situation (eg, broad range of information for newly diagnosed, safety net information for experienced patients)	Professionals considered that reputable information such as inhaler technique videos, treatment information, why and how the medication should be taken were important for patient self-management
Provision of action plan	What to do when condition gets worse	Most patients wanted a (digital) action plan to remind them what to do when they forgot the agreed actions when their conditions were getting worse	Most clinicians suggested an action plan to remind about medication adjustment and agreed actions if patients' condition was getting worse
Regular clinical review	Routine review reminder; remote options.	Most patients wanted reminders for the yearly review. Some preferred web-based or teleconsultation consultation for regular reviews to save travel time	Most professionals thought reminders would encourage attendance, and agreed remote consultations were convenient though not always clinically appropriate
Monitoring condition with feedback	Logging asthma symptoms, peak flow, and medication use.	Most patients wanted automatic logging with intelligent feedback and the facility to transfer data to the hospital or general practitioner practice. Reminders could be useful though ideally only generated when their asthma was bad. Some patients wanted alerts when they had increased use of rescue inhaler	Most clinicians thought objective data would help them assess status and help patients understand their triggers. They were skeptical that reminders would improve adherence to logging. Flagging excess or increasing use of recuse medication could alert patients and professionals to poor control
Practical support with adherence	Medication reminders and support	Some patients thought reminders to take medication were useful if they were busy or forgetful but would not change opinions. Automatic prompting reordering of medication was wanted. Most patients wanted flu vaccine alerts	Clinicians wanted medication adherence logs and agreed with low-medication alerts to facilitate reordering. Warning about overordering were also important. Flu vaccination reminders should pop up in both patients and clinicians' system
Provision of equipment	Smart devices	Most patients wanted to try smart devices (inhalers, peak flows, activity trackers)	Clinicians generally interested in how the whole technology system, as opposed to how individual smart device can support patients
Provision of easy access to support when needed	Panic button for emergency	A few patients suggested this would be helpful because it could be difficult to speak during exacerbations	Clinicians generally considered this was a duplicate emergency system, though might be useful for brittle high-risk asthma
Communication with health care professionals	Emails, texts, and WhatsApp messages	Most patients wanted flexibility to ask quick follow up questions, and a patient with hearing problems found WhatsApp useful	Clinicians agreed with a flexible approach to reviews, including text services for quick follow up questions, though resources would be needed
Training for everyday activities	Air pollution/pollen high alert	Most patients wanted environmental features, and some young patients (16-25 years old) suggested it could identify triggers and help them to plan their day	Clinicians were less interested in these data, though some thought environmental information could help patients to understand (and avoid) triggers
Training for self-management activities	Incorrect inhaler technique alert	Most patients thought it was good to be prompted when their inhaler technique was incorrect	Most clinicians wanted inhaler technique checks with real-time alerts for patients recorded for discussion at a review
Training for psychological strategies	Breathing exercise	Some patients wanted breathing exercises to keep themselves calm and considered it would help their asthma	Most clinicians were keen to encourage patients to do breathing exercises to improve asthma symptoms
Social support	Social media and alerts	One patient wanted a friend alerted when she was admitted to the hospital	One clinician suggested a social media page to enable sharing of experiences
Lifestyle advice and support	Physical activity and weight loss	Some patients wanted to connect asthma logs with their activity tracker, while others would use a weight watching facility	One clinician suggested an individualized fitness program for people with severe asthma, and 2 clinicians wanted individualized weight management plans

Survey Findings and Triangulation With Interview Themes

The responses to the survey are illustrated with a bubble plot (Figure 2). In keeping with the priorities of the interviewees, the largest bubbles represent monitoring of symptoms (prioritized by 25/140, 17.9%) and peak flow measurement (33/140, 23.6%). For most respondents (24/32, 75.0%), this was a task to be undertaken at home, though some considered that symptom monitoring could be useful at home, work, or leisure

locations. Monitoring environmental asthma triggers (pollen, humidity, air temperature) was prioritized by 23.6% (33/140).

Figure 2. Information that patients want internet-of-things system features to capture. FeNO: fractional exhaled nitric oxide.



Qualitative Perceptions of Specific Features

Smart Devices for Monitoring Asthma Control

Some patients explained that a peak flow was "useful" and "needed" to help clinicians assess their asthma. Most professional interviewees agreed with this and thought that a log of peak flows, symptom scores, and medication use would engage patients and could inform assessment of control and management strategies.

> *I see particularly important is logging asthma symptoms and medication use. Peak flow, maybe less so, but the symptoms and medication use I think are very important. although peak flow is far from perfect, it just shows that the patient has engaged to some degree, that they're seeing differences, whether it's in the morning or evening or when they...after exercise and they're engaged with that, so I know when I ask them some questions about their symptom experience that they're thoughtful about it.* [Health care professional 7, general practitioner]

However, in contrast to the priority attached to logging, most patients acknowledged that in reality they checked their peak flow "rarely" or "only when their asthma was getting worse." The reasons for not measuring every day were varied. Some "forgot," while others felt "weird doing it in front of people," but many suggested it was unnecessary as they knew their asthma and could assess status by how they felt.

> *Not very often. I do sometimes, if I get chest infections or if I get tight chested or if it's not, if I just feel there's something wrong, then I test what my peak flow is, just to see what it is.* [Patient 4, 26-45 years old, male]

> *I'm a big believer in logging things so that if you've got a history... I mean to log things, but I forget, so it would be good to have something there.* [Patient 5, >65 years old, male]

Several suggestions were made about how an IoT system could help overcome this discrepancy. A word that was used frequently was "automatic." Most patients wanted an IoT system that could automatically log their asthma condition. They were interested in trying smart devices (eg, smart peak flow meter); some explicitly mentioned that this would enable them to capture data automatically though they were not clear how the devices could capture their asthma status without some effort on their part. A device with an automated data transfer feature still requires the user to blow the air into the meter, and patients acknowledged that to do this regularly would require motivation such as a request from researchers or their clinician. Patients had different opinions on using voice assistants, smart watches, smart fabrics, smart purifiers, human robots, and robot pets to collect additional data automatically:

> *Smart peak flow would be the way forward I think, it really would, because it's recording it (my peak flow). If you say to me record it every day, I would record it every day...I very rarely use it (the current mechanical peak flow). But if this smart peak flow is going to record my peak flow at that particular time each day or twice a day, that's only going to be good to help me and to help my doctor understand, you know.* [Patient 2, >65 year old, male]

> *It would be good to have something that I could just quickly tell it what's happening. Probably not when I'm having an asthma attack because I can't talk.* [Patient 1, 46-65 years old, female]

> *I don't have my inhalers with me and my other sensors in the office, so for me, given I'm going in and out of meetings and various other things and travelling about and stuff, it's probably easier if I just have the watch and any clothing or something with sensors on it that can do it, in an ideal world.* [Patient 4, 26-45 years old, male]

> *(Voice assistant) It's not very secure, I don't like that.* [Patient 5, >65 years old, male]

How the data were used was an important motivator. Most patients said it was "good" or even "essential" that data from smart peak flow meters and inhalers were transferred automatically to the system so these logs could be correlated with other data and displayed graphically on a mobile phone. Some suggested that they would log peak flows if their clinicians asked them to do so.

Clinicians were keen to see a record of medication use to enable assessment of adherence. They favored automated logging of this information via a smart inhaler as that was perceived to be more accurate.

> *It would be very useful if the patient is logging their asthma symptoms and peak flow, medication use...because that's then helping us to adjust on treatment. I would love to know if they were taking it (prescribed medication) every day like they're telling me they are.* [Health care professional 4, prescribing support pharmacist]

Feedback and Advice

Patients in the interviews suggested a broad range of information and advice that could be usefully provided through an IoT system such as a timely alert when their asthma control changed; the amount of medication to be taken according to asthma control to reduce medication side effects; the numbers of doses taken and remaining in the inhaler, ideally with an option to order a repeat prescription when the medication was running low; and correct inhaler technique.

The clinician interviewees concurred with these priorities, especially the need for detecting poor inhaler technique and linking to information about correct use of an inhaler. They also saw value in feedback supporting treatment adjustment according to an action plan.

Environmental Data

Outdoor environmental data such as pollen, humidity, and air temperature were considered "good to know" and "useful" data by many patients. Patients who were newly diagnosed with asthma wanted to learn which environmental factors affected their asthma, whereas those who already knew what triggered their asthma wanted to use daily environmental data to plan their day.

> *(The outdoor environmental data) could maybe suggest like how likely they are to be actual triggers as opposed to me just thinking. Then if it suggests that I'm really triggered by something then I could put more effort to avoid that.* [Patient 10, 16-25 years old, female]

These data were of less interest to clinicians though some suggested that, together with indoor triggers, outdoor pollution, exercise intensity, weight, peak flow, and symptoms, data could provide real-time feedback and help patients understand which factors affected their asthma in order to avoid them.

Discussion

Principal Results

Both patients and clinicians expressed their interest in automated monitoring with real-time feedback within an IoT system that could support a wide range of self-management tasks. In the qualitative interviews, patients mostly wanted the system to log peak flow, asthma symptoms, and environmental triggers (pollen, humidity, air temperature); provide advice on relevant actions or medication adjustment to suit different levels of asthma control; and provide alerts about the number of doses of medication remaining and inhaler technique. The questionnaire responses quantified these preferences with the most wanted features being monitoring peak flow (33/140, 23.6%), environmental asthma triggers (pollen, humidity, air temperature) (33/140, 23.6%), and symptoms (25/140, 17.9%). Clinicians wanted automated objective logs about patient condition that they could access during a consultation. Patients considered that easy access to information and clinical advice, such as text or email communication via the system, provided a feeling of safety, while clinicians appreciated the potential reduction in face-to-face consultations because it would reduce patient travel time and the use of health service resources. Lifestyle logs (fitness regimes or weight control) were wanted by some patents but were of less interest to clinicians.

Strengths and Limitations

Our study identified the preferred IoT features for patients and clinicians and the type of data that they wanted the system to collect; however, there are some limitations. First, due to limited resources and time, we excluded children under 16 years old from patient interviews though we included experienced pediatricians to provide insights on the needs of children with asthma and their carers. Second, we did not manage to interview any patients who were newly diagnosed with asthma (within the previous year) who may well have had specific needs for information and support. Similarly, all our participants were familiar with downloading and using apps, suggesting that our recruitment strategy of using social media reached a technologically experienced population. Although this approach will have resulted in some perspectives being overlooked, the sample included in the study were demographically diverse, with a range of experiences of living with asthma, and could provide a range of perspectives on the potential of technological support. Increasingly, the global population is becoming more familiar with digital communication. However, we reached data saturation [41,42] in the qualitative analysis, and the web-based questionnaire, which attracted a broader range of patients, provided findings that were consistent with the qualitative data.

Third, though the features that participants said that they wanted were related to participants' past experience, we also used images of emerging technologies and stimulated discussion about novel features that could potentially contribute to managing their asthma.

Reflexivity

The researchers had engineering and research experience in developing asthma apps. To reduce the influence on the interview findings of the researcher's background, the coding and interpretation of results were discussed with study team members from different backgrounds and with different experiences, including general practitioners, a patient, and a technology developer. This range of expertise enabled the study to present a balanced interpretation.

Comparison With Published Literature

IoT is an option to support self-management. Asthma action plans, asthma education, and regular consultations with clinicians are core component of effective self-management; which aims to support patients achieve good control by recognizing when their asthma is getting worse and responding promptly and appropriately [43]. The features that patients wanted from the IoT system resonated with these aims (ie, such as facilitating learning about asthma and its triggers, logging asthma status, providing alerts to highlight deterioration and offering feedback advising on treatment adjustment and other actions). Features such as monitoring how much medication is left and inhaler technique can be delivered by smart inhalers that are already on the market.

There is, however, an anomaly. Most patients wanted features that enabled regular monitoring of peak flow and symptoms, but the literature suggests that this is rarely, if ever, achieved in real life [7]. For example, in the context of chronic obstructive pulmonary disease, patients who are more severely affected by their disease (ie, those who might find monitoring most helpful) were less likely to use communication technologies such as mobile phones, text messaging, email, and video chat [44]. Our interviewees acknowledged this discrepancy by explaining that they actually only logged their asthma control when they were concerned about increasing symptoms, but that, most of the time, they did not see the necessity of monitoring because they already knew the status of their asthma. Reminders are unlikely to solve the problem of nonadherence to logging because, when patients did not complete logs, it was typically intentional and not due to forgetfulness [45]. A solution suggested by both patients and professionals was to simplify data collection by automating the process and increase motivation by providing useful graphical feedback or linking with professional advice, though even this requires some input from the patient. Further advances may require the development of systems that that silently monitor use of rescue medication with a smart inhaler that requires no input by the patient and only alerts patients in the event of unusual behavior patterns (eg, increased usage) with advice from their action plan on jhow to regain control.

Monitoring is also influenced by whether the patients' clinician is interested in the results. Home monitoring with feedback has shown promise as a self-management strategy for people with hypertension [46] and diabetes [47]. The sense of having the ongoing interest and support of a clinician is a factor in maintaining motivation to self-manage [48]. Our interviewees felt "safe" if they had an easy way (text or email) to contact clinicians when they had concerns or needed clarifications. A recent review [49] of patients with asthma and chronic obstructive pulmonary disease revealed similar preferences for accessible support. Similarly, patients with type 2 diabetes wanted technology that allowed 2-way text communication with their clinicians.[50]. The professional interviewees in our study considered that, when text and email are offered as options, they have the potential to reduce face-to-face consultations and save health care resources.

Conclusion

An IoT system can encompass the range of components needed to support asthma self-management. Patients and clinicians preferred features that monitor asthma status (preferably using automated silent monitoring), provide timely advice in line with an agreed upon asthma action plan, and allow observation of environmental factors in relation to asthma control. Our technologically literate participants appreciated the ability to connect to asthma information, as well as easy access to clinicians by text or email. Sustained use was acknowledged as a challenge. Large-scale evaluation of usability, health outcomes, and resource implications are needed to realize the potential benefits of silent monitoring connected IoT systems.

Acknowledgments

CYH received funding from (Scotland) Chief Scientist Office and Asthma UK Innovation (grant CSO-AUK-2018-03) to plan and carry out the study. BM and HP received funding from European Institute of Innovation and Technology (grant 18113). The views expressed in this publication are those of the authors and not necessarily those of the Chief Scientist Office or Asthma UK.

We are grateful to patient and professional participants for taking part in this study; Asthma UK, AUKCAR, SHARE, and REACH for help with patient recruitment; the National Health Service Research Scotland Primary Care Network, Primary Care Respiratory Society, and Lothian Respiratory Managed Clinical Network for help with professional recruitment; and AUKCAR Patient Public Involvement group members for advice on the interview materials.

Authors' Contributions

CYH and HP designed the study. CYH extracted data. HP and synthesized the data. HP is the study guarantor. CYH and HP wrote the manuscript. BM reviewed the final manuscript. OF commented on the findings from patient perspective, and MB commented on implications from technology perspective. All authors approved the final version of the manuscript.

Conflicts of Interest

MB is managing director of Tactuum Ltd. OF contributes in a lay capacity to Teva Pharmaceuticals, AstraZeneca, and WEGOHealth.

Multimedia Appendix 1
Features and screenshots.
[DOCX File , 221 KB - jmir_v23i4e22432_app1.docx]

Multimedia Appendix 2
Detailed topic guide.
[PDF File (Adobe PDF File), 179 KB - jmir_v23i4e22432_app2.pdf]

Multimedia Appendix 3
Web-based questionnaire.
[PDF File (Adobe PDF File), 110 KB - jmir_v23i4e22432_app3.pdf]

Multimedia Appendix 4
Perceived usefulness of wanted feature.
[PDF File (Adobe PDF File), 139 KB - jmir_v23i4e22432_app4.pdf]

References

1. Chronic respiratory diseases: asthma. World Health Organization. 2020. URL: https://www.who.int/news-room/q-a-detail/asthma [accessed 2020-12-26]
2. Gibson PG, Powell H, Coughlan J, Wilson AJ, Abramson M, Haywood P, et al. Self-management education and regular practitioner review for adults with asthma. Cochrane Database Syst Rev 2003(1):CD001117. [doi: 10.1002/14651858.CD001117] [Medline: 12535399]
3. Pinnock H, Parke HL, Panagioti M, Daines L, Pearce G, Epiphaniou E, PRISMS and RECURSIVE groups. Systematic meta-review of supported self-management for asthma: a healthcare perspective. BMC Med 2017 Mar 17;15(1):64 [FREE Full text] [doi: 10.1186/s12916-017-0823-7] [Medline: 28302126]
4. Levy M, Andrews R, Buckingham R. Why asthma still kills. Royal College of Physicians. URL: https://www.rcplondon.ac.uk/sites/default/files/why-asthma-still-kills-full-report.pdf [accessed 2020-12-26]
5. Pearce G, Parke HL, Pinnock H, Epiphaniou E, Bourne CLA, Sheikh A, et al. The PRISMS taxonomy of self-management support: derivation of a novel taxonomy and initial testing of its utility. J Health Serv Res Policy 2016 Apr;21(2):73-82. [doi: 10.1177/1355819615602725] [Medline: 26377727]
6. Pinnock H. Connecting professionals and patients: how technology can support asthma self-management. In: Proceedings of Respiratory Drug Delivery Europe 2017. 2017 Presented at: Respiratory Drug Delivery Europe; April 25-28; Nice, France p. 43-52.
7. Hui CY, Walton R, McKinstry B, Pinnock H. Time to change the paradigm? a mixed method study of the preferred and potential features of an asthma self-management app. Health Informatics J 2020 Jun;26(2):862-879 [FREE Full text] [doi: 10.1177/1460458219853381] [Medline: 31195877]
8. Internet of Things (IoT) ecosystem study. IEEE. 2015. URL: https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/iot_ecosystem_exec_summary.pdf [accessed 2020-12-26]
9. The Things Network. The Things Industries. URL: https://www.thethingsnetwork.org/ [accessed 2021-04-02]
10. Raspberry Pi. Raspberry Pi Foundation. URL: https://www.raspberrypi.org/ [accessed 2020-12-26]
11. Arduino. Arduino. URL: https://www.arduino.cc/ [accessed 2020-12-26]
12. Minerva R, Biru A, Rotondi D. Towards a definition of the Internet of Things (IoT). IEEE Internet Initiative. URL: https://iot.ieee.org/images/files/pdf/IEEE_IoT_Towards_Definition_Internet_of_Things_Revision1_27MAY15.pdf [accessed 2020-12-26]
13. Overview of the internet of things. International Telecommunication Union. URL: https://www.itu.int/rec/T-REC-Y.2060-201206-I [accessed 2020-12-26]
14. What is the Internet of Things? IBM Corp. URL: https://www.ibm.com/blogs/internet-of-things/what-is-the-iot/ [accessed 2020-12-26]
15. Ejaz W, Anpalagan A. Internet of Things for Smart Cities: Technologies, Big Data and Security. Cham: Springer; 2019.

16. Riazul Islam SM, Kwak D, Humaun Kabir M, Hossain M, Kwak KS. The Internet of Things for health care: a comprehensive survey. IEEE Access 2015;3:678-708. [doi: 10.1109/access.2015.2437951]
17. Saheb T, Izadi L. Paradigm of IoT big data analytics in the healthcare industry: a review of scientific literature and mapping of research trends. Telemat Inform 2019 Aug;41:70-85. [doi: 10.1016/j.tele.2019.03.005]
18. Thuemmler C, Bai C. Health 4.0: Application of Industry 4.0 Design Principles in Future Asthma Management. Cham: Springer; 2017:23-37.
19. Thilakarathne NN, Kagita MK, Gadekallu DTR. The role of the Internet of Things in health care: a systematic and comprehensive study. Int J Eng Manag Res 2020 Aug 31;10(4):145-159. [doi: 10.31033/ijemr.10.4.22]
20. Deshkar S, Thanseeh R, Menon V. A review on IoT based m-Health systems for diabetes. Int J Comput Sci Telecommun 2017;8(1):13-18 [FREE Full text]
21. Baali H, Djelouat H, Amira A, Bensaali F. Empowering technology enabled care using IoT and smart devices: a review. IEEE Sensors J 2018 Mar 1;18(5):1790-1809. [doi: 10.1109/jsen.2017.2786301]
22. MyAirCoach. The myAirCoach Consortium. 2015. URL: https://www.myaircoach.eu/content/what-myaircoach-project [accessed 2020-12-26]
23. Bui AAT, Hosseini A, Rocchio R, Jacobs N, Ross MK, Okelo S, et al. Biomedical real-time health evaluation (BREATHE): toward an mHealth informatics platform. JAMIA Open 2020 Jul;3(2):190-200 [FREE Full text] [doi: 10.1093/jamiaopen/ooaa011] [Medline: 32734159]
24. Smartinhaler for asthma. National Institute for Health and Care Excellence. 2017 Jan 11. URL: https://www.nice.org.uk/advice/mib90/chapter/The-technology [accessed 2020-12-26]
25. Carpenter DM, Jurdi R, Roberts CA, Hernandez M, Horne R, Chan A. A review of portable electronic spirometers: implications for asthma self-management. Curr Allergy Asthma Rep 2018 Aug 25;18(10):53. [doi: 10.1007/s11882-018-0809-3] [Medline: 30145683]
26. Massaroni C, Nicolo A, Sacchetti M, Schena E. Contactless methods for measuring respiratory rate: a review. IEEE Sensors J 2020:1-1 [FREE Full text] [doi: 10.1109/jsen.2020.3023486]
27. Greiwe J, Nyenhuis SM. Wearable technology and how this can be implemented into clinical practice. Curr Allergy Asthma Rep 2020 Jun 06;20(8):36 [FREE Full text] [doi: 10.1007/s11882-020-00927-3] [Medline: 32506184]
28. Tag: asthma. Wearable technology. URL: https://www.wearable-technologies.com/tag/asthma/ [accessed 2020-12-26]
29. Harnan SE, Tappenden P, Essat M, Gomersall T, Minton J, Wong R, et al. Measurement of exhaled nitric oxide concentration in asthma: a systematic review and economic evaluation of NIOX MINO, NIOX VERO and NObreath. Health Technol Assess 2015 Oct;19(82):1-330 [FREE Full text] [doi: 10.3310/hta19820] [Medline: 26484874]
30. VanZeller C, Williams A, Pollock I. Comparison of bench test results measuring the accuracy of peak flow meters. BMC Pulm Med 2019 Apr 08;19(1):74 [FREE Full text] [doi: 10.1186/s12890-019-0837-3] [Medline: 30961573]
31. Tinschert P, Jakob R, Barata F, Kramer J, Kowatsch T. The potential of mobile apps for improving asthma self-management: a review of publicly available and well-adopted asthma apps. JMIR Mhealth Uhealth 2017 Aug 02;5(8):e113 [FREE Full text] [doi: 10.2196/mhealth.7177] [Medline: 28768606]
32. Shuwandy ML, Zaidan BB, Zaidan AA, Albahri AS. Sensor-based mhealth authentication for real-time remote healthcare monitoring system: a multilayer systematic review. J Med Syst 2019 Jan 06;43(2):33. [doi: 10.1007/s10916-018-1149-5] [Medline: 30612191]
33. Harst L, Lantzsch H, Scheibe M. Theories predicting end-user acceptance of telemedicine use: systematic review. J Med Internet Res 2019 May 21;21(5):e13117 [FREE Full text] [doi: 10.2196/13117] [Medline: 31115340]
34. Hui CY, Walton R, McKinstry B, Jackson T, Parker R, Pinnock H. The use of mobile applications to support self-management for people with asthma: a systematic review of controlled studies to identify features associated with clinical effectiveness and adherence. J Am Med Inform Assoc 2017 May 01;24(3):619-632 [FREE Full text] [doi: 10.1093/jamia/ocw143] [Medline: 27694279]
35. Hanlon P, Daines L, Campbell C, McKinstry B, Weller D, Pinnock H. Telehealth interventions to support self-management of longterm conditions: a systematic metareview of diabetes, heart failure, asthma, chronic obstructive pulmonary disease, and cancer. J Med Internet Res 2017 May 17;19(5):e172 [FREE Full text] [doi: 10.2196/jmir.6688] [Medline: 28526671]
36. Hanson WE, Creswell JW, Clark VLP, Petska KS, Creswell JD. Mixed methods research designs in counseling psychology. J Couns Psychol 2005 Apr;52(2):224-235. [doi: 10.1037/0022-0167.52.2.224]
37. 2019/20 General Medical Services (GMS) contract Quality and Outcomes Framework (QOF). National Health Service. 2019. URL: https://www.england.nhs.uk/wp-content/uploads/2019/05/gms-contract-qof-guidance-april-2019.pdf [accessed 2020-12-26]
38. McKinstry B, Sullivan FM, Vasishta S, Armstrong R, Hanley J, Haughney J, et al. Cohort profile: the Scottish Research register SHARE. a register of people interested in research participation linked to NHS data sets. BMJ Open 2017 Feb 01;7(2):e013351 [FREE Full text] [doi: 10.1136/bmjopen-2016-013351] [Medline: 28148535]
39. REACH. Asthma UK Centre for Applied Research. URL: https://www.aukcar.ac.uk/reach [accessed 2020-12-26]
40. Fonteyn ME, Kuipers B, Grobe SJ. A description of think aloud method and protocol analysis. Qual Health Res 2016 Jul 01;3(4):430-441. [doi: 10.1177/104973239300300403]

41. Guest G, Bunce A, Johnson L. How many interviews are enough? Field Methods 2016 Jul 21;18(1):59-82. [doi: 10.1177/1525822x05279903]
42. Francis JJ, Johnston M, Robertson C, Glidewell L, Entwistle V, Eccles MP, et al. What is an adequate sample size? operationalising data saturation for theory-based interview studies. Psychol Health 2010 Dec;25(10):1229-1245. [doi: 10.1080/08870440903194015] [Medline: 20204937]
43. BTS/SIGN British guideline on the management of asthma. British Thoracic Society. 2019. URL: https://brit-thoracic.org.uk/quality-improvement/guidelines/asthma/ [accessed 2020-12-26]
44. Witry M, Comellas A, Simmering J, Polgreen P. The association between technology use and health status in a chronic obstructive pulmonary disease cohort: multi-method study. J Med Internet Res 2018 Apr 02;20(4):e125 [FREE Full text] [doi: 10.2196/jmir.9382] [Medline: 29610113]
45. Horne R, Cooper V, Wileman V, Chan A. Supporting adherence to medicines for long-term conditions. Eur Psychologist 2019 Jan;24(1):82-96. [doi: 10.1027/1016-9040/a000353]
46. Shahaj O, Denneny D, Schwappach A, Pearce G, Epiphaniou E, Parke HL, et al. Supporting self-management for people with hypertension: a meta-review of quantitative and qualitative systematic reviews. J Hypertens 2019 Feb;37(2):264-279. [doi: 10.1097/HJH.0000000000001867] [Medline: 30020240]
47. Captieux M, Pearce G, Parke HL, Epiphaniou E, Wild S, Taylor SJC, et al. Supported self-management for people with type 2 diabetes: a meta-review of quantitative systematic reviews. BMJ Open 2018 Dec 14;8(12):e024262 [FREE Full text] [doi: 10.1136/bmjopen-2018-024262] [Medline: 30552277]
48. Hui CY, McKinstry B, Walton R, Pinnock H. A mixed method observational study of strategies to promote adoption and usage of an application to support asthma self-management. J Innov Health Inform 2019 Jan 09;25(4):243-253 [FREE Full text] [doi: 10.14236/jhi.v25i4.1056] [Medline: 30672405]
49. O'Connell S, McCarthy VJ, Savage E. Self-management support preferences of people with asthma or chronic obstructive pulmonary disease: a systematic review and meta-synthesis of qualitative studies. Chronic Illn 2019 Aug 19:1742395319869443. [doi: 10.1177/1742395319869443] [Medline: 31426658]
50. Sahin C, Courtney KL, Naylor PJ, E Rhodes R. Tailored mobile text messaging interventions targeting type 2 diabetes self-management: a systematic review and a meta-analysis. Digit Health 2019;5:2055207619845279 [FREE Full text] [doi: 10.1177/2055207619845279] [Medline: 31041110]

Abbreviations

AUKCAR: Asthma UK for Applied Research
IoT: internet of things
NHS: National Health Service
REACH: Register for Asthma Research
SHARE: Scottish Health Research Register

Edited by G Eysenbach; submitted 11.07.20; peer-reviewed by M Witry, L Kayser; comments to author 03.09.20; revised version received 23.01.21; accepted 15.02.21; published 13.04.21.

Please cite as:
Hui CY, McKinstry B, Fulton O, Buchner M, Pinnock H
Patients' and Clinicians' Visions of a Future Internet-of-Things System to Support Asthma Self-Management: Mixed Methods Study
J Med Internet Res 2021;23(4):e22432
URL: https://www.jmir.org/2021/4/e22432
doi:10.2196/22432
PMID:33847592

Review

Blockchain Personal Health Records: Systematic Review

Hao Sen Andrew Fang[1], MBBS, MMed, MTech; Teng Hwee Tan[2], MBBS, FRANZCR; Yan Fang Cheryl Tan[3], MBBS, MMed, FCFP; Chun Jin Marcus Tan[2], MBBS, MMed, FRCOphth, MTech

[1]SingHealth Polyclinics, Singapore, Singapore

[2]National University Health System, Singapore, Singapore

[3]SingHealth Community Hospitals, Singapore, Singapore

Corresponding Author:
Hao Sen Andrew Fang, MBBS, MMed, MTech
SingHealth Polyclinics
167, Jalan Bukit Merah
Connection One, Tower 5, #15-10
Singapore, 150167
Singapore
Phone: 65 93690001
Email: andrew.fang.h.s@singhealth.com.sg

Abstract

Background: Blockchain technology has the potential to enable more secure, transparent, and equitable data management. In the health care domain, it has been applied most frequently to electronic health records. In addition to securely managing data, blockchain has significant advantages in distributing data access, control, and ownership to end users. Due to this attribute, among others, the use of blockchain to power personal health records (PHRs) is especially appealing.

Objective: This review aims to examine the current landscape, design choices, limitations, and future directions of blockchain-based PHRs.

Methods: Adopting the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-analyses) guidelines, a cross-disciplinary systematic review was performed in July 2020 on all eligible articles, including gray literature, from the following 8 databases: ACM, IEEE Xplore, MEDLINE, ScienceDirect, Scopus, SpringerLink, Web of Science, and Google Scholar. Three reviewers independently performed a full-text review and data abstraction using a standardized data collection form.

Results: A total of 58 articles met the inclusion criteria. In the review, we found that the blockchain PHR space has matured over the past 5 years, from purely conceptual ideas initially to an increasing trend of publications describing prototypes and even implementations. Although the eventual application of blockchain in PHRs is intended for the health care industry, the majority of the articles were found in engineering or computer science publications. Among the blockchain PHRs described, permissioned blockchains and off-chain storage were the most common design choices. Although 18 articles described a tethered blockchain PHR, all of them were at the conceptual stage.

Conclusions: This review revealed that although research interest in blockchain PHRs is increasing and that the space is maturing, this technology is still largely in the conceptual stage. Being the first systematic review on blockchain PHRs, this review should serve as a basis for future reviews to track the development of the space.

(J Med Internet Res 2021;23(4):e25094) doi:10.2196/25094

KEYWORDS

blockchain; personal health records; electronic health records; distributed ledger; systematic review

Introduction

Background

Personal health records (PHRs) are a form of electronic health records (EHRs). PHRs are unique in that patients themselves can access, manage, and share their health information [1]. The benefits of PHRs include patient empowerment, which leads to improved outcomes and reduced health care costs [2,3]. Although interest in PHRs has been increasing, their adoption remains low [4,5]. One of the oft-cited reasons is related to privacy and security concerns owing to an increasing trend of

health information breaches [6,7]. Another reason is the lack of perceived usefulness to patients [7].

Blockchain technology was introduced through Bitcoin in 2008 [8]. It is considered a general-purpose technology and has since been successfully applied across several different industries [9,10]. In the health care industry, EHRs were found to be the most commonly used case for blockchain applications [11-14]. Compared with conventional data management methods that rely on on-premise data servers or third-party cloud services, blockchain's distributed ledger technology offers a novel alternative. This could potentially address the privacy and security concerns surrounding EHRs [15]. Specifically for application to PHRs, blockchain also has the ability to decentralize control and incorporate incentive mechanisms through smart contracts, which can further entice its general use and increase adoption [16]. These advantages, among others, have motivated efforts to test the feasibility and implement blockchain PHRs [17-19].

The research space in which EHRs and blockchain intersect is still in its infancy, with the first blockchain EHR introduced in 2016 [20]. Systematic reviews covering this space so far have considered EHRs as a collective entity. Mayer et al [21] provided an overview of the ecosystem of blockchain EHRs while also proposing a taxonomy for the space. Shuaib et al [22] looked at the main areas of focus when implementing a blockchain EHR and the remaining issues to be addressed, whereas Vazirani et al [23] assessed the feasibility of blockchain as a method of managing health care records efficiently.

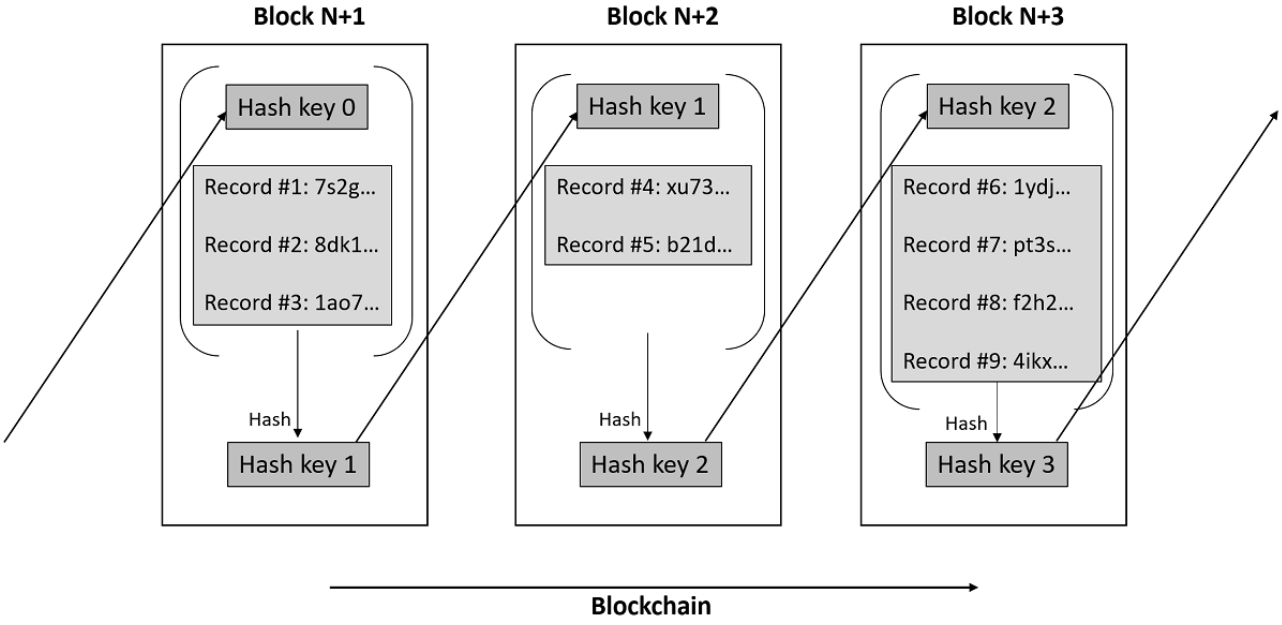
Given that one of the inherent properties of blockchain is its decentralized nature, in which data ownership is placed in the hands of individual users, some have proposed that blockchain may be more suitably applied to PHRs specifically rather than EHRs in general [19,24-26]. In this paper, we aim to systematically review the following: (1) the current landscape and trends of blockchain-based PHRs (blockchain PHRs), (2) the attributes of various blockchain PHRs that have been described, and (3) the current limitations and future directions for blockchain PHRs. To the best of our knowledge, this is the first systematic review examining blockchain with PHRs. We hope that this review will serve as a useful reference, especially for those intending to develop a blockchain PHR and for future reviews in this area.

To provide more context for subsequent sections of this paper, we will first explain pertinent blockchain concepts and take the opportunity to introduce some terminology specific to blockchain. This is by no means an exhaustive explanation of blockchain.

What Is a Blockchain?

A blockchain can be thought of as a shared (or distributed) database that is spread across multiple sites and participants. For new data to be added to a blockchain, they are first compiled into a *block*, which is simply a collection of records to be added to the database. The block is then combined with some data (a *hash key*) from the previous block through a cryptographic technique called *hashing* before it is added. As it combines the previous block's hash key, each new block is tied to all its predecessors in the form of a chain—hence the term *blockchain* (Figure 1).

Figure 1. Illustration of how blocks of data are linked together in a blockchain through hashing. To add a new record (eg, Record #7) to the blockchain, this is first grouped with other records (Records #6, #8, and #9). The group of records is then combined with a hash key from the previous block (hash key 2) and then put through a hashing algorithm to produce a new hash key (hash key 3). The new records, along with hash keys 2 and 3, are now part of a new block (Block N+3) that has been added to the blockchain. This process continues as new records are added.



Types of Blockchains and Their Properties

Before data can be added to a blockchain, its users need to agree or reach *consensus*. This is achieved through a *consensus algorithm*. A well-known consensus algorithm is the proof of work (PoW) algorithm. PoW is used in the Bitcoin and Ethereum blockchain network protocols [8,27]. In the PoW algorithm, users (also known as *miners*) compete in computational tasks to reach consensus. The winning miner of each block's task is usually given a reward [28].

Blockchains can be classified into the following three types, depending on which participants are allowed in the consensus algorithm [28]:

1. Public: anyone can participate in the consensus algorithm. Examples include Bitcoin and Ethereum [8,27].
2. Consortium: a select (or permissioned) group of entities can participate in the consensus algorithm. Examples include Hyperledger Fabric (HF), Quorum, and Corda [29-31].
3. Private: only a single entity operates the consensus algorithm and controls the addition of new data.

Public blockchains are sometimes referred to as permissionless blockchains, whereas consortium and private blockchains are collectively termed permissioned blockchains.

The three types of blockchains differ in the following properties:

1. Decentralization: unlike traditional databases that are owned by a specific entity, a decentralized blockchain can allow every user to own the data collectively. Using the illustration in Figure 1 as an example, a decentralized blockchain would contain all the records, but only one user owns records #1, #3, #4, #6, and #8, and another user separately owns records #2, #5, #7, and #9.
2. Immutability: because of the underlying chain structure, once data have been added to the blockchain, they cannot be tampered with. Changing a record would alter the hash key and effectively cause a break in the chain.
3. Transparency (with privacy): the entire blockchain can be made publicly viewable while preserving privacy by masking each individual record using cryptography. To unmask one's own records, a private key is required.

Table 1 provides a summary of the different types of blockchains and their properties, with an example of each type in the health care setting.

Table 1. Comparison of public, consortium, and private blockchains.

Variables	Type of blockchain		
	Public	Consortium	Private
Participation in consensus protocol	Anyone	Select (or permissioned) group of entities	A single entity
Decentralization	Yes	Partial	No
Immutability	Tamperproof	Could be tampered with	Could be tampered with
Transparency	Public	Can be public or restricted	Can be public or restricted
Example in health care setting	A transnational, open EMR[a] system in which anyone (eg, health care institutions, patients) may choose to contribute their own resources to maintain health records of all patients who use the EMR system	A national EMR system in which selected health care institutions collectively maintain the health care records of their patients	An institution-based EMR system in which only a single institution maintains the health records of its own patients

[a]EMR: electronic medical record.

Scalability and Smart Contracts

Finally, we will briefly explain the two concepts of *scalability* and *smart contracts*, which will be relevant to subsequent parts of this paper.

Scalability refers to the capacity of a blockchain to store and process transactions. It generally relates to the size and frequency of transactions a blockchain can handle. For example, Bitcoin's block size is limited to 1 megabyte, and each block is added every 10 minutes. This translates to a rate of approximately 7 transactions per second. Various solutions have been proposed to improve scalability. One such solution is to store data *off-chain* (instead of on-chain), and another solution is to use *side-chains* (linked to the main chain) to enable larger transaction volumes to be processed in parallel. Given that health care data are estimated to reach as much as 2314 exabytes generated yearly by 2020, it is crucial for almost all blockchain-based health care applications to achieve a certain level of scalability [32].

Smart contracts are programmable computer rules. Blockchain is a digital database that allows for the implementation of smart contracts, which can be automatically triggered to execute when predefined conditions are satisfied. For example, a smart contract can be programmed to issue tokens on the blockchain each time a user records his or her blood pressure. These tokens can then be used to pay for health care services. Such smart contracts can thus potentially be used to enable incentive structures to encourage certain positive user behaviors.

In this systematic review, particularly focused on the blockchain component of blockchain PHRs, we will pay particular attention to the (1) type of blockchain, (2) scalability solutions, and (3) smart contract–based incentive structures.

Methods

Study Design

While conducting and reporting this systematic literature review, the guidelines described in the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-analyses) statement were adopted [33]. This type of literature review was selected because the goal was to identify articles on blockchain PHRs and to summarize the current landscape, design choices, limitations, and future directions. Unlike a meta-analysis, this review did not require any data synthesis. Quality assessment was not performed because the intention was to achieve a collective understanding of the efforts and ideas rather than judging the quality of various blockchain PHRs.

The presented systematic review was carried out by defining the following activities:

1. Research questions
2. Search strategy
3. Article selection
4. Data abstraction

Research Questions

For this review, there were 3 research questions we aimed to address:

1. What are the current landscape and trends of blockchain PHRs in terms of interest groups, geography, and maturity level?
2. What were the key design decisions made for the blockchain PHRs described?
3. What are the current limitations faced by blockchain PHRs and future directions?

Search Strategy

The following search string was used: “blockchain” AND (“health record*” OR “medical record*” OR “*EHR*” OR “*EMR*” OR “*PHR*”). Articles in the following databases were searched: (1) ACM, (2) IEEE Xplore, (3) MEDLINE, (4) ScienceDirect, (5) Scopus, (6) SpringerLink, (7) Web of Science, and (8) Google Scholar. For databases whose search engines did not enable the use of wildcards, the search was widened to include abstracts and keywords, and Microsoft Excel was subsequently used to filter the returned list by applying the search string to the titles.

As the space is still in its infancy stage, Google Scholar was included as a search database to incorporate relevant gray literature in this review. This decision was supported by systematic reviews by Holbl et al [34] and Kuo et al [35] on blockchain in the health care domain, which had found valuable information residing in gray literature.

Article Selection

Once the articles were obtained, we applied the following inclusion and exclusion criteria to select articles for the final review. The inclusion criteria were as follows: a health record system that had (1) a patient-facing component and (2) used blockchain in its health record system. The exclusion criteria were as follows: (1) duplicate articles, (2) review articles, (3) articles that did not have full text available, and (4) articles whose full text was not in English.

The selection was performed in a stepwise manner. First, duplicate articles returned from multiple databases were excluded. Second, the titles of the articles were reviewed and those that were not relevant to the topic were discarded. Third, the abstracts of the articles were reviewed and those whose main focus was not on blockchains and EHR or PHR and those that were review articles were also discarded. Those that looked at EHRs at this stage were retained because some EHRs would have a patient-facing component but might not have been explicitly mentioned in the title or abstract. Finally, the full text was reviewed and those that did not have a PHR element in the EHR were discarded. At this stage, those that did not have full text available or whose full text was not in English were also excluded.

Data Abstraction

For data abstraction, a standardized data collection form was developed using Microsoft Excel. A full-text review of each selected study was performed independently by 3 reviewers who are knowledgeable about blockchain and health records. For discrepancies in the abstracted data, the reviewers performed a repeat review of the articles together to reach a consensus.

For the interest groups, author affiliations, publishers, and publications were used as a proxy. As this space is situated at the intersection of computer science (CS), engineering, and medicine, we classified the publications into either (1) CS or engineering, (2) medical, or (3) general. For maturity level, the classification used by Chukwu et al [12] was modified, and the projects were classified as *concept/model/framework*, *prototypes*, and *pilots or implementations*. A prototype was considered to have both a working front-end and back-end system, and a pilot or implementation had to be a product that was released for use in the real world. If an article described systems at multiple levels of maturity (eg, a framework and a prototype), only the more mature level described was abstracted.

Many design choices must be made when developing a blockchain PHR. To keep this review manageable, the review focused on high-level design decisions [36]. To ensure a comprehensive list of possible design parameters, the *PHR taxonomy* proposed by Roehrs et al [37] and *EHR in a Blockchain taxonomy* proposed by Mayer et al [21] were used as starting points. Next, through a consensus-driven process of elimination, 10 design parameters were selected for abstraction. These were (1) blockchain type, (2) data storage, (3) scaling solution, (4) incentive smart contract, (5) PHR type, (6) data owner, (7) read and write ability, (8) semantic standards, (9) privacy standards, and (10) user interface (UI).

For limitations and future directions, the issues and areas for improvement brought up across the articles reviewed were identified, consolidated, and presented as a list of unique issues. We did not delve into a more in-depth analysis such as ranking the unique issues because the frequency of mention was not necessarily associated with importance or criticality. Moreover, the articles may not have fully listed all their limitations, as it was not their primary aim.

In total, 23 data elements were extracted from each article. Table 2 provides a complete list of the extracted data elements and a description of each element.

Table 2. List of data elements extracted from the selected articles.

Types of data elements	Description
General	
Author	First author's last name
Title	Title of the article
Year	Publication year of the article
Country	First author's affiliated country
Type	Type of article (eg, journal article, conference paper, book chapter, and whitepaper)
Publisher	Name of publisher of the article
Blockchain	
Name of blockchain PHR[a]	Name of the blockchain PHR (if any)
Maturity	Maturity level of the blockchain PHR described (ie, concept/framework, prototype, and pilot/implementation)
Blockchain type	Type of blockchain (ie, public, consortium, private, or a combination)
Blockchain name	Name of the blockchain used (if any)
Data storage	Type of data storage mechanism (ie, on-chain, off-chain, or hybrid)
Scaling solution	Type of scaling solution used in the blockchain PHR (if any)
Incentive smart contract	Was a smart contract used to incentivize use of the blockchain PHR? (yes or no)
PHR	
PHR type	Type of PHR (ie, standalone or tethered to an existing EMR[b] system)
Data owner	Party that owned the data from the blockchain PHR (ie, patient, provider, or both)
Read-write access (for patients)	Was the patient given read and/or write access in the blockchain PHR? (yes or no)
Read-write access (for providers)	Was the provider given read and/or write access in the blockchain PHR? (yes or no)
Read-write access (for other parties)	Were other parties given read and/or write access in the blockchain PHR? (yes or no)
Semantic standard	Type of semantic standard adopted (eg, HL-7[c] and FHIR[d])
Privacy standard	Type of privacy standard adopted (eg, HIPAA[e] and GDPR[f])
User interface	Modality of accessing the blockchain PHR (ie, web, mobile, or desktop application)
Additional	
Limitations	Current limitations of the blockchain PHR
Future directions	Future directions and opportunities described

[a]PHR: personal health record.

[b]EMR: electronic medical record.

[c]HL-7: health level 7.

[d]FHIR: Fast Healthcare Interoperability Resource.

[e]HIPAA: Health Insurance Portability and Accountability Act.

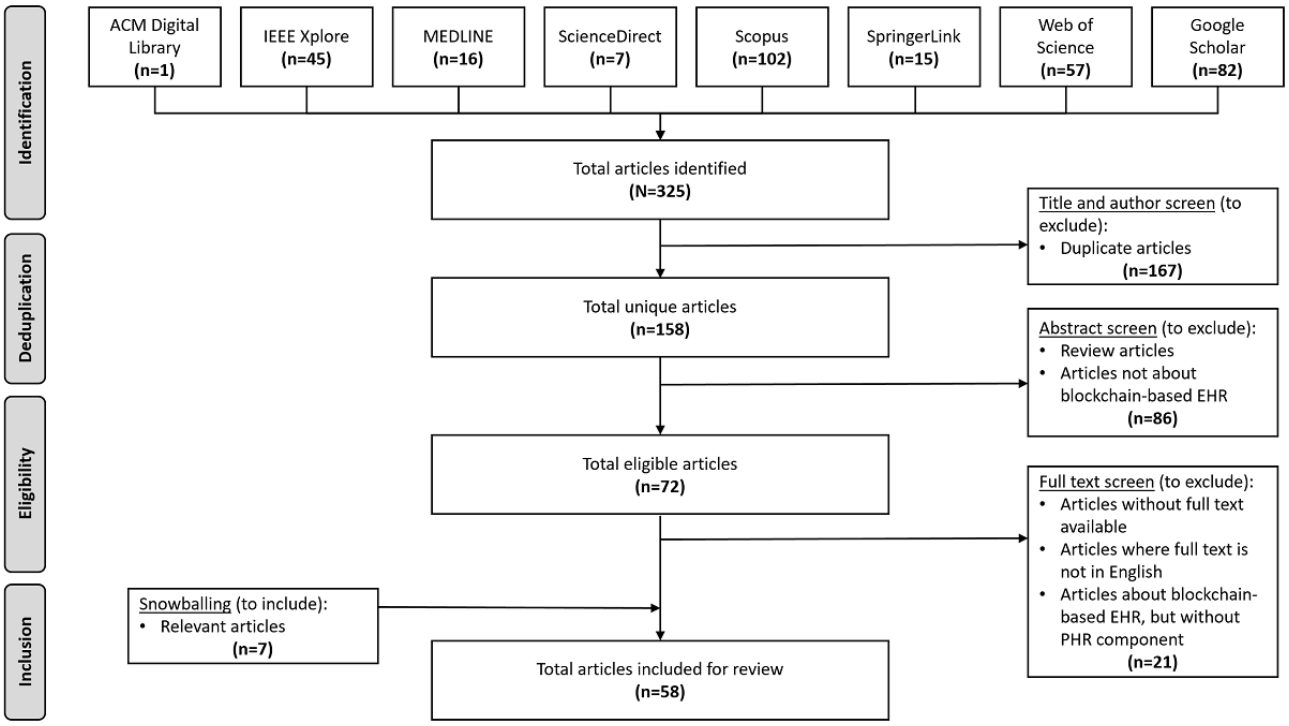
[f]GDPR: General Data Protection Regulation.

Results

Overview of Articles

The search performed on July 6, 2020, yielded 325 articles, of which 158 were unique articles. From the article selection process, 51 articles were selected for review. An additional 7 articles were added via snowballing (review of the references from the included articles) of the full texts screened (Figure 2).

Figure 2. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-analyses) flow diagram of the article selection process. The title and author screen involved removing duplicate articles that had the same title and authors. The abstract screen involved reviewing article abstracts to remove review articles and those not related to blockchain and electronic health records. The full text screen involved reviewing the full articles to exclude those that did not meet the inclusion and exclusion criteria, and those whose full text was not available or in English. ACM: Association of Computing Machinery; EHR: electronic health record; IEEE: Institute of Electrical and Electronics Engineers; PHR: personal health record.



A total of 58 studies were included in the final review [17,19,37-92]. The complete list of articles, with identifiers used in this study, are presented in Table 3. The completed data collection form for these articles can be found in Multimedia Appendix 1. An overview of the articles with the publication year, publisher, article type, country, and interest group is presented in Multimedia Appendix 2 [17,19,37-92].

Table 3. List of articles included in the final review.

Article identifier	Authors	Article title
A01	Burniske [37]	How blockchain technology can enhance EHR[a] interoperability
A02	McFarlane et al [38]	Patientory: A Healthcare Peer-to-Peer EMR[b] Storage Network
A03	Roehrs et al [39]	OmniPHR: A distributed architecture model to integrate personal health records
A04	Badr et al [40]	Multi-tier blockchain framework for IoT[c]-EHRs systems
A05	Boiani [41]	Blockchain based electronic health record management for mass crisis scenarios: A feasibility study
A06	Chen et al [42]	Blockchain-Based Medical Records Secure Storage and Medical Service Framework
A07	Dagher et al [43]	Ancile: Privacy-preserving framework for access control and interoperability of electronic health records using blockchain technology
A08	Dubovitskaya et al [44]	Secure and Trustable Electronic Medical Records Sharing using Blockchain
A09	Gebremedhin [45]	Blockchain as a Technology to Facilitate Privacy and Better Health Record Management
A10	Lippman et al [46]	MedRec: Patient Control of Medical Record Distribution
A11	Medicalchain [47]	Medicalchain
A12	Rouhani et al [48]	MEDICHAIN: A Secure Decentralized Medical Data Asset Management System
A13	Thwin and Vasupongayya [49]	Blockchain Based Secret-Data Sharing Model for Personal Health Record System
A14	Vora et al [50]	BHEEM[d]: A Blockchain-Based Framework for Securing Electronic Health Records
A15	Zhang and Poslad [51]	Blockchain Support for Flexible Queries with Granular Access Control to Electronic Medical Records (EMR)
A16	Abouzahra [52]	Using blockchain technology to enhance the use of personal health records
A17	Alkhushayni et al, [53]	Blockchain technology applied to electronic health records
A18	Ray Chawdhuri [54]	Patient Privacy and Ownership of Electronic Health Records on a Blockchain
A19	Ciampi [55]	A Blockchain Architecture for the Italian EHR System
A20	Daraghmi et al [56]	MedChain: A Design of Blockchain-Based System for Medical Records Access and Permissions Management
A21	Donawa et al [57]	Scaling Blockchains to Support Electronic Health Records for Hospital Systems
A22	Hang et al [58]	A novel EMR integrity management based on a medical blockchain platform in hospital
A23	Harika et al [59]	Blockchain technology for managing an architectural model of decentralized medical record
A24	Huang et al [60]	MedBloc: A Blockchain-Based Secure EHR System for Sharing and Accessing Medical Data
A25	Hylock and Zeng [61]	A Blockchain Framework for Patient-Centered Health Records and Exchange (HealthChain): Evaluation and Proof-of-Concept Study
A26	Jiang et al [62]	Patients-Controlled Secure and Privacy-Preserving EHRs Sharing Scheme Based on Consortium Blockchain
A27	Koushik et al [63]	Performance Analysis of BlockChain-based Medical Records Management System
A28	Lee [64]	PHR[e] system using blockchain technology
A29	MediBloc [65]	MediBloc Technical Whitepaper
A30	MediLOT [66]	MediLOT Whitepaper
A31	Nchinda et al [67]	MedRec: A Network for Personal Information Distribution
A32	Nguyen et al [68]	Blockchain for Secure EHRs Sharing of Mobile Cloud Based E-Health Systems
A33	Park et al [17]	Is Blockchain Technology Suitable for Managing Personal Health Records? Mixed-Methods Study to Test Feasibility
A34	Rajput et al [69]	EACMS: Emergency Access Control Management System for Personal Health Record Based on Blockchain
A35	Reen et al [70]	Decentralized patient centric e-Health record management system using blockchain and IPFS[e]

Article identifier	Authors	Article title
A36	Sangeetha [71]	Electronic Health Record System using Blockchain
A37	Shahnaz et al [72]	Using Blockchain for Electronic Health Records
A38	Shekhawat [73]	Cloud-chain: Revamp Health Record System Using Blockchain
A39	Thwin and Vasupongayya [74]	Blockchain-Based Access Control Model to Preserve Privacy for Personal Health Record Systems
A40	Tian [75]	Blockchain-based secure medical record sharing system
A41	Toshniwal et al [76]	PACEX: Patient-centric EMR exchange in Healthcare Systems using Blockchain
A42	Wang et al [77]	Blockchain-Based Personal Health Records Sharing Scheme With Data Integrity Verifiable
A43	Wang et al [78]	Cloud-Assisted EHR Sharing with Security and Privacy Preservation via Consortium Blockchain
A44	Wu and Du [79]	Electronic medical record security sharing model based on blockchain
A45	Al Goni et al [80]	A P2P Optimistic Fair-Exchange Scheme for Personal Health Records Using Blockchain Technology
A46	Arunkumar and Kousalya [81]	Blockchain-Based Decentralized and Secure Lightweight E-Health System for Electronic Health Records
A47	Aswin et al [82]	Design of AYUSH: A blockchain-based health record management system
A48	Cao et al [83]	Hybrid blockchain–based privacy-preserving electronic medical records sharing scheme across medical information control system
A49	Charanya et al [84]	Sefra: A secure framework to manage eHealth records using blockchain technology
A50	Kavathekar and Patil [85]	Data sharing and privacy-preserving of medical records using blockchain
A51	Kim et al [86]	Design of Secure Protocol for Cloud-Assisted Electronic Health Record System Using Blockchain
A52	Kung et al [87]	Personal Health Record in FHIR[f] Format Based on Blockchain Architecture
A53	Lee et al [19]	An Architecture and Management Platform for Blockchain-Based Personal Health Record Exchange: Development and Usability Study
A54	Sharma et al [88]	Secure Cloud Storage Architecture for Digital Medical Record in Cloud Environment using Blockchain
A55	Sharma and Balamurugan [89]	Preserving the Privacy of Electronic Health Records using Blockchain
A56	Tith et al [90]	Application of Blockchain to Maintaining Patient Records in Electronic Health Record for Enhanced Privacy, Scalability, and Availability
A57	Verdonck and Poels [91]	Architecture and value analysis of a blockchain-based electronic health record permission management system
A58	Wu et al [92]	Secure Personal Health Records Sharing Based on Blockchain and IPFS[g]

[a]EHR: electronic health record.

[b]EMR: electronic medical record.

[c]IoT: internet of things.

[d]BHEEM: Blockchain-based framework for efficient storage and maintenance of electronic health records

[e]PHR: personal health record.

[f]FHIR: Fast Healthcare Interoperability Resource.

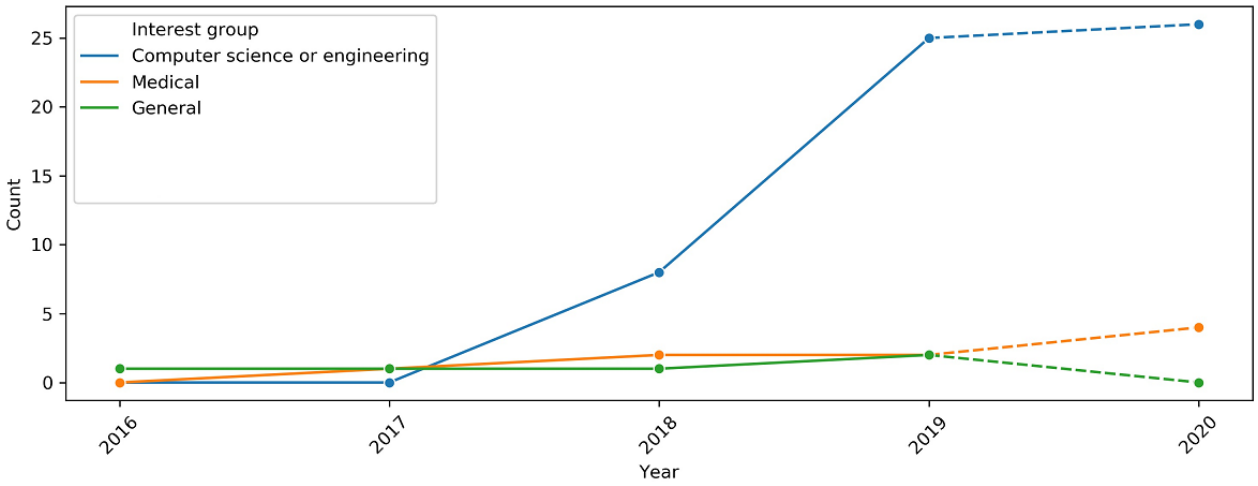
[g]IPFS: Interplanetary File System.

Current Landscape and Trends of Blockchain PHRs

Interest Group

The level of academic interest in the space has been rising, supported by an increasing trend in the number of published articles since 2016. In terms of interest groups, 45 articles were CS-or engineering-related publications or from CS- or engineering-related authors. Seven were published in medical journals, all of which were related to medical informatics. Of the 6 remaining articles that were classified as *General*, 5 were whitepapers. The articles from the CS or engineering interest group showed a sharp rise from 2017 to 2019 and may have started to plateau, whereas those from medical journals have been following a gradual, steady increase since 2016 (Figure 3).

Figure 3. Trend of blockchain personal health record articles by interest group. The trend from 2019 to 2020 (represented by dashed lines) is a projection because only data from the first half of the year 2020 was available at the time of the search. Count refers to the number of articles published in that year.



Geographic Distribution

The articles originated from 23 different countries. The majority were from India (n=13), United States (n=9), China (n=8), and South Korea (n=5), with Canada, Switzerland, Taiwan, and Thailand having 2 articles each and the remaining countries having 1 article each (Figure 4). Although the research interest in blockchain PHR is multinational, there clearly are a few countries that are leading the pack. Among these leading countries, there has been an increasing number of publications from India over the years, whereas China, South Korea, and the United States have shown a slowing trend. Apart from these countries, the aggregated output from the rest of the countries is also increasing (Figure 5).

Figure 4. Distribution of articles published by geography. The number of articles refers to the total number of articles selected for the final review.

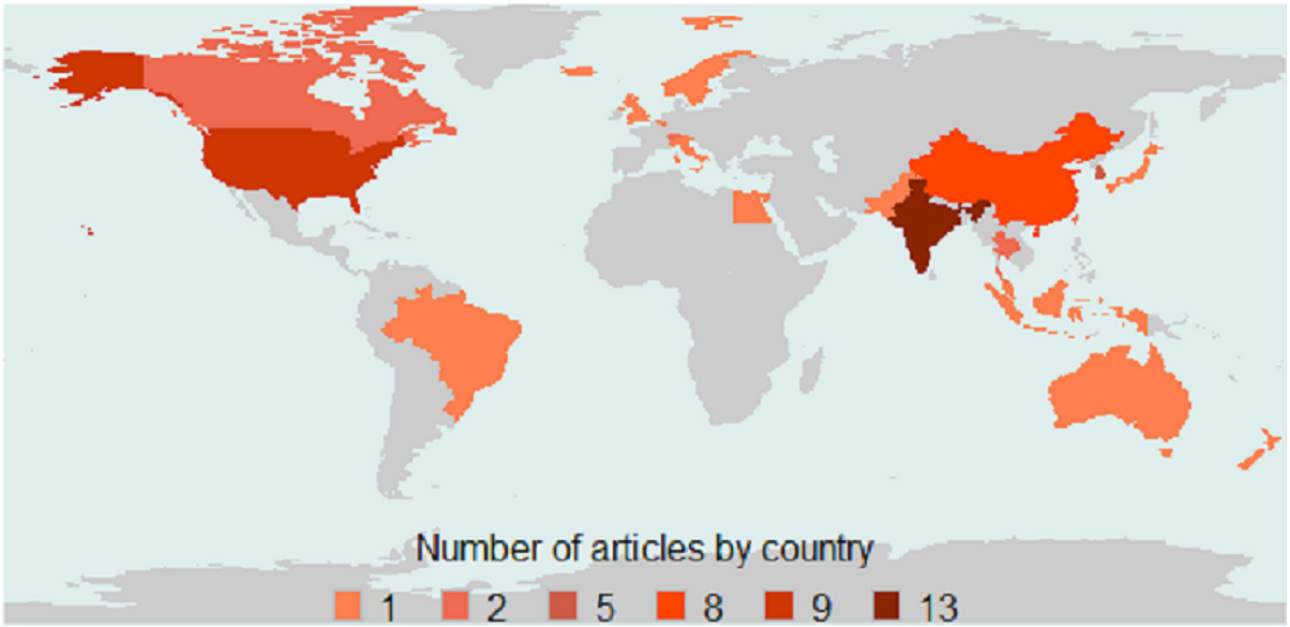
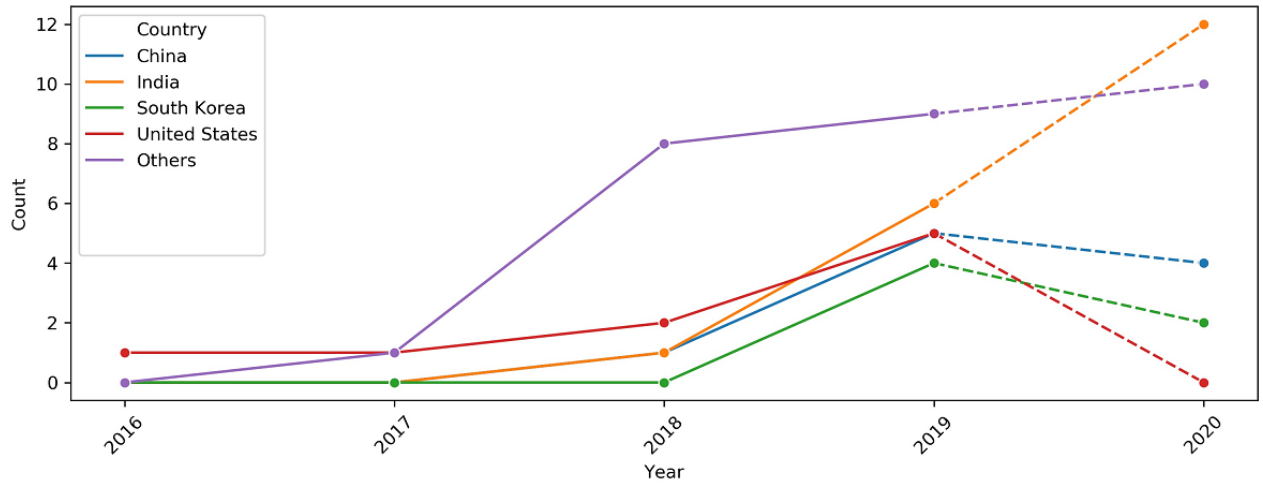


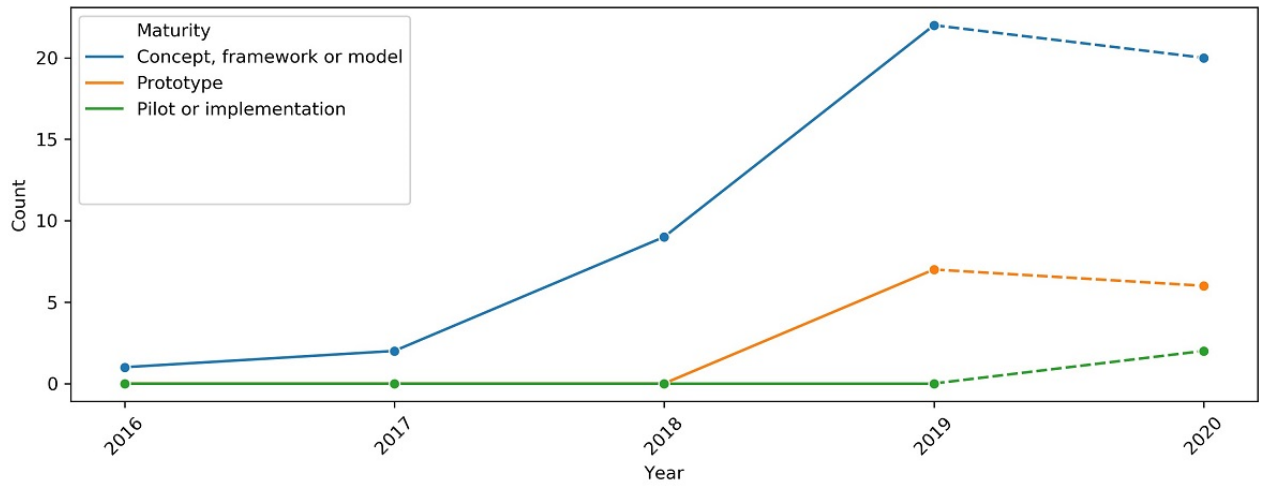
Figure 5. Trend of blockchain personal health record articles by country. Only countries with 5 or more articles in the final review were plotted individually. The other countries were grouped under an Others category. The trend from 2019 to 2020 (represented by dashed lines) is a projection because only data from the first half of the year 2020 was available at the time of the search. Count refers to the number of articles published in that year.



Maturity Level

The blockchain PHR space is maturing, with the proportion of articles describing prototypes showing an upward trend (Figure 6). In addition, the first paper to describe an implementation was also published in the first half of 2020 by Lee et al [19]. Their blockchain PHR implementation was deployed across Southeast Asia via an information network and became the first PHR management platform for cross-regional medical data exchange.

Figure 6. Trend of blockchain personal health record maturity. Note that the trend from 2019 to 2020 (represented by dashed lines) is a projection because only data from the first half of the year 2020 was available at the time of search. Count refers to the number of articles published in that year (2018).



Key Design Choices for Blockchain PHRs

Blockchain Attributes

Most blockchain PHRs are described using a private (n=24) or consortium (n=22) blockchain, whereas 4 others used a public-permissioned hybrid design. Only 4 cases of using a public blockchain were described. In the remaining 4 cases, the blockchain type was not clearly stated. An Ethereum-based blockchain was the most commonly used (n=26), with HF being the next most common (n=20). Among these, 3 articles used both Ethereum and HF.

For data storage, the majority used off-chain data storage (n=40), 14 stored EHR data on-chain, and 4 described hybrid data storage. For off-chain storage, 10 articles, all from 2019 onward, used the Interplanetary File System (IPFS). In terms of other scaling solutions, 9 articles considered new consensus algorithms such as Proof-of-Authority, 4 used a tiered-chain architecture, and 1 used both side-chain and algorithmic methods to improve the blockchain scaling capacity.

Among the articles, 5 described an incentive structure in the blockchain PHR using smart contracts. Four of these were whitepapers, which proposed incentivizing stakeholders through the issuing of tokens (digital currency of value) from smart contracts. In these cases, once an action warranting compensation had taken place, the smart contract automatically triggered the issuance of tokens. Table 4 provides additional

details of the tokens and how they can be earned and used as part of the incentive structure. Unlike the others, Daraghmi et al [56] proposed a novel, nonmonetary incentive. Their system kept score using *degrees* based on the effort in maintaining the quality of records and creating new blocks. Those with higher degrees would have a lower probability of performing the computation task of creating new blocks. In this way, it is meant to achieve fairness and sustainability of the system.

Table 4. Incentive structure proposed by blockchain personal health record systems.

Article identifier	Token name	Party compensated	Compensation	Token use
A02	Patientory issued tokens	Provider	• On the basis of how effective provider ensures improvement in care quality and outcomes	• Renting storage space on the platform • Execution of smart contracts
A11	MedToken	Patients	• Sharing of personal health data on health care data marketplace	• Lower insurance premiums • Payment for use of applications (eg, tele-consultations)
A29	MediBloc coin	Provider	• Performing computational task of producing blocks	• Tradable for monetary value
A30	LOT[a] token	Patients	• Contribute data • Compliance to health care recommendations	• Analytics services (eg, personalized health reports) • Payment to retail and pharmaceutical partners

[a]LOT: token used in the MediLOT system.

PHR Attributes

A total of 18 articles described a tethered blockchain PHR that interfaced with an existing electronic medical record (EMR) system. All of these were of the *Concept/Framework/Model* maturity level. Those that were prototypes, pilots, or implementations were all standalone PHR systems.

In the majority (n=45) of the articles, the patient was the data owner. Of the remaining articles, providers were data owners in 9 of them, whereas 2 had both patients and providers as owners. It was unclear who the data owner was in the last 2 articles.

In most articles, both patients and providers had read and write abilities. Most blockchain PHRs granted providers with both read and write abilities (n=40), and only 4 blockchain PHRs did not grant providers any read or write abilities. Table 5 is a matrix representing the distribution of read and write capabilities for patients and doctors among the various articles and article codes refer to article identifier in Table 3.

Table 5. Matrix of various read and write models among the blockchain personal health records reviewed.

Provider	Patient		
	Read only	Write only	Read and write
Read only	—[a]	—	A04, A15, A32, A42, A48, A50, A52, A58
Write only	A07, A14, A18, A31	—	—
Read and write	A01, A05, A08, A10, A12, A17, A19, A20, A23, A24, A27, A28, A30, A33, A40, A41, A43, A47, A49, A56	A25, A51	A02, A03, A06, A09, A11, A21, A22, A26, A34, A35, A37, A38, A39, A44, A45, A46, A53, A54, A55, A57
Neither	—	A16	A13, A29, A36

[a]Not available.

Most articles did not mention the adoption of any semantic standard. For those that did, the 2 standards mentioned were Fast Healthcare Interoperability Resource (FHIR) and health level 7 (HL-7) in 5 and 2 articles, respectively. Similarly, most did not mention adopting any privacy standards. For those that did, 4 mentioned compliance with the Health Insurance Portability and Accountability Act (HIPAA), 1 with the General Data Protection Regulation (GDPR), and 1 with both the HIPAA and GDPR.

Among the blockchain PHRs that were either prototypes or implementations, 9 developed a web UI, whereas 2 had both a mobile phone application UI and a desktop UI.

Current Limitations of and Future Directions for Blockchain PHRs

Current Limitations

Most of the current limitations can be grouped into 1 of the following 3 main categories: (1) scalability, (2) privacy, and (3) usability. Scalability issues pertained to the inability of

blockchain PHR to store large file sizes such as medical images [44,53,54] or to the slowness in confirming transactions, especially with the incorporation of streaming data from internet of things devices [45,76].

The inability of blockchain PHRs to ensure full privacy has been highlighted in a few articles. Although records on the blockchain are encrypted, there are possible means to infer the information, such as through blockchain analysis [17,43,54]. Another privacy issue raised was the inability to erase one's records, as blockchains are inherently immutable [17,70]. This limitation would make it difficult for blockchain PHRs to comply with privacy regulations such as the GDPR, which stipulates data subjects' right to erasure (Article 17 of the GDPR).

One of the usability limitations was the affordability of the blockchain PHR, as each transaction typically required users to pay a transaction fee [45,71]. Another practical usability issue described by Charanya et al [84] was that, unlike conventional PHRs that had password recovery mechanisms, patients would not be able to access their records if they lost their private keys on blockchain PHRs. Incapacitated or unconscious patients also present a similar problem with blockchain PHRs that do not have built-in access control when emergency health care providers would need permission to access records.

Apart from these 3 main categories, there were other limitations inherent to certain types of popular blockchains such as Ethereum. For example, Gebremedhin [45] highlighted that Solidity (Ethereum's programming language) was unable to implement nested string data types, whereas Kung et al [86] mentioned the need to batch upload data in a certain file format as a limitation of their Ethereum-based PHR.

Future Directions

The current limitations provide direction to some future work areas for blockchain PHRs. Scalability solutions have already been studied and experimented on, such as Proof-of-Authority and the novel Byzanthine fault tolerance (BFT) consensus mechanisms [44,56,67]. Other methods include enhancing the blockchain architecture through tiered-chain [40,64] or side-chain structures [57]. Although privacy solutions were more limited in our review, we came across one by Reen et al [70] who proposed storing InterPlantary Naming System records instead of the conventional hash of the medical records directly on the blockchain. In this way, users may retain the ability to revoke access to the record if desired.

Many suggestions have been made to improve the usability of the system. These suggestions could be grouped into (1) user experience, (2) integration with existing systems, and (3) compliance with regulations and development of governance processes. Table 6 summarizes the suggestions proposed in the articles reviewed.

Table 6. Suggestions for improving blockchain personal health record usability from the selected articles.

Suggestion	Article identifiers
User experience	
Improving user interface of blockchain PHRs[a]	A10, A31, A55
Biometric user authentication	A40
Allowing next-of-kin or caregiver to access records if patient grants access or is incapable of self-access	A40
Incorporating incentives for users	A56
Incorporating analytics capabilities for personal health insights and management	A32, A53
Adding on payment functions for health care services	A37, A55
Integration with existing systems	
Integrating with existing EMR[b] systems	A07, A08, A19, A20, A52,
Adopting health care data standards	A03, A09, A17, A53
Integrating with IoT[c] devices	A32
Integrating with open, public blockchain systems	A17, A18, A33
Compliance with regulations and development of governance processes	
Complying with regulations on health care data privacy	A07, A18
Developing governance processes for the blockchain PHRs	A38

[a]PHR: personal health record.

[b]EMR: electronic medical record.

[c]IoT: internet of things.

Apart from improving usability, another aspect of future work is the validation of blockchain PHRs. Among the areas for validation, several articles suggested data validation when data were transferred to off-chain storage [77], security validation [48,58,79], and real-world validation in terms of cost-effectiveness [52,53,71]. Validating these components

would be relevant to obtain stakeholder and user confidence in deciding where to implement and adopt blockchain PHRs.

Discussion

Principal Findings

In this first ever systematic review on blockchain PHRs, we adopted a broad search strategy across medical and CS and engineering databases and included gray literature. We focused on the scope of blockchain PHRs to allow for more targeted data abstraction. Through our study, we found that there was a growing interest in blockchain PHRs and that the space has been steadily maturing over the past few years, albeit still much in the conceptual stage. As the space is still fairly new, a lion's share of the research and innovation has been happening at the technical level to discover new ways to solve problems. This is evidenced by the overwhelming proportion of articles that have come from the CS and engineering domain.

One of the major areas regarding blockchain PHRs that is still undergoing much research is scalability. We came across a few ideas such as Proof-of-Authority, novel BFT consensus mechanisms, and other modified blockchain architectures such as tiered-chains and side-chains [40,44,56,57,64,67]. Apart from blockchain PHR teams working on this, the space may also benefit from parallel innovations from the larger blockchain ecosystem. As Ethereum is looking forward to a new version release (version 2), it is considering various scaling solutions, of which *rollups* is a strong contender [93]. *Rollups* solution essentially involves keeping transaction data on-chain while pushing the computational load off-chain. If adopted into Ethereum 2.0, this could automatically benefit many Ethereum-based PHRs.

Although some areas are actively evolving, others are beginning to consolidate. As found in other systematic reviews, most blockchain PHR project teams have gravitated toward Ethereum and HF as their blockchains of choice [22]. In addition, in terms of data storage, we see more projects opting for IPFS as a complementary off-chain data store for their blockchain PHRs [68,72,83,92]. Outside of this review, we are also aware that there are efforts happening in other public blockchains. An example is NEO, whose core developers are developing a similarly distributed, decentralized object storage network known as NEO file storage system (NeoFS), which will seamlessly integrate with its native blockchain [94,95]. We did not come across any NEO-based PHRs in this review. NeoFS could potentially be a game changer, so it would be interesting to track its development in this area.

In this review, we also identified some current limitations that blockchain PHRs need to address. We broadly classified them into scalability, privacy, and usability limitations. In addition to identifying the current limitations, this review also revealed some possible solutions. For example, to address the privacy issue of inferring information from chain analysis, Ray Chawdhuri [54] introduced zero-knowledge provable mixing, whereas Park et al [17] proposed the zero-knowledge succinct non-interactive argument of knowledge technique. Another example is the solution of using biometric authentication mentioned by Tian [75] to address the issues of verifiable user authentication and patients losing their private keys. Medicalchain has also described an emergency bracelet that can be scanned, giving access to essential health information in unconscious patients who are unable to access their private keys [47].

The first blockchain PHR has already been piloted, and this will undoubtedly augur a move of the space toward deployment [19]. With this in mind, blockchain PHRs will need to comply with the privacy standards within the jurisdictions they intend to become operational. In addition, to enable integration with existing health care EMR systems, it is necessary to design blockchain PHRs that follow established semantic standards such as HL-7 and FHIR. Looking further ahead, to realize true decentralization, it may be necessary to consider building a PHR atop public blockchains.

Finally, in terms of geographic interest, we found that although interest in blockchain PHR was multinational, there were obvious leaders in this space. Looking deeper among the leading countries, we noticed that since 2018 there has been an increase in publications from India, whereas those from China, South Korea, and the United States started to level off or decrease. A possible reason for this could be that in 2017, in the midst of an initial coin offering (ICO) fever that drove unusually high interest in blockchain, the latter 3 countries' relevant authorities had issued bans or indicated legal restrictions on ICO activities with stiff penalties [96-98]. This may suggest further research into the different factors, including sociopolitical, economic, and cultural factors, which could significantly impact the development of this space. In terms of interest groups, our findings should also provide a sense of where most of the developments are occurring, and this may guide government and private sector funders in their allocation of resources.

Limitations

We acknowledge that this review is not exhaustive and that there are many other areas that were excluded. These areas include other smart contract uses, performance evaluation, and the type of vocabulary standard such as Systematized Nomenclature of Medicine Clinical Terms, 10th revision of the International Classification of Diseases, and Logical Observation Identifiers Names and Codes. We also recognize that greater detail about the read and write models could be studied, such as their validity periods and whether other stakeholders (eg, researchers and insurance companies) were given access. Future reviews should consider delving deeper into these areas.

Furthermore, despite our best efforts to capture as much material available as possible, we are aware that the exclusion of articles whose full text was not in English would have limited the scope of this review. In addition, there may also be other developments in this space that have not been made publicly available for commercial or other reasons.

Conclusions

This cross-disciplinary systematic review on the blockchain PHR space has revealed that as of now, much of the development is still in the conceptual stage. However, there is a trend of growth and maturation. We believe that this provides

consolidated evidence for researchers to continue following this space and, more optimistically, to spur them to contribute ideas and efforts to accelerate its development. Those in the medical informatics community will undoubtedly play an increasingly larger role in the development and implementation of blockchain PHRs, especially when the need to integrate with EMR systems and adopt health care data standards becomes more prominent. In addition, as the first systematic review covering blockchain PHRs, we expect this to be an important basis for subsequent reviews to track how the space has progressed in the future.

Acknowledgments

The authors would like to thank their families and work organizations—SingHealth Polyclinics, SingHealth Community Hospitals, and the National University Hospital—for their support.

Authors' Contributions

AF had the original idea for this study. AF, MT, and CT designed the review. MT obtained the relevant articles. AF, TT, and MT reviewed the articles and abstracted the data. AF wrote the first draft of this paper, and all the authors subsequently assisted in revising the work and have approved the final version.

Conflicts of Interest

The authors did not receive any funding for this work and declare no conflicts of interest. However, the authors would like to highlight that MT is the cofounder of MediLOT, which published one of the articles reviewed in this paper.

Multimedia Appendix 1
List of articles with data abstracted.
[XLSX File (Microsoft Excel File), 29 KB - jmir_v23i4e25094_app1.xlsx]

Multimedia Appendix 2
Overview of the articles with the article identifier, publication year, publisher, type of article, country and interest group.
[DOCX File , 15 KB - jmir_v23i4e25094_app2.docx]

References

1. Tang PC, Ash JS, Bates DW, Overhage JM, Sands DZ. Personal health records: definitions, benefits, and strategies for overcoming barriers to adoption. J Am Med Inform Assoc 2006 Mar;13(2):121-126 [FREE Full text] [doi: 10.1197/jamia.M2025] [Medline: 16357345]
2. Price M, Bellwood P, Kitson N, Davies I, Weber J, Lau F. Conditions potentially sensitive to a personal health record (PHR) intervention, a systematic review. BMC Med Inform Decis Mak 2015;15:32 [FREE Full text] [doi: 10.1186/s12911-015-0159-1] [Medline: 25927384]
3. Andrikopoulou E, Scott P, Herrera H, Good A. What are the important design features of personal health records to improve medication adherence for patients with long-term conditions? A systematic literature review. BMJ Open 2019 Sep 26;9(9):e028628. [doi: 10.1136/bmjopen-2018-028628]
4. Dameff C, Clay B, Longhurst CA. Personal Health Records: More Promising in the Smartphone Era? JAMA 2019 Jan 29;321(4):339-340. [doi: 10.1001/jama.2018.20434] [Medline: 30633300]
5. Alanazi A, Anazi YA. The Challenges in Personal Health Record Adoption. Journal of Healthcare Management 2019;64(2):104-109. [doi: 10.1097/jhm-d-17-00191]
6. Hossain MM, Hong YA. Trends and characteristics of protected health information breaches in the United States. 2019 Presented at: AMIA Annu Symp Proc. 2020; 2019; Washington p. 1081-1090.
7. Abd-alrazaq AA, Bewick BM, Farragher T, Gardner P. Factors that affect the use of electronic personal health records among patients: A systematic review. International Journal of Medical Informatics 2019 Jun;126:164-175. [doi: 10.1016/j.ijmedinf.2019.03.014]
8. Nakamoto S. Bitcoin: A Peer-to-Peer Electronic Cash System. SSRN Journal 2008:1-9. [doi: 10.2139/ssrn.3440802]
9. Kane E. Is Blockchain a General Purpose Technology? SSRN Journal 2017 Mar 11:1-27. [doi: 10.2139/ssrn.2932585]
10. Tapscott D, Tapscott A. Blockchain Revolution: How the Technology Behind Bitcoin and Other Cryptocurrencies Is Changing the World. Blockchain Revolution: Penguin Random House; 2018.
11. Agbo C, Mahmoud Q, Eklund J. Blockchain Technology in Healthcare: A Systematic Review. Healthcare 2019 Apr 04;7(2):56. [doi: 10.3390/healthcare7020056]
12. Chukwu E, Garg L. A Systematic Review of Blockchain in Healthcare: Frameworks, Prototypes, and Implementations. IEEE Access 2020;8:21196-21214. [doi: 10.1109/access.2020.2969881]
13. Drosatos G, Kaldoudi E. Blockchain Applications in the Biomedical Domain: A Scoping Review. Computational and Structural Biotechnology Journal 2019;17:229-240. [doi: 10.1016/j.csbj.2019.01.010]

14. Hasselgren A, Kralevska K, Gligoroski D, Pedersen SA, Faxvaag A. Blockchain in healthcare and health sciences—A scoping review. International Journal of Medical Informatics 2020 Feb;134:104040. [doi: 10.1016/j.ijmedinf.2019.104040]
15. Adelmeyer M, Meier P, Teuteberg F. Security and Privacy of Personal Health Records in Cloud Computing Environments - An Experimental Exploration of the Impact of Storage Solutions and Data Breaches. In: Security and Privacy of Personal Health Records in Cloud Computing Environments - An Experimental Exploration of the Impact of Storage Solutions and Data Breaches.: AIS Electronic Library; 2019 Presented at: Wirtschaftsinformatik 2019; 23-Feb-2019 to 29-Feb-2019; Siegen.
16. Xuan S, Zheng L, Chung I, Wang W, Man D, Du X, et al. An incentive mechanism for data sharing based on blockchain with smart contracts. Computers & Electrical Engineering 2020 May;83:106587. [doi: 10.1016/j.compeleceng.2020.106587]
17. Park YR, Lee E, Na W, Park S, Lee Y, Lee J. Is Blockchain Technology Suitable for Managing Personal Health Records? Mixed-Methods Study to Test Feasibility. J Med Internet Res 2019 Feb 08;21(2):e12533 [FREE Full text] [doi: 10.2196/12533] [Medline: 30735142]
18. Roehrs A, da Costa CA, Righi RR, de Oliveira KS. Personal health records: a systematic literature review. J Med Internet Res 2017 Jan 06;19(1):e13 [FREE Full text] [doi: 10.2196/jmir.5876] [Medline: 28062391]
19. Lee H, Kung H, Udayasankaran JG, Kijsanayotin B, B Marcelo A, Chao LR, et al. An Architecture and Management Platform for Blockchain-Based Personal Health Record Exchange: Development and Usability Study. J Med Internet Res 2020 Jun 09;22(6):e16748 [FREE Full text] [doi: 10.2196/16748] [Medline: 32515743]
20. Azaria A, Ekblaw A, Vieira T, Lippman A. MedRec: Using Blockchain for Medical Data Access and Permission Management. MedRec: Using Blockchain for Medical Data Access and Permission Management. Innd International Conference on Open and Big Data (OBD); 2016 Presented at: International Conference on Open and Big Data; 22-24 Aug 2016; Vienna p. 2016-2012. [doi: 10.1109/obd.2016.11]
21. A M, C C, R R. Electronic health records in a Blockchain: A systematic review. Electronic health records in a Blockchain: A systematic review 2020:1-16 [FREE Full text] [doi: 10.1177/1460458219866350]
22. Shuaib, Saleous, Shuaib, Zaki. Blockchains for Secure Digitized Medicine. JPM 2019 Jul 13;9(3):35. [doi: 10.3390/jpm9030035]
23. Vazirani AA, O'Donoghue O, Brindley D, Meinert E. Implementing Blockchains for Efficient Health Care: Systematic Review. J Med Internet Res 2019 Feb 12;21(2):e12439 [FREE Full text] [doi: 10.2196/12439] [Medline: 30747714]
24. Angeles R. Blockchain-Based Healthcare: Three Successful Proof-of-Concept Pilots Worth Considering. Blockchain-Based Healthcare: Three Successful Proof-of-Concept Pilots Worth Considering 2018;27(3):38.
25. Rijmenam DV. How Blockchain Will Give Consumers Ownership of their Data. Medium. URL: https://medium.com/@markvanrijmenam/how-blockchain-will-give-consumers-ownership-of-their-data-3e90020107e6 [accessed 2020-07-05]
26. Dickson B. How blockchain solves the complicated data-ownership problem. The Next Web. URL: https://thenextweb.com/contributors/2017/08/17/blockchain-solves-complicated-data-ownership-problem/ [accessed 2020-07-05]
27. Ethereum W. Ethereum Whitepaper. Ethereum. URL: https://ethereum.org [accessed 2020-07-07]
28. Zheng Z, Xie S, Dai H, Chen X, Wang H. An Overview of Blockchain Technology: Architecture, Consensus, and Future Trends. In: IEEE International Congress on Big Data (BigData Congress). An Overview of Blockchain Technology: Architecture, Consensus, and Future Trends. InIEEE International Congress on Big Data (BigData Congress); 2017 Presented at: IEEE International Congress on Big Data (BigData Congress); 25-30 Jun 2017; Honolulu p. A. [doi: 10.1109/bigdatacongress.2017.85]
29. Hyperledger. Hyperledger Whitepaper. Hyperledger project. URL: https://www.hyperledger.org/ [accessed 2020-07-07]
30. Quorum Whitepaper. Quorum project. URL: https://www.goquorum.com/ [accessed 2020-07-07]
31. Corda |OSBPFB. Corda Whitepaper. Corda project. URL: https://www.corda.net/ [accessed 2020-07-07]
32. Banks MA. Sizing up big data. Nat Med 2020 Jan 13;26(1):5-6. [doi: 10.1038/s41591-019-0703-0]
33. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. J Clin Epidemiol 2009 Oct;62(10):1006-1012. [doi: 10.1016/j.jclinepi.2009.06.005] [Medline: 19631508]
34. Hölbl M, Kompara M, Kamišalić A, Nemec Zlatolas L. A Systematic Review of the Use of Blockchain in Healthcare. Symmetry 2018 Oct 10;10(10):470. [doi: 10.3390/sym10100470]
35. Kuo T, Zavaleta Rojas H, Ohno-Machado L. Comparison of blockchain platforms: a systematic review and healthcare examples. J Am Med Inform Assoc 2019 May 01;26(5):462-478. [doi: 10.1093/jamia/ocy185] [Medline: 30907419]
36. Keutzer K, Newton A, Rabaey J, Sangiovanni-Vincentelli A. System-level design: orthogonalization of concerns and platform-based design. IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst 2020;19(12):1523-1543. [doi: 10.1109/43.898830]
37. Burniske C. How blockchain technology can enhance EHR operability. 3iQ. 2016. URL: https://3iq.ca/how-blockchain-technology-can-enhance-electronic-health-record-ehr-operability/ [accessed 2021-04-01]
38. McFarlane C, Beer M, Brown J, Prendergast N. Patientory: A Healthcare Peer-to-Peer EMR Storage Network v1.0. Patientory. URL: [accessed 2021-04-01]
39. Roehrs A, da Costa CA, da Rosa Righi R. OmniPHR: a distributed architecture model to integrate personal health records. J Biomed Inform 2017 Jul;71:70-81 [FREE Full text] [doi: 10.1016/j.jbi.2017.05.012] [Medline: 28545835]

40. Badr S, Gomaa I, Abd-Elrahman E. Multi-tier Blockchain Framework for IoT-EHRs Systems. Procedia Computer Science 2018;141:159-166. [doi: 10.1016/j.procs.2018.10.162]
41. Boiani F. Blockchain Based Electronic Health Record Management For Mass Crisis Scenarios?: A Feasibility Study. undefined. Published 2018. Accessed July 14, 2020. /paper/Blockchain-Based-Electronic-Health-Record-For-Mass-Boiani/1218ed976d8fb5ad33096933defc3ce87cc0 2018:6488.
42. Chen Y, Ding S, Xu Z, Zheng H, Yang S. Blockchain-Based Medical Records Secure Storage and Medical Service Framework. J Med Syst 2018 Nov 22;43(1):1. [doi: 10.1007/s10916-018-1121-4]
43. Dagher GG, Mohler J, Milojkovic M, Marella PB. Ancile: Privacy-preserving framework for access control and interoperability of electronic health records using blockchain technology. Sustainable Cities and Society 2018 May;39:283-297. [doi: 10.1016/j.scs.2018.02.014]
44. Dubovitskaya A, Xu Z, Ryu S, Schumacher M, Wang F. Secure and Trustable Electronic Medical Records Sharing using Blockchain. 2018 Presented at: AMIA Annu Symp Proc. 2018; 2018; San Francisco p. 650-659.
45. Gebremedhin T. Blockchain as a Technology to Facilitate Privacy and Better Health Record Management. Published online September 1. 2018. URL: https://bora.uib.no/handle/1956/19621 [accessed 2020-07-14]
46. MedRec: Patient Control of Medical Record Distribution - IEEE Blockchain Initiative. IEEE Blockchain Initiative. URL: https://blockchain.ieee.org/technicalbriefs/july-2018/medrec-patient-control-of-medical-record-distribution [accessed 2020-07-14]
47. Medicalchain W. Medicalchain Whitepaper. Medicalchain. URL: https://medicalchain.com/en/whitepaper/ [accessed 2020-07-14]
48. Rouhani S, Butterworth L, Simmons A, Humphery D, Deters R. MediChainTM: A Secure Decentralized Medical Data Asset Management System. 2018 Presented at: IEEE Int Conf Internet Things; 2018; Halifax p. A. [doi: 10.1109/cybermatics_2018.2018.00258]
49. Thwin T, Vasupongayya S. Blockchain Based Secret-Data Sharing Model for Personal Health Record System. Inth International Conference on Advanced Informatics: Concept Theory and Applications (ICAICTA); 2018 Presented at: 5th International Conference on Advanced Informatics: Concept Theory and Applications (ICAICTA); 2018; Thailand p. 2018-2015. [doi: 10.1109/icaicta.2018.8541296]
50. Vora J, Nayyar A, Tanwar S. BHEEM: A Blockchain-Based Framework for Securing Electronic Health Records. In: IEEE.: IEEE; 2018 Presented at: 2018 IEEE Globecom Workshops; 2018; Abu Dhabi p. A. [doi: 10.1109/glocomw.2018.8644088]
51. Zhang X, Poslad S. Blockchain Support for Flexible Queries with Granular Access Control to Electronic Medical Records (EMR). 2018 Presented at: IEEE International Conference on Communications (ICC); 2018; Kansas City p. 2018. [doi: 10.1109/icc.2018.8422883]
52. Abouzahra M. Using Blockchain Technology to Enhance the Use of Personal Health Records. In: AMCIS Proc. 2019 Presented at: AMCIS; 2019; Cancun p. 1 URL: https://aisel.aisnet.org/amcis2019/healthcare_it/healthcare_it/6
53. Alkhushayni S, Al-Zaleq D, Kengne N. Blockchain Technology applied to Electronic Health Records. EasyChair 2019:34-42. [doi: 10.29007/2x3r]
54. Ray CD. Patient Privacy Ownership of Electronic Health Records on a Blockchain. In: Lecture Notes in Computer Science. New York: Springer; 2019:95-111.
55. Ciampi M. A blockchain architecture for the Italian EHR system. In: ThinkMind Digital Library. 2019 Presented at: The Fourth International Conference on Informatics and Assistive Technologies for Health-Care, Medical Support and Wellbeing (HEALTHINFO 2019); 2019; Valencia URL: https://www.researchgate.net/publication/337946548_A_blockchain_architecture_for_the_Italian_EHR_system
56. Daraghmi E, Daraghmi Y, Yuan S. MedChain: A Design of Blockchain-Based System for Medical Records Access and Permissions Management. IEEE Access 2019;7:164595-164613. [doi: 10.1109/access.2019.2952942]
57. Donawa A, Orukari I, Baker C. Scaling Blockchains to Support Electronic Health Records for Hospital Systems. In: IEEE. 2019 Presented at: IEEE 10th Annual Ubiquitous Computing, Electronics Mobile Communication Conference (UEMCON); 2019; New York p. 2019. [doi: 10.1109/uemcon47517.2019.8993101]
58. Hang, Choi, Kim. A Novel EMR Integrity Management Based on a Medical Blockchain Platform in Hospital. Electronics 2019 Apr 25;8(4):467. [doi: 10.3390/electronics8040467]
59. Harika M, Rahmadika S, Ramdania DR. Blockchain technology for managing an architectural model of decentralized medical record. In: J. Phys.: Conf. Ser. 2019 Dec 14 Presented at: J Phys Conf Ser; 2019; Medan p. 077027. [doi: 10.1088/1742-6596/1402/7/077027]
60. Huang J, Qi Y, Asghar M, Meads A. MedBloc: A Blockchain-Based Secure EHR System for Sharing and Accessing Medical Data. In: Tu Y-C. MedBloc: A Blockchain-Based Secure EHR System for Sharing and Accessing Medical Data. Inth IEEE International Conference On Trust, Security And Privacy In Computing And Communications/13th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE); 2019 Presented at: 18th IEEE International Conference On Trust, Security And Privacy In Computing And Communications; 2019; New Zealand. [doi: 10.1109/trustcom/bigdatase.2019.00085]
61. Hylock RH, Zeng X. A Blockchain Framework for Patient-Centered Health Records and Exchange (HealthChain): Evaluation and Proof-of-Concept Study. J Med Internet Res 2019 Aug 31;21(8):e13592. [doi: 10.2196/13592]

62. Jiang S, Wu H, Wang L. Patients-Controlled Secure and Privacy-Preserving EHRs Sharing Scheme Based on Consortium Blockchain. In: IEEE. 2019 Presented at: IEEE Global Communications Conference (GLOBECOM); 2019; Hawaii. [doi: 10.1109/globecom38437.2019.9013220]
63. Koushik A, Jain B, Menon N, Lohia D, Chaudhari SB. Performance Analysis of BlockChain-based Medical Records Management System. 2019 Presented at: 4th International Conference on Recent Trends on Electronics, Information, Communication Technology (RTEICT); 2019; Bengaluru p. 2019-2014. [doi: 10.1109/rteict46194.2019.9016812]
64. Lee SY. PHR System using Blockchain Technology. International Journal of Advanced Trends in Computer Science and Engineering 2019:1-6 [FREE Full text] [doi: 10.30534/ijatcse/2019/84862019]
65. Medibloc. Medibloc Technical Whitepaper. MediBloc. 2019. URL: https://github.com/medibloc/whitepaper [accessed 2020-07-15]
66. MediLOT | Transforming Healthcare For All. MediLOT. URL: https://medilot.com/ [accessed 2020-07-15]
67. Nchinda N, Cameron A, Retzepi K, Lippman A. MedRec: A Network for Personal Information Distribution. MedRec: A Network for Personal Information Distribution. InInternational Conference on Computing, Networking and Communications (ICNC); 2019 Presented at: International Conference on Computing, Networking and Communications (ICNC); 2019; Hawaii p. A. [doi: 10.1109/iccnc.2019.8685631]
68. Nguyen DC, Pathirana PN, Ding M, Seneviratne A. Blockchain for Secure EHRs Sharing of Mobile Cloud Based E-Health Systems. IEEE Access 2019;7:66792-66806. [doi: 10.1109/access.2019.2917555]
69. Rajput AR, Li Q, Taleby Ahvanooey M, Masood I. EACMS: Emergency Access Control Management System for Personal Health Record Based on Blockchain. IEEE Access 2019;7:84304-84317. [doi: 10.1109/access.2019.2917976]
70. Reen G, Mohandas M, Venkatesan S. Decentralized Patient Centric e- Health Record Management System using Blockchain and IPFS. In: IEEE. 2019 Presented at: InIEEE Conference on Information and Communication Technology; 2019; Pakistan p. 2019. [doi: 10.1109/cict48419.2019.9066212]
71. Sangeetha R. Electronic Health Record System using Blockchain. Int Res J Multidiscip Technovation 2019 Mar 25;1(2):57-61. [doi: 10.34256/irjmt1927]
72. Shahnaz A, Qamar U, Khalid A. Using Blockchain for Electronic Health Records. IEEE Access 2019;7:147782-147795. [doi: 10.1109/access.2019.2946373]
73. Shekhawat G. Cloud-chain: Revamp Health Record System Using Blockchain. Website. 2019. URL: [accessed 2020-07-14]
74. Thwin T, Vasupongayya S. Blockchain-Based Access Control Model to Preserve Privacy for Personal Health Record Systems. Security and Communication Networks 2019:1-16 [FREE Full text] [doi: 10.1155/2019/8315614]
75. Tian X. Blockchain-based secure medical record sharing system. Website. 2019. URL: http://eprints.utar.edu.my/3435/ [accessed 2020-07-15]
76. Toshniwal B, Podili P, Reddy R, Kataoka K. PACEX: PAtient-Centric EMR eXchange in Healthcare Systems using Blockchain. In: IEEE. PACEX: PAtient-Centric EMR eXchange in Healthcare Systems using Blockchain. InIEEE 10th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON); 2019 Presented at: IEEE 10th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON); 2019; Vancouver p. 2019. [doi: 10.1109/iemcon.2019.8936258]
77. Wang S, Zhang D, Zhang Y. Blockchain-Based Personal Health Records Sharing Scheme With Data Integrity Verifiable. IEEE Access 2019;7:102887-102901. [doi: 10.1109/access.2019.2931531]
78. Wang Y, Zhang A, Zhang P, Wang H. Cloud-Assisted EHR Sharing With Security and Privacy Preservation via Consortium Blockchain. IEEE Access 2019;7:136704-136719. [doi: 10.1109/access.2019.2943153]
79. Sihua W, Jang D. Electronic medical record security sharing model based on blockchain. In: Proceedings of the 3rd International Conference on Cryptography, Security and Privacy. 2019 Presented at: ICCSP '19; 2019; Japan. [doi: 10.1145/3309074.3309079]
80. Goni N, Ahmed S, Ibrahim A. A P2P Optimistic Fair-Exchange (OFE) Scheme for Personal Health Records Using Blockchain Technology. In: Springer.: Springer; 2020 Presented at: 3rd International Conference on Wireless, Intelligent and Distributed Environment for Communication; 2020; Toronto p. 1-21. [doi: 10.1007/978-3-030-44372-6_1]
81. Arunkumar B, Kousalya G. Blockchain-Based DecentralizedSecure Lightweight E-Health System for Electronic Health Records. In: Intelligent Systems, Technologies and Applications. New York: Springer; 2020:273-289.
82. Aswin A, Basil K, Viswan V, Reji B, Kuriakose B. Design of AYUSH: A Blockchain-Based Health Record Management System. In: Inventive Communication and Computational Technologies. Singapore: Springer Singapore; 2020.
83. Cao Y, Sun Y, Min J. Hybrid blockchain-based privacy-preserving electronic medical records sharing scheme across medical information control system. Measurement and Control 2020:1-14 [FREE Full text] [doi: 10.1177/0020294020926636]
84. Charanya RR. SeFra: A Secure Framework to Manage eHealth Records Using Blockchain Technology. International Journal of E-Health and Medical Communications (IJEHMC) 2020:1-5. [doi: 10.4018/ijehmc.2020010101]
85. Kavathekar S, Patil R. Data Sharing and Privacy Preserving of Medical Records Using Blockchain. In: Lecture Notes on Data Engineering and Communications Technologies. New York: Springer; 2020:65-72.
86. Kim M, Yu S, Lee J, Park Y, Park Y. Design of Secure Protocol for Cloud-Assisted Electronic Health Record System Using Blockchain. Sensors 2020 May 21;20(10):2913. [doi: 10.3390/s20102913]

87. Kung H, Cheng Y, Lee H, Hsu C. Personal Health Record in FHIR Format Based on Blockchain Architecture. In: Lecture Notes in Electrical Engineering. New York: Springer; 2020:1776-1788.
88. Sharma S, Mishra A, lala A, Singhai D. Secure Cloud Storage Architecture for Digital Medical Record in Cloud Environment Using Blockchain. SSRN Journal 2020:1-5. [doi: 10.2139/ssrn.3565922]
89. Sharma Y, Balamurugan B. Preserving the privacy of electronic health records using blockchain. Procedia Comput Sci 2020;173(1):171-180. [doi: 10.1016/j.procs.2020.06.021]
90. Tith D, Lee JS, Suzuki H, Wijesundara WM, Taira N, Obi T, et al. Application of blockchain to maintaining patient records in electronic health record for enhanced privacy, scalability, and availability. Healthc Inform Res 2020 Jan;26(1):3-12 [FREE Full text] [doi: 10.4258/hir.2020.26.1.3] [Medline: 32082695]
91. Verdonck M, Poels G. Architecture and Value Analysis of a Blockchain-Based Electronic Health Record Permission Management System. In: Proceedings of 14th International Workshop on Value Modelling and Business Ontologies.: Ghent University; 2020 Presented at: 14th International Workshop on Value Modelling and Business Ontologies; Jan 16, 2020; Brussels, Belgium URL: https://biblio.ugent.be/publication/8654183
92. Wu X, Han Y, Zhang M, Zhu S. Secure personal health records sharing based on blockchain and IPFS. In: Trusted Computing and Information Security. Communications in Computer and Information Science. New York: Springer; 2020.
93. Layer. Ethereum. 2020. URL: https://ethereum.org
94. Neo Smart Economy. NEO team. URL: https://neo.org/ [accessed 2020-07-15]
95. NeoFS data storage, payment, and auditing demoed at Consensus 2019. Neo News Today. 2020. URL: https://neonewstoday.com/development/neofs-data-storage-payment-and-auditing-demoed-at-consensus-2019/ [accessed 2020-07-15]
96. China Bans Digital Coin Offers as Celebrities Like Paris Hilton Tout Them. Walls Street Journal. URL: https://www.wsj.com/articles/some-digital-coins-are-up-2-800-what-could-go-wrong-1504522801 [accessed 2020-12-24]
97. Kim C. South Korea bans raising money through initial coin offerings. Reuters. URL: https://www.reuters.com/article/us-southkorea-bitcoin-idUSKCN1C408N.Published September 29, 2017 [accessed 2020-12-24]
98. SEC Issues Investigative Report Concluding DAO Tokens, a Digital Asset, Were Securities. SEC.hov. 2020. URL: https://www.sec.gov/news/press-release/2017-131 [accessed 2021-04-01]

Abbreviations

BFT: Byzanthine fault tolerance
CS: computer science
EHR: electronic health record
EMR: electronic medical record
FHIR: Fast Healthcare Interoperability Resource
GDPR: General Data Protection Regulation
HF: Hyperledger Fabric
HIPAA: Health Insurance Portability and Accountability Act
HL-7: health level 7
ICO: initial coin offering
IPFS: Interplanetary File System
NeoFS: NEO file storage system
PHR: personal health record
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-analyses
PoW: proof of work
UI: user interface

Edited by R Kukafka; submitted 17.10.20; peer-reviewed by CA da Costa, TT Kuo, JH Lee; comments to author 21.12.20; revised version received 25.12.20; accepted 16.03.21; published 13.04.21.

Please cite as:
Fang HSA, Tan TH, Tan YFC, Tan CJM
Blockchain Personal Health Records: Systematic Review
J Med Internet Res 2021;23(4):e25094
URL: https://www.jmir.org/2021/4/e25094
doi:10.2196/25094
PMID:33847591

Original Paper

Geosocial Networking Dating App Usage and Risky Sexual Behavior in Young Adults Attending a Music Festival: Cross-sectional Questionnaire Study

Shirali Garga[1], MBBS; Meryl Thomas[1], MBBS; Ashneet Bhatia[1], MBBS; Aidan Sullivan[1], MBBS; Franklin John-Leader[2], BSc, MSW, MA; Sabrina Pit[1,3,4], Dip(OHS), MSc, PhD

[1]University Centre for Rural Health, Western Sydney University, Lismore, Australia

[2]Harm Reduction and Health Promotion Programs, HIV and Related Programs (HARP), North Coast Public Health, Mid-North Coast Local Health District, Lismore, Australia

[3]University Centre for Rural Health, University of Sydney, Lismore, Australia

[4]NSW Rural Doctors Network, Newcastle, Australia

Corresponding Author:
Sabrina Pit, Dip(OHS), MSc, PhD
University Centre for Rural Health
Western Sydney University
62 Uralba Street
Lismore, 2480
Australia
Phone: 61 429455720
Email: S.Pit@westernsydney.edu.au

Abstract

Background: Despite the prevalent use of geosocial networking dating apps (GNDAs), there is limited research on their impact on sexual health outcomes among young music festivals attendees.

Objective: This study aims to explore the use of GNDAs and risky sexual behaviors of young adults attending a music festival.

Methods: The music festival attendees (N=862) completed a cross-sectional questionnaire study encompassing demographics, dating app use, and risky sexual behaviors in the past year. Associations between these variables were estimated using bivariate and multivariate logistic regression analyses.

Results: Of the respondents, 51.9% (448/862) had used GNDAs in the previous year. Compared with people who had 1 partner, people who had 2-5 sexual partners in the previous year had almost 7 times the odds of using dating apps (odds ratio [OR] 6.581, 95% CI 4.643-9.328) and those who had more than 5 partners had 14 times the odds of using dating apps (OR 14.294, 95% CI 8.92-22.906). Condom users were more likely to be app users (P<.001), as were those who relied on emergency Plan B (P=.002), but people using hormonal contraception were less likely to use dating apps (P=.004). After adjusting for sexual orientation and relationship status, those having casual sex had 3.096 (95% CI 2.225-4.307; P<.001) times the odds of using dating apps and those having multiple sexual partners had 3.943 (95% CI 2.782-5.588; P<.001) times the odds of using dating apps. Similarly, after adjusting for sexual orientation, relationship status, and number of sexual partners, people who had no discussions before having sex about sexually transmitted infections (STIs) or boundaries were more likely to use dating apps (OR 1.755, 95% CI 1.232-2.500; P=.002). Those who perceived the risk of having sex without contraception to be *very high* had 2.486 (95% CI 2.213-5.096; P=.01) times the odds of using dating apps than those who perceived *no risk*. Compared with those who perceived no risk, people who thought that the risk of having multiple sexual partners was *low* to *high* had 1.871 (95% CI 1.024-3.418; P=.04) times the odds of using dating apps. A significant number of app users (389/440, 88.4%) indicated that GNDAs should promote safe sex.

Conclusions: This study identified that festival goers engaging in certain high-risk sexual behaviors, including casual sex, having multiple sexual partners, and having sex without discussion about STI status and boundaries, are more likely to use dating apps. Festival goers who perceived sex without any form of contraception, having sex while drunk, and having multiple sexual partners as risky were more likely to be app users. Policy makers and GNDA developers should acknowledge the vulnerability of their users to adverse sexual health outcomes and use GNDAs as a platform to promote risk-reduction practices.

(J Med Internet Res 2021;23(4):e21082) doi:10.2196/21082

KEYWORDS

sexual health; mobile apps; young adults; music festival

Introduction

Background

Geosocial networking dating apps (GNDAs) provide users with a web-based platform to make social, romantic, or sexual connections. These apps connect users to potential partners based on their location. Such platforms have become increasingly popular since the launch of Tinder in 2012, which has grown to over 10 million users per day, with similar apps (Bumble and OkCupid) joining the market [1]. In Australia, web-based dating has become the second most preferred way to meet new partners after introductions from family and friends, surpassing more traditional methods of meeting [2]. A Dutch study on the primary motivations for Tinder use showed that in addition to dating, GNDAs may be used to ease communication and obtain self-worth or validation and for excitement or trendiness [1].

Young adults, the primary users of GNDAs [3], have a higher tendency toward risky sexual behaviors. For example, they often lack safe-sex discussions before intercourse [4], a significant risk factor for sexually transmitted infections (STIs). Consequently, in Australia, young adults (15-24 years) account for 50% of newly acquired STIs [5].

GNDA use has been directly associated with risky sexual behaviors. A meta-analysis concluded that GNDA users were more likely to contract STIs [6]. Similarly, Shapiro et al [7] found that Tinder users were more likely to report 5 or more previous sexual partners and a positive STI screen. This is concerning, given the high rate of GNDA use and promotion at music festivals [8]. Furthermore, a recent review of 99 studies on GNDAs and sexual health identified that the most common theme was risky sexual behaviors [9] and tended to focus on the lesbian, gay, bisexual, transgender, and queer (LGBTQ+) community [9].

Music festival attendees have also demonstrated significant risky sexual behaviors. An Australian study found that half of the music festival attendees reported taking illicit drugs in the last 12 months and had engaged in risky sexual behaviors such as sex without condoms, casual sex, and a lack of STI screening [10]. The environment of a music festival is particularly risky for dating app use because of the proximity of hundreds of young people, many of whom use dating apps and do not practice safe sex. Therefore, the proximity of congregated youth creates an environment conducive for increased app use and, in turn, greater potential for risky sexual behavior. There is evidence to suggest that GNDAs can be used for safe-sex health promotion, and several reviews have identified the potential effectiveness of mobile phone (or app based) interventions on delivering safe-sex information among young people [11,12]. Despite this potential, a 2016 study investigated 60 dating apps to determine whether they included any sexual health content. Huang et al [13] found that only 9 dating apps had sexual health content and 7 of these only targeted men having sex with men.

Objective

In summary, GNDA use is prevalent among young adults, and festival goers are a higher risk group, but the association between sexual health and dating app use among young people attending festivals has, to our knowledge, not been investigated yet. Therefore, this study aims to explore GNDA use and risky sexual behavior among young adults attending a music festival.

Methods

Study Design

This study was a cross-sectional survey. People aged 18 years and above attending a music festival in 2019 in New South Wales, Australia, were eligible to participate in the study.

Data Collection

Data were collected at a major 3-day music festival in 2019 in New South Wales, Australia. Recruitment was conducted by 5 of the authors who were on-site during the festival. Attendees who approached a permanent health promotion stall within the campgrounds were invited to take part and people within the target demographic were approached. Those approached were screened by age and given a participant information survey. The information sheet outlined the study aims, what they would be asked to do, the benefits of taking part and potential downsides, how the study was paid for, data storage, an explanation about study withdrawal, and how to contact the ethics committee or researchers. If willing to participate, paper surveys were distributed, completed anonymously, and placed in closed boxes for confidentiality. Survey completion was obtained as consent. Owing to the size, timing, and location of the festival, the data were collected via convenience sampling. The number of individuals who declined to participate was not recorded. No visibly intoxicated people were allowed to complete the study, and no incentives were provided. The data were entered into a Microsoft Excel spreadsheet.

Survey Development

The questionnaire development was guided by public health and sexual health experts and included questions about demographics, sexual behavior, risk perception, and dating apps. Risky sexual behaviors were selected from the Safer Sex Behavior Questionnaire (SSBQ) by Dorio et al [14] and were used to develop questions in relation to dating apps and sexual health. Akin to the SSBQ, the questionnaire covered condom use, impact of intravenous drug use history in a sexual partner, drinking alcohol before sex, and discussion about safe sex with a partner before sexual activity. The SSBQ covered more risky sexual behaviors, such as engagement in anal intercourse and sexual intercourse on the first date. However, these were excluded to reduce survey fatigue, as it was necessary to keep

the survey relatively brief. The survey was piloted (n=13) with university students, representing the target population and refined before being approved by the Western Sydney University Human Research Ethics Committee (H11327). The survey questions used in this study are listed in Multimedia Appendix 1.

Outcome Measures

GNDA use and nonuse were determined by asking whether they had used a dating app in the last 12 months. Dating app users were also asked: "Should dating apps share educational messages about safe sex?"

The survey asked about reasons for use and sexual behavior with partners meeting off GNDAs. However, data were not captured in this study as it did not allow the comparison of users versus nonusers.

All participants were asked about their age, gender, sexual orientation, relationship, whether they had engaged in sexual activity in the last 12 months, and the number of sexual partners in the last 12 months. The survey explained the definition of *sexual activity* to be defined as oral and/or penetrative sex.

Participants were also asked what forms of contraception they used regularly, including *none*; *condoms*; *Plan B (morning-after pill)*; *the pill*; *intrauterine devices(IUDs), such as Mirena and Paragard*; *spermicide*; *Implanon—the rod;pulling out method*; or other methods. These were then categorized as follows for further analyses: *no contraception at all*, *condom use*, *Plan B*, *pulling out*, and *hormonal contraception*, with the latter including *the pill*, *IUD (Mirena and Paragard)*, *spermicide*, and *Implanon—the rod*.

With respect to sexual health behavior, participants were first asked about their actual behavior and then about their risk perception of this behavior. Participants were asked if they had engaged in the following 8 behaviors in the last 12 months: sex without any form of contraception; sex without condom but with other contraception, for example, the pill; casual sex; having sex while drunk; lack of discussion about STI status and sexual boundaries before sexual activity; having multiple sexual partners; having unprotected sex with a partner who has ever injected drugs; and finally having sex with a partner who has an STI. For each behavior, participants were asked to rate their perception of sexual risk on a 5-point Likert scale, ranging from no risk to very high risk.

Data Analysis

Data were analyzed using SPSS, version 25. Bivariate analysis was used to compare categorical demographics, contraception, specific risky sexual behaviors, and related risk perceptions with dating app use. Statistical significance was defined as P=.05. Logistic regression analyses were used to estimate crude odds ratios (ORs) to determine the factors associated with using GNDAs. Contraceptive and specific sexual behavioral variables that were statistically significantly associated with GNDA use in bivariate analyses were then analyzed in a multivariate model, which was adjusted for sexual orientation, relationship status, and number of sexual partners.

Results

Demographics

About half (448/862, 51.9%) of the participants had used GNDAs within the past 12 months. A significant number of dating app users (389/440, 88.4%) indicated that GNDAs should promote safe sex. The study (N=862) was predominantly completed by those between the ages of 21 and 24 years (395/862, 45.8%), closely followed by those aged between 18 and 21 years (383/862, 44.4%; Table 1). The highest proportion of respondents were female (558/860, 64.9%), heterosexual (770/862, 89.3%), and sexually active (813/862, 94.3%). The majority were in exclusive relationships (362/858, 42.2%), whereas the rest were single and not dating (310/858, 36.1%) or casually dating or in open relationships (186/858, 21.7%). Of the sexually active participants (n=812), most had 2 to 5 sexual partners within the past 12 months (341/812, 41.9%), followed by those who had one sexual partner (311/812, 38.3%).

People identified as LGBTQ+ had almost twice the odds (OR 1.846, 95% CI 1.175-2.900; P=.008) of using GNDAs as compared with their heterosexual counterparts. Participants in exclusive relationships had a lower proportion of app usage (109/362, 30.1%) compared with casual daters (141/186, 75.8%) and singles (195/310, 62.9%). Casual daters were 7.273 times (95% CI 4.857-10.891; P<.001) more likely to have used GNDAs within 12 months compared with those in exclusive relationships. Participants who were *single and not dating* were 3.936 times (95% CI 2.853-5.430; P=.003) more likely to use GNDAs.

People who had more sexual partners were more likely to be dating app users. Of the participants with 1 partner, only 21.9% (68/311) were GNDA users, compared with 64.8% (221/341) of those with 2 to 5 partners and 80% (128/160) of those with more than 5 partners. Compared with those with only 1 partner in the last 12 months, participants with 2 to 5 partners had 6.581 times the odds (95% CI 4.643-9.328; P<.001) of using dating apps, whereas people with more than 5 partners had 14 times the odds of using dating apps (OR 14.294, 95% CI 8.92-22.906; P<.001).

Table 1. Logistic regression of geosocial networking dating app users and demographics (N=862).

Characteristics	Total, n (%)	Users, n (%)	Nonusers, n (%)	Dating app users, crude odds ratio (95% CI)	*P* value
Age (years) (N=862)					
18-20	383 (44.4)	192 (50.1)	191 (49.9)	1.054 (0.657-1.691)	.83
21-24	395 (45.8)	215 (54.4)	180 (45.6)	1.253 (0.782-2.007)	.35
25-30	84 (9.7)	41 (48.8)	43 (51.2)	Ref[a]	N/A[b]
Gender (n=860)					
Female	558 (64.9)	291 (52.2)	267 (47.8)	Ref	N/A
Male	302 (35.1)	156 (51.7)	146 (48.3)	1.020 (0.771-1.350)	.89
Sexual orientation (N=862)					
Heterosexual	770 (89.3)	388 (50.4)	382 (49.6)	Ref	N/A
Nonheterosexual				1.846 (1.175-2.900)	.008
Homosexual	25 (2.9)	19 (76.0)	6 (24.0)		
Bisexual	54 (6.3)	32 (59.3)	22 (40.7)		
Other	13 (1.5)	9 (69.2)	4 (30.8)		
Relationship status (n=858)					
Single, not dating	310 (36.1)	195 (62.9)	115 (37.1)	3.936 (2.853-5.430)	.003
Casual dating or open relationship	186 (21.7)	141 (75.8)	45 (24.2)	7.237 (4.857-10.891)	<.001
Exclusive	362 (42.2)	109 (30.1)	253 (69.9)	Ref	N/A
Sexually active (N=862)					
Yes	813 (94.3)	417 (51.3)	396 (48.7)	Ref	N/A
No	49 (5.7)	31 (63.3)	18 (36.7)	1.635 (0.900-2.907)	.11
Sexual partners in the past 12 months (n=812)					
1	311 (38.3)	68 (21.9)	243 (78.1)	Ref	N/A
2-5	341 (42.0)	221 (64.8)	120 (35.2)	6.581 (4.643-9.328)	<.001
5+	160 (19.7)	128 (80.0)	32 (20.0)	14.294 (8.92-22.906)	<.001

[a]Ref: reference (this is the comparison group used to establish the odds ratio).

[b]N/A: not applicable.

Risky Sexual Behavior and User Status

Contraception

Table 2 shows that participants who used condoms had 1.914 times (95% CI 1.448-2.531; *P*<.001) the odds of using GNDAs. Participants who relied on emergency contraception as Plan B were more likely to use dating apps (OR 2.357, 95% CI 1.381-4.087; *P*=.002), whereas those who used hormonal contraception such as IUDs, the oral contraceptive pill, and the rod were less likely to use dating apps (OR 0.660, 95% CI 0.496-0.879; *P*=.004).

Table 3 shows that after adjusting for sexual orientation, relationship status, and number of sexual partners, no association was found between any form of contraception use and dating app use.

Table 2. Logistic regression of geosocial networking dating app users and sexual behaviors.

Behavior	Total, n (%)	Users, n (%)	Nonusers, n (%)	Dating app users, crude odds ratio (95% CI)	*P* value
Contraception (n=811)					
No contraception at all	76 (15.7)	44 (57.9)	32 (43.3)	1.334 (0.828-2.162)	.24
Condom use	417 (51.4)	247 (59.2)	170 (40.8)	1.914 (1.448-2.531)	<.001
Plan B	67 (8.3)	47 (70.1)	20 (29.9)	2.357 (1.381-4.087)	.002
Pulling out	211 (26)	114 (54)	97 (46)	1.152 (0.841-1.578)	.38
Hormonal	505 (62.3)	240 (47.5)	265 (52.5)	0.660 (0.496-0.879)	.004
Specific sexual behaviors (n=788)					
Sex without any form of contraception	385 (48.9)	205 (53.2)	180 (46.8)	1.227 (0.928-1.623)	.15
Sex without a condom but with other contraception	524 (66.5)	262 (50)	262 (50)	0.927 (0.690-1.246)	.62
Casual sex	413 (52.4)	280 (67.8)	133 (32.8)	4.567 (3.383-6.167)	<.001
Sex while drunk	581 (73.7)	300 (51.6)	281 (48.4)	1.165 (0.848-1.600)	.35
No discussion about STI[a] status and/or boundaries before sex	278 (35.3)	189 (70)	89 (30)	3.024 (2.223-4.113)	<.001
Multiple sexual partners	292 (37.1)	221 (75.7)	71 (24.3)	5.592 (4.043-7.736)	<.001
Having unprotected sex with a PWID[b]	60 (7.6)	38 (63.3)	22 (36.7)	1.756 (1.018-3.028)	.04
Having sex with a partner with an STI	52 (6.6)	35 (67.3)	17 (32.7)	2.110 (1.161-3.835)	.01

[a]STI: sexually transmitted infection.

[b]PWID: person who injects drugs.

Table 3. Multivariate logistic regression models of geosocial networking dating app use and sexual behaviors.

Characteristics	Dating app users, adjusted odds ratio (95% CI)	*P* value	Coefficient of determination (R^2)	Chi-square (*df*)
Contraception (n=811)				
Condom use[a]	1.334 (0.958-1.859)	.09	0.341	2.899 (1)
Plan B[a]	1.276 (0.698-2.331)	.42	0.338	0.639 (1)
Hormonal[a]	0.844 (0.601-1.186)	.33	0.338	0.955 (1)
Specific sexual behaviors (n=788)				
Casual sex[b]	3.096 (2.225-4.307)	<.001	0.272	45.346 (1)
No discussion about STI[c] status and/or boundaries before sex[a]	1.755 (1.232-2.500)	.002	0.357	9.705 (1)
Multiple sexual partners[b]	3.943 (2.782-5.588)	<.001	0.295	62.261 (1)
Having unprotected sex with an PWID[a,d]	1.247 (0.649-2.399)	.51	0.345	0.444 (1)
Having sex with a partner with an STI[a]	1.645 (0.815-3.319)	.16	0.347	1.988 (1)

[a]Adjusted for sexual orientation, relationship status, and number of sexual partners.

[b]Adjusted for sexual orientation and relationship status.

[c]STI: sexually transmitted infection.

[d]PWID: person who injects drugs.

Specific Sexual Behaviors

Table 2 shows that participants who had casual sex (*P*<.001), those who had no discussion about STIs or boundaries before having sex (*P*<.001), those with multiple sexual partners (*P*<.001), those having unprotected sex with a person who injects drugs (*P*=.04), and those having sex with a partner with an STI (*P*=.01) had higher odds of using GNDAs. Table 3 shows that after adjusting for sexual orientation and relationship status, those having casual sex have 3.096 times (95% CI 2.225-4.307;

P<.001) the odds of using dating apps and those having multiple sexual partners have 3.943 times (95% CI 2.782-5.588; *P*<.001) the odds of using dating apps compared with those who have not. Similarly, after adjusting for sexual orientation, relationship status, and number of sexual partners, people who had no discussions before having sex about STIs or boundaries were more likely to use dating apps (OR 1.755, 95% CI 1.232-2.500; *P*=.002). However, no association was found between dating app usage and having unprotected sex with a person who injects drugs or having sex with a person with an STI, after adjusting for sexual orientation, relationship status, and number of sexual partners (Table 3).

Risk Perception of Risky Sexual Behavior and User Status

Table 4 shows that those who perceived the risk of having sex without any form of contraception to be *very high* had 2.486 times (95% CI 2.213-5.096; *P*=.01) the odds of using dating apps than those who saw *no risk*.

Those who felt that the risk of having sex while drunk to be *low to high* had 1.659 times (95% CI 1.067-2.581; *P*=.03) the odds of using dating apps compared with those who saw *no risk*. Similarly, those who felt the risk of having sex when drunk was *very high* were 2.151 times (95% CI 1.087-4.256; *P*=.03) more likely to use dating apps compared with those who felt there was no risk. Finally, people who thought that the risk of having multiple sexual partners was *low to high* had 1.871 times (95% CI 1.024-3.418; *P*=.04) the odds of using dating apps compared with people who thought having multiple partners posed *no risk*.

Table 4. Logistic regression of geosocial networking dating app users and the perceptions of risky sexual behavior.

Behavior and risk perception	Total, n (%)	Users, n (%)	Nonusers, n (%)	Dating app users, crude odds ratio (95% CI)	*P* value
Sex without any form of contraception (n=628)					
No risk	40 (6.4)	15 (37.5)	25 (62.5)	Ref[a]	N/A[b]
Low to high risk	436 (69.4)	214 (49.1)	222 (50.9)	1.607 (0.825-3.131)	.16
Very high risk	152 (24.2)	91 (59.9)	61 (40.1)	2.486 (2.213-5.096)	.01
Sex without condom but with other contraception (n=663)					
No risk	69 (10.4)	28 (40.6)	41(59.4)	Ref	N/A
Low to high risk	561 (84.6)	288 (51.3)	273 (48.7)	1.545 (0.929-2.568)	.09
Very high risk	33 (5)	17 (51.5)	16 (48.5)	1.556 (0.675-3.585)	.30
Casual sex (n=610)					
No risk	94 (15.4)	52 (55.3)	42 (43.7)	Ref	N/A
Low to high risk	484 (79.3)	265 (54.8)	219 (45.2)	0.977 (0.627-1.524)	.92
Very high risk	32 (5.2)	20 (62.5)	12 (37.5)	1.346 (0.591-3.066)	.48
Sex while drunk (n=700)					
No risk	96 (13.7)	38 (39.6)	58 (60.4)	Ref	N/A
Low to high risk	551 (78.7)	287 (52.1)	264 (47.9)	1.659 (1.067-2.581)	.03
Very high risk	53 (7.6)	31 (58.5)	22 (41.5)	2.151 (1.087-4.256)	.03
No discussion about STIs[c] and/or boundaries before sex (n=580)					
No risk	40 (6.9)	18 (45)	22 (55)	Ref	N/A
Low to high risk	370 (63.8)	196 (53)	174 (47)	1.377 (0.715-2.652)	.34
Very high risk	170 (29.3)	98 (57.6)	72 (42.4)	1.664 (0.832-3.327)	.15
Multiple sexual partners (n=581)					
No risk	48 (8.3)	20 (41.7)	28 (58.3)	Ref	N/A
Low to high risk	465 (80.3)	266 (57.2)	199 (42.8)	1.871 (1.024-3.418)	.04
Very high risk	68 (11.7)	37 (54.4)	31 (45.6)	1.671 (0.792-3.524)	.18
Having unprotected sex with an IVDU[d] (n=497)					
No risk	46 (9.3)	22 (47.8)	24 (52.2)	Ref	N/A
Low to high risk	212 (42.7)	102 (48.1)	110 (51.9)	1.012 (0.534-1.915)	.97
Very high risk	239 (48.1)	131 (54.8)	108 (45.2)	1.323 (0.703-2.490)	.39
Having sex with a partner with an STI (n=493)					
No risk	42 (8.5)	21 (50)	21 (50)	Ref	N/A
Low to high risk	93 (18.9)	47 (50.5)	46 (49.5)	1.022 (0.493-2.118)	.95
Very high risk	358 (72.6)	191 (53.4)	167 (46.6)	1.144 (0.603-2.168)	.68

[a]Ref: reference (this is the comparison group used to establish the odds ratio).

[b]N/A: not applicable.

[c]STI: sexually transmitted infection.

[d]IVDU: intravenous drug user.

Discussion

Principal Findings

This study explored the link between risky sexual behaviors, risk perceptions, and GNDA use among festival goers. After adjusting for confounders, statistically significant associations existed between GNDA use and lack of discussion about safe sex, engaging in casual sex, and having multiple sexual partners. Festival goers who perceived sex without any form of contraception, having sex while drunk, and having multiple sexual partners as risky were more likely to be app users. A high proportion of dating app users (389/440, 88.4%) also thought that GNDAs should promote safe sex.

Condom users were more likely to use dating apps, which can be seen as a protective factor; however, this was not significant after adjusting for confounders. People having unprotected sex with a person who injects drugs and having sex with a partner with an STI were about twice as likely to use dating apps. After adjusting for confounders, this effect was no longer apparent; however, this may have been because of the relatively small number of people in these categories. Given the prevalence of risky sexual behavior in this music festival population, it would be appropriate for GNDA companies and public health experts to include targeted health promotion interventions on such platforms [1,9,15].

Most participants were heterosexual (770/862, 89.3%), directly reflective of the general Australian population, where 11% identify as LGBTQ+ [16]. The LGBTQ+ group were almost twice as likely to use dating apps, a rate comparable with recent literature [3,4,9]. Rosenfeld et al [17] stated that over 60% of same-sex couples met on the web in 2008 and 2009 in the United States, positing a *thin dating market* for LGBTQ+ individuals. It is not surprising that those who are in an exclusive relationship are less likely to use dating apps, whereas those with multiple sexual partners are much more likely to use dating apps.

This study found that music festival attendees who engage in casual sex, do not discuss safe sex, have multiple partners, have unprotected sex with intravenous drug users, or someone with an STI are more likely to use dating apps. Peter and Valkenburg [18] had a similar finding that individuals who scored highly on *sexual permissiveness*—having more tolerant attitudes and higher engagement in risky sexual behaviors—were more likely to use GNDAs. Chan [19] also demonstrated that GNDA users have more casual relationships and sexual partners.

This high rate of casual sex in GNDA users attending music festivals is significant because of the increased STI risk [20]. In Australia, the rates of notifiable STIs such as gonorrhea and chlamydia have risen significantly in recent years [21], implying that there is a risk involved with having multiple sexual partners. In this study, people with multiple sexual partners were more likely to use dating apps. However, the effective use of barrier protection greatly diminishes the likelihood of STI contraction, with the male condom offering 90% protection against gonorrhea [22]. Those who perceived the risk of sex with no form of contraception to be *very high* were 2.5 times more likely to use dating apps and those using condoms were almost twice as likely to use dating apps, potentially signaling that dating app users are aware of the risks involved. A 2016 study found that young Australians have high rates of condom failures, with 48% experiencing the condom slipping off during intercourse and because of high rates of inconsistent or incorrect use [23]. In addition, the cost of condoms impedes their use, even in developed countries [24]. Therefore, a practical method to minimize STI contraction could be to promote awareness about correct use as opposed to increasing risk perception and the services that offer free condoms such as Aboriginal medical services and Family Planning New South Wales [25].

Furthermore, despite perceiving no contraception use to be at a *very high* risk, GNDA users at the festival had low rates of hormonal contraception and relied on emergency contraception. MacPhail et al [26] found that despite 75% of Australian university students having positive attitudes toward condom use, only 50% used condoms during their last encounter. Thus, an increasing perception of risk may not always translate to safer sexual behaviors. Given that Australia's abortion rates are among the highest in the developed world [27], it is important to have greater health promotion on practical methods of reducing unintentional pregnancies. Those who did not use efficacious hormonal contraceptive options such as the oral contraceptive pill and IUDs [28] had higher odds of using GNDAs. Despite the high uptake of barrier contraception, its efficacy tends to be user dependent, as mentioned earlier [23]. Distributing information encouraging longer-acting contraceptive devices and correct contraceptive use via dating apps may be an effective means of curbing accidental pregnancies [28,29]. In addition, this study found that 37.1% (292/788) of sexually active respondents had multiple sexual partners in the last 12 months and that 48.9% (385/788) reported having not used any form of contraception. This is comparable with the study by Lim et al [30], which was conducted at the Big Day Out festival in 2007, with 48% of respondents reporting multiple partners in the last 3 months and 43% not using a condom because of substance abuse. Thus, it seems that over the past 13 years, no significant reduction in engagement in risky sexual behaviors among Australian music festival populations can be noted, and it remains an area in need of targeted sexual health promotion strategies.

Limitations and Strengths

Study limitations included the use of a binary measure of dating app usage in the last 12 months, self-report, selection bias, and the inability to show causal effects because of the cross-sectional design. A longitudinal study method should be conducted to determine the relationship between GNDA use and sexual health behaviors and outcomes over time. Although the survey used evidence-based, risky sexual behaviors [14], the SSBQ was abridged to ensure respondents remained engaged and avoid survey fatigue. This may be problematic, as the study may not have captured the full breadth of risky sexual behaviors within the target demographic. Respondents were willing to complete the survey because the topic is of interest to the study population; however, we did not record how many people declined to partake.

Another study limitation is the potential influence of relationship status and sexual orientation on dating app usage. Our study found that LGBTQ+ people are twice as likely as heterosexuals to use dating apps. This may influence the results about the performance of risky sexual behaviors, as it has been shown that people of sexual minorities, especially men who have sex with men, engage in risky sexual behaviors. Similarly, an increased number of sexual partners and not being in an exclusive relationship were predictors of GNDA use. However, this was adjusted for in multivariate analyses. It should also be noted that participants could have been under the influence of drugs or alcohol, the data were therefore collected early in the day, and people who were perceived to be under the influence were excluded. Finally, the terminologies *casual sex* and *multiple sexual partners* may be considered vague and interpreted differently between participants. Thus, the results

from questions in which these terms were used need to be interpreted with caution.

Practical Implications

An overwhelming number of festival attendees who were GNDA users (389/440, 88.4%) indicated that safe-sex messaging should be included in GNDAs. Our study also demonstrated that young festival attendees engaging in more high-risk sexual behaviors were more likely to be app users. Therefore, as GNDA use at music festivals is promoted by app companies [8], it is recommended to focus on safe-sex messaging. This provides a platform for health strategists to target at-risk demographics. However, health professionals should be aware that a 2018 Australian survey found that although 88.7% of those aged between 18 and 29 years noticed sexual health promotion, only 40.9% considered it relevant and only 32% felt an increase in knowledge [15]. Therefore, the success of safe-sex messaging relies on whether the material will be relevant to users and increase their knowledge of risky sexual behavior. However, as stated earlier, increasing the perception of risk may not always translate to safer sexual behaviors. Despite relatively high levels of risk perception, health promotion messages to lower risk among music festival patrons should be further explored. Areas of future research include examining which GNDAs have the riskiest user bases and which strategies are most effective for harm reduction in this population. Focus groups with festival participants could be conducted to help ensure that messaging is relevant and targeted.

Major GNDAs such as Tinder and Bumble run web-based promotional subsidiaries, including *SwipeLife* (Tinder) [31], *The Buzz* (Bumble) [32], and blogs that incorporate safe-sex articles. As stated previously, less than 19% of heterosexual app users saw safe-sex messages in dating apps [9], despite existing literature identifying the importance of web-based safe sexual health messages to users [9,15,33]. Safe-sex messages may therefore need to appear more prominently within the apps themselves. This is further confirmed by our study, in which an overwhelming majority of participants were in favor of in-app sexual health resources. The fact that Tinder has released coronavirus safety messages in March 2020 demonstrates their ability and willingness to use health promotion messages [31]. In addition, intervention research for safer dating app use has also emerged [34]. Finally, there is evidence that mobile phone interventions can be successful in delivering safe-sex messages [35,36] and strengthens the argument that safe-sex messaging could appear more prominently within dating apps. However, a recent systematic review on new digital media interventions for sexual health promotion among young people reported that it should be taken into account that the technology itself does not necessarily lead to success [11]. The authors suggest that interventions should use high-quality, evidence-based content that engages with young people. Formative research [37] among Swedish youth on the development of a mobile phone app to promote safe sex identified that the following features would engage youth and therefore useful for the app development: *Condom Obstacles and Solutions; Quiz; Games; Self-Refection; Challenges; Stories by Peers (stories from peers and information from a doctor); Condom Tips, Pep Talk, and Boosting; and Random Facts*. Further guidelines are available for complex messaging in health promotion when developing interactive eHealth apps [38].

Conclusions

This study identified that festival goers engaging in certain high-risk sexual behaviors, including casual sex, having multiple sexual partners, and having sex without discussion about STI status and boundaries, are more likely to use dating apps. Festival attendees who perceived sex without any form of contraception, having sex while drunk, and having multiple sexual partners as risky were more likely to be app users. A high proportion of dating app users support the notion that GNDAs should promote safe sex. The results of this study contribute to the growing body of knowledge surrounding the changing landscape of dating, sexual behaviors, and health impacts in the era of GNDAs. Policy makers and GNDA developers should acknowledge the vulnerability of their users to adverse sexual health outcomes and use GNDAs as a platform to promote risk-reduction practices.

Acknowledgments

The study group would like to thank the Positive Adolescent Sexual Health-North Coast Youth Sexual Health Consortium for facilitating data collection.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Survey questions.
[DOCX File , 17 KB - jmir_v23i4e21082_app1.docx]

References

1. Sumter SR, Vandenbosch L, Ligtenberg L. Love me Tinder: untangling emerging adults? motivations for using the dating application Tinder. Telemat Inform 2017;34(1):67-78 [FREE Full text] [doi: 10.1016/j.tele.2016.04.009]
2. November 2017: online dating - relationships Australia. Relationships.org.au. 2020. URL: https://www.relationships.org.au/what-we-do/research/online-survey/november-2017-online-dating [accessed 2020-06-28]

3. Carpenter CJ, McEwan B. The players of micro-dating: individual and gender differences in goal orientations toward micro-dating apps. First Monday 2016;21(5) [FREE Full text] [doi: 10.5210/fm.v21i5.6187]
4. Grello CM, Welsh DP, Harper MS. No strings attached: the nature of casual sex in college students. J Sex Res 2006;43(3):255-267 [FREE Full text] [doi: 10.1080/00224490609552324]
5. National Centre in HIV Epidemiology and Clinical Research. Sydney: University of NSW; 2010. URL: https://kirby.unsw.edu.au/sites/default/files/kirby/report/SERP_2010-Annual-Surveillance-Report.pdf [accessed 2020-06-28]
6. Wang H, Zhang L, Zhou Y, Wang K, Zhang X, Wu J, et al. The use of geosocial networking smartphone applications and the risk of sexually transmitted infections among men who have sex with men: a systematic review and meta-analysis. BMC Public Health 2018 Oct 16;18(1):1178 [FREE Full text] [doi: 10.1186/s12889-018-6092-3] [Medline: 30326887]
7. Shapiro GK, Tatar O, Sutton A, Fisher W, Naz A, Perez S, et al. Correlates of tinder use and risky sexual behaviors in young adults. Cyberpsychol Behav Soc Netw 2017 Dec;20(12):727-734. [doi: 10.1089/cyber.2017.0279] [Medline: 29211500]
8. Costello L. Tinder's new 'festival mode' for Splendour in the Grass is now officially live. Pedestrian.tv. 2019. URL: https://www.pedestrian.tv/sex-dating/tinder-festival-mode-live-splendour-in-the-grass/[accessed [accessed 2020-06-28]
9. Albury K, Bryon P, Walshe J. Safety, risk and wellbeing on dating apps: final report. Melbourne: Swinburne University of Technology 2019; 2019. URL: https://apo.org.au/node/268156 [accessed 2020-06-28]
10. Jenkinson R, Bowring A, Dietze P, Hellard M, Lim MSC. Young risk takers: alcohol, illicit drugs, and sexual practices among a sample of music festival attendees. J Sex Transm Dis 2014;2014:- [FREE Full text] [doi: 10.1155/2014/357239] [Medline: 26316974]
11. Wadham E, Green C, Debattista J, Somerset S, Sav A. New digital media interventions for sexual health promotion among young people: a systematic review. Sex Health 2019 Apr;16(2):101-123. [doi: 10.1071/SH18127] [Medline: 30819326]
12. Eleuteri S, Rossi R, Tripodi F, Fabrizi A, Simonelli C. Sexual health in your hands: how the smartphone apps can improve your sexual wellbeing? Sexologies 2018 Jul 1;27(3):57-60 [FREE Full text] [doi: 10.1016/j.sexol.2018.04.004]
13. Huang ET, Williams H, Hocking JS, Lim MS. Safe sex messages within dating and entertainment smartphone apps: a review. JMIR Mhealth Uhealth 2016 Nov 08;4(4):e124 [FREE Full text] [doi: 10.2196/mhealth.5760] [Medline: 27826133]
14. DiIorio C, Parsons M, Lehr S, Adame D, Carlone J. Measurement of safe sex behavior in adolescents and young adults. Nurs Res 1992;41(4):203-208. [Medline: 1408860]
15. Adam PCG, de Wit JBF, Ketsuwan I, Treloar C. Sexual health-related knowledge, attitudes and practices of young people in Australia Results from the 2018 Debrief Survey among heterosexual and non-heterosexual participants. UNSW Sydney. Sydney: Centre for Social Research in Health; 2019. URL: http://unsworks.unsw.edu.au/fapi/datastream/unsworks:55103/bin2f1e798c-2e44-4a48-aec3-9a7617568015?view=true [accessed 2020-06-28]
16. Face the facts: Lesbian, gay, bisexual, trans and intersex people. Australian Human Rights Commission 2015. URL: https://www.humanrights.gov.au/our-work/education/face-facts-lesbian-gay-bisexual-trans-and-intersex-people [accessed 2020-07-03]
17. Rosenfeld MJ, Thomas RJ. Searching for a mate: the rise of the internet as a social intermediary. Am Sociol Rev 2012;77(4):523-547 [FREE Full text] [doi: 10.1177/0003122412448050]
18. Peter J, Valkenburg PM. Who looks for casual dates on the internet? A test of the compensation and the recreation hypotheses. New Media & Society 2016 Jun 30;9(3):455-474 [FREE Full text] [doi: 10.1177/1461444807076975]
19. Chan C. Swipe left on disease: the effects of contextual factorsdating app usage on sexual risk behaviours among young adults. Br Med J 2017;93(2):- [FREE Full text] [doi: 10.1136/sextrans-2017-053264.521]
20. Lyons HA. Heterosexual casual sex and STI diagnosis: a latent class analysis. Int J Sex Health 2017 Jul 15;29(1):32-47 [FREE Full text] [doi: 10.1080/19317611.2016.1210711] [Medline: 29276549]
21. Department of Health. Pregnancy Care Guidelines: 41 Gonorrhoea. Australian Government. URL: https://www.health.gov.au/resources/pregnancy-care-guidelines/part-g-targeted-maternal-health-tests/gonorrhoea [accessed 2020-06-28]
22. Marfatia YS, Pandya I, Mehta K. Condoms: past, present, and future. Indian J Sex Transm Dis AIDS 2015;36(2):133-139 [FREE Full text] [doi: 10.4103/0253-7184.167135] [Medline: 26692603]
23. Hall KM, Brieger DG, De Silva SH, Pfister BF, Youlden DJ, John-Leader F, et al. Errors and predictors of confidence in condom use amongst young australians attending a music festival. J Sex Transm Dis 2016;2016:- [FREE Full text] [doi: 10.1155/2016/6054870] [Medline: 27957381]
24. Sarkar NN. Barriers to condom use. Eur J Contracept Reprod Health Care 2008 Jun 06;13(2):114-122. [doi: 10.1080/13625180802011302] [Medline: 18465472]
25. NSW Health. Where Can I buy Condoms? Playsafe. URL: https://playsafe.health.nsw.gov.au/condoms/where-to-get-condoms/ [accessed 2020-07-03]
26. MacPhail C, Dune T, Dillon G, Rahman S, Khanam R, Jenkins L, et al. Knowledge and attitudes to sexual health and STI testing for students at an Australian Regional University: a cross-sectional study. J Aus N Z Stud Serv Asso. 2017. URL: https://ro.uow.edu.au/sspapers/2933/ [accessed 2021-03-29]
27. Mazza D, Bateson D, Frearson M, Goldstone P, Kovacs G, Baber R. Current barriers and potential strategies to increase the use of long-acting reversible contraception (LARC) to reduce the rate of unintended pregnancies in Australia: an expert roundtable discussion. Aust N Z J Obstet Gynaecol 2017 Apr;57(2):206-212. [doi: 10.1111/ajo.12587] [Medline: 28294293]

28. Goldhammer DL, Fraser C, Wigginton B, Harris ML, Bateson D, Loxton D, et al. What do young Australian women want (when talking to doctors about contraception)? BMC Fam Pract 2017 Mar 15;18(1):35 [FREE Full text] [doi: 10.1186/s12875-017-0616-2] [Medline: 28298197]
29. Coombe J, Harris ML, Wigginton B, Lucke J, Loxton D. Contraceptive use at the time of unintended pregnancy: findings from the contraceptive use, pregnancy intention and decisions study. Aust Fam Physician 2016 Nov;45(11):842-848 [FREE Full text] [Medline: 27806456]
30. Lim MSC, Hellard ME, Aitken CK, Hocking JS. Sexual-risk behaviour, self-perceived risk and knowledge of sexually transmissible infections among young Australians attending a music festival. Sex Health 2007 Mar;4(1):51-56. [doi: 10.1071/sh06031] [Medline: 17382039]
31. Pfeuffer C. How to have the STI/STD talk without making it weird. Swipelife. URL: https://swipelife.tinder.com/post/safer-sex [accessed 2020-07-03]
32. The Buzz. Bumble. URL: https://bumble.com/the-buzz/ [accessed 2020-07-03]
33. Bailey J, Mann S, Wayal S, Hunter R, Free C, Abraham C, et al. Sexual health promotion for young people delivered via digital media: a scoping review. Southampton (UK): NIHR J Library; 2015.
34. Wong WCW, Song L, See C, Lau STH, Sun WH, Choi KWY, et al. Using crowdsourcing to develop a peer-led intervention for safer dating app use: pilot study. JMIR Form Res 2020 Apr 21;4(4):- [FREE Full text] [doi: 10.2196/12098] [Medline: 32314975]
35. Gold J, Aitken CK, Dixon HG, Lim MSC, Gouillou M, Spelman T, et al. A randomised controlled trial using mobile advertising to promote safer sex and sun safety to young people. Health Educ Res 2011 Oct;26(5):782-794. [doi: 10.1093/her/cyr020] [Medline: 21447750]
36. Steinberg A, Griffin-Tomas M, Abu-Odeh D, Whitten A. Evaluation of a mobile phone app for providing adolescents with sexual and reproductive health information, New York City, 2013-2016. Public Health Rep 2018;133(3):234-239 [FREE Full text] [doi: 10.1177/0033354918769289] [Medline: 29664696]
37. Nielsen A, Bågenholm A, De Costa A. Development of a mobile phone app to promote safe sex practice among youth in Stockholm, Sweden: qualitative study. JMIR Form Res 2020 Jan 28;4(1):- [FREE Full text] [doi: 10.2196/12917] [Medline: 32012038]
38. Heffernan KJ, Chang S, Maclean ST, Callegari ET, Garland SM, Reavley NJ, et al. Guidelines and recommendations for developing interactive ehealth apps for complex messaging in health promotion. JMIR Mhealth Uhealth 2016 Mar 09;4(1):e14 [FREE Full text] [doi: 10.2196/mhealth.4423] [Medline: 26860623]

Abbreviations

GNDA: geosocial networking dating app
IUD: intrauterine device
LGBTQ+: lesbian, gay, bisexual, transgender, and queer
OR: odds ratio
SSBQ: Safer Sex Behavior Questionnaire
STI: sexually transmitted infection

Edited by R Kukafka; submitted 04.06.20; peer-reviewed by S Moukalled, B Ma; comments to author 25.06.20; revised version received 02.07.20; accepted 25.02.21; published 15.04.21.

Please cite as:
Garga S, Thomas M, Bhatia A, Sullivan A, John-Leader F, Pit S
Geosocial Networking Dating App Usage and Risky Sexual Behavior in Young Adults Attending a Music Festival: Cross-sectional Questionnaire Study
J Med Internet Res 2021;23(4):e21082
URL: https://www.jmir.org/2021/4/e21082
doi:10.2196/21082
PMID:33856354

Original Paper

Characteristics of Online Health Care Services From China's Largest Online Medical Platform: Cross-sectional Survey Study

Xuehan Jiang[1*], PhD; Hong Xie[2*], MM; Rui Tang[1], MSc; Yanmei Du[2], MM; Tao Li[2], MD; Jinsheng Gao[2], MD; Xiuping Xu[2], MM; Siqi Jiang[2], BE; Tingting Zhao[1], MSc; Wei Zhao[1], PhD; Xingzhi Sun[1], PhD; Gang Hu[1], MA; Dejun Wu[2], MA; Guotong Xie[1], PhD

[1]Ping An Healthcare Technology, Beijing, China

[2]Ping An Healthcare and Technology Company Limited, Shanghai, China

[*]these authors contributed equally

Corresponding Author:
Hong Xie, MM
Ping An Healthcare and Technology Company Limited
17/F Building B, No. 166 Kaibin Road, Xuhui District
Shanghai
China
Phone: 86 18951670324
Email: xiehong858@jk.cn

Abstract

Background: Internet hospitals in China are in great demand due to limited and unevenly distributed health care resources, lack of family doctors, increased burdens of chronic diseases, and rapid growth of the aged population. The COVID-19 epidemic catalyzed the expansion of online health care services. In recent years, internet hospitals have been rapidly developed. Ping An Good Doctor is the largest, national online medical entry point in China and is a widely used platform providing online health care services.

Objective: This study aims to give a comprehensive description of the characteristics of the online consultations and inquisitions in Ping An Good Doctor. The analyses tried to answer the following questions: (1) What are the characteristics of the consultations in Ping An Good Doctor in terms of department and disease profiles? (2) Who uses the online health services most frequently? and (3) How is the user experience of the online consultations of Ping An Good Doctor?

Methods: A total of 35.3 million consultations and inquisitions over the course of 1 year were analyzed with respect to the distributions of departments and diseases, user profiles, and consulting behaviors.

Results: The geographical distribution of the usage of Ping An Good Doctor showed that Shandong (18.4%), Yunnan (15.6%), Shaanxi (7.2%), and Guangdong (5.5%) were the provinces that used it the most; they accounted for 46.6% of the total consultations and inquisitions. In terms of department distribution, we found that gynecology and obstetrics (19.2%), dermatology (17.0%), and pediatrics (14.4%) were the top three departments in Ping An Good Doctor. The disease distribution analysis showed that, except for nondisease-specific consultations, acute upper respiratory infection (AURI) (4.1%), pregnancy (2.8%), and dermatitis (2.4%) were the most frequently consulted diseases. In terms of user profiles, females (60.4%) from 19 to 35 years of age were most likely to seek consultations online, in general. The user behavior analyses showed that the peak times of day for online consultations occurred at 10 AM, 3 PM, and 9 PM. Regarding user experience, 93.0% of users gave full marks following their consultations. For some disease-related health problems, such as AURI, dermatitis, and eczema, the feedback scores were above average.

Conclusions: The prevalence of internet hospitals, such as Ping An Good Doctor, illustrated the great demand for online health care services that can go beyond geographical limitations. Our analyses showed that nondisease-specific issues and moderate health problems were much more frequently consulted about than severe clinical conditions. This indicated that internet hospitals played the role of the family doctor, which helped to relieve the stress placed on offline hospitals and facilitated people's lives. In addition, good user experiences, especially regarding disease-related inquisitions, suggested that online health services can help solve health problems. With support from the government and acceptance by the public, online health care services could develop at a fast pace and greatly benefit people's daily lives.

(J Med Internet Res 2021;23(4):e25817) doi:10.2196/25817

KEYWORDS

eHealth; internet hospital; China; online health care services; mHealth; COVID-19; digital health; app; online consultation; user experience

Introduction

Background

Internet hospital is an innovative type of hospital where professional physicians provide health care services on the internet. The online health care services at these hospitals include, but not limited to, health-related consultations, disease diagnoses, medication prescription, and chronic disease management [1].

In this study, internet hospitals generally refer to all applications that provide online health services for disease consultation and treatment through information technology. Services by internet hospitals include a combination of offline medical treatment and online follow-up consultation, telemedicine, online medical consultation, and online health management. Internet hospitals are sometimes referred to worldwide as telehealth [2]. In the United States, each state's laws, regulations, and Medicaid program policies for telehealth differ significantly. However, the common aspect they share is that most states and Washington, DC, provide reimbursement via Medicaid for some form of live video as a fee-for-service [3]. In the European Union, most countries have no formal definition of telemedicine services [4]. There is an online consultation system in South West England that allows adult patients to contact their general practitioner, but evidence indicates that the use of e-consultations is very low [5]. Compared with developed countries, telehealth could be more meaningful in developing countries because of limited medical resources and poor health care services [5,6].

Internet hospitals have been rapidly established in China in recent years [7-9]. As of May 2019, there were 158 internet hospitals in China [10], and as of October 28, 2020, there were about 900 of them operating in China [11]. The expansion of internet hospitals is due to the following five reasons.

First, health care resources are limited and unevenly distributed by geography in China [12]. By 2019, there were only 2.77 licensed physicians per 1000 people [13]; for city residents, there were 4.10 licensed physicians per 1000 people, while that number was 1.96 for rural residents [13]. There are 3-tier health care systems in China [14,15]. Primary health care providers are usually community based and are expected to play the role of general practitioner and to perform health care management. Secondary and tertiary health care providers include more specialists and focus on more complicated clinical problems. However, tertiary health care providers are mainly located in the eastern cities of China where the economies are more developed. The top 100 hospitals in China have mostly been located in big cities, such as Beijing and Shanghai, and provincial capitals [16]. As a result, residents in rural areas or western cities have had limited access to high-quality health care services [16].

Second, family doctor systems in China have been underdeveloped [17,18]. As the family doctor plays a gatekeeping role in developed countries, it is convenient for one to seek health care services for mild problems. Although the Chinese government launched a series of policies and regulations to accelerate family doctors' contracting services during the last decades, family doctor systems are still at an early stage in China [17]. The number of general practitioners was not sufficient due to the large population, wide geographic area, and uneven distribution of health care resources. As a consequence, the effect of family doctor systems was compromised in practice [19]. With low public awareness of diseases, patients with moderate symptoms also visited the tertiary hospitals, putting more stress on health care resources. According to the China Health Statistics Yearbook 2020, tertiary hospitals covered 53.5% of patient visits in China, while primary hospitals only accounted for 6.0% [13]. Internet hospitals are easy to access and serve as a supplement to family doctors. Therefore, internet hospitals meet the population's need for convenient access to professional medical help in China.

Third, clinical data sharing has been hindered by unconnected hospital systems. Such information islands result in patients having to endure repeated examinations when changing to a new hospital. However, with internet hospitals, patients are able to upload their existing examination results, which avoids wasting clinical resources.

Fourth, with the increasing burden of medical insurance, the Chinese government proposed the *Healthy China 2030* plan [20] and put more focus on health-driven management instead of traditional disease-driven treatment. For the increasing aged population, health care providers should intervene at the onset of chronic diseases and take advantage of artificial intelligence technology to manage the disease at the same time.

Last but not least, after the outbreak of COVID-19, offline treatment channels were blocked and the Chinese government had adopted a series of administrative measures to encourage the development of internet hospitals [21]. Unnecessary face-to-face contact was avoided with internet hospitals, providing safer and more convenient health care services than in-person, offline hospitals. As suggested by recent work, internet hospitals helped control the COVID-19 epidemic [22,23] and made access to health care services more convenient [24].

In terms of the types of initiators, internet hospitals in China can be divided into government-led, hospital-led, and enterprise-led services [8]. In 2012, the Guangdong Second Provincial General Hospital built the first internet hospital in China, which belonged to the first type of internet hospital (ie, government-led) [14]. Compared with government-led and hospital-led internet hospitals, enterprise-led internet hospitals have the advantages of stronger capabilities in market

exploration and faster product iterations. Enterprises are more open to advanced technology and pay more attention to improving the efficiency of consultation.

Ping An Good Doctor has been one of the leading companies of enterprise-led internet hospitals [21]. It provides health care services via a mobile app of the same name, *Ping An Good Doctor*. Ping An Good Doctor aimed to create a one-stop, whole-process, online-to-offline service platform and integrated online health service platform with offline health services, such as private clinics, pharmacies, health checkup and test centers, and so on [25]. It was rapidly developed, as there was an urgent need for internet hospitals in China. The use of Ping An Good Doctor was nationwide. There were 346 million registered users, more than 1800 staff members working for in-house medical teams, and about 10,000 external experts by the middle of 2020. There were more than 820 million online consultations and inquisitions in total from 2014 until now, which covered a wide range of departments and diseases. The number of cumulative visits during the COVID-19 epidemic—from the period of January 20 to February 10, 2020—reached 1.11 billion [26]. Besides Ping An Good Doctor, Ali and JD are also major players in internet health care in China [27]. However, Ping An Good Doctor has the largest average daily consultation volume [28-30].

Any user with a cell phone can use the health care services provided by the Ping An Good Doctor app anytime and anywhere. Users can ask any health-related questions or seek health care services whether or not they have a specific health problem. In general, a dialogue in the absence of health problems is referred to as a consultation, while seeking clinical help for specific health problems is called an inquisition. In the latter scenario, users are expected to submit their chief complaint and other clinical materials about their health status; the physician would then make the diagnosis and provide corresponding health care services, including prescriptions if needed. Medications, if needed, would be delivered if the users chose to buy them online. All consultations and inquisitions are documented for further quality examination. At the end of the process, users are also asked to score the consultation, where 1 stands for the least satisfaction and 5 for the most satisfaction.

Study Objectives

Despite the rapid development of internet hospitals in China, there exist limited studies examining the characteristics of the departments and diseases of internet hospitals. To fill this gap, we carried out analyses regarding the nationally utilized Ping An Good Doctor platform, aiming to provide a comprehensive description of online medical care in China. Our primary focus was on answering the following three questions: (1) What are the characteristics of the consultations and inquisitions in Ping An Good Doctor in terms of department and disease profiles? (2) Who most often use the platform's online health services? and (3) How do users experience the online consultations of Ping An Good Doctor?

Methods

Data Collection

To eliminate the influence of seasonal patterns, we analyzed the online consultations and inquisitions of Ping An Good Doctor from the past year (ie, from August 2019 to August 2020). Here, consultation referred to the asking of health-related questions, while inquisition meant being treated online for health problems. For all consultations and inquisitions, four types of information were extracted: (1) demographic information about the clients, (2) health care–related variables, (3) details about the consultation process, and (4) consultation evaluations collected after the consultation finished. A detailed description of all variables involved in this study is listed in Table 1.

Table 1. Variables considered in this study.

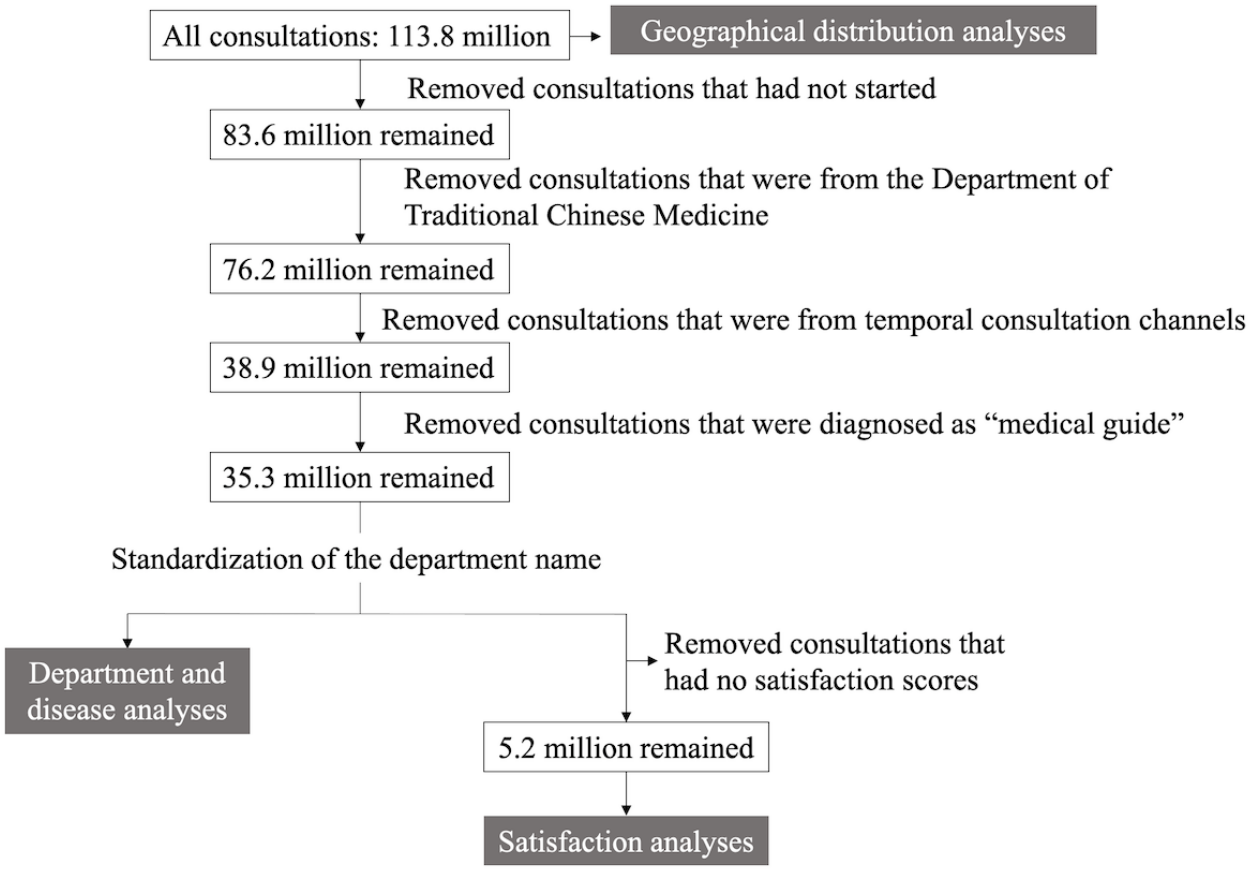
Type of information and variables	Description of variables	Responses (N=113,805,518[a]), n (%)
Demographic information		
Gender	Gender of the user	111,844,894 (98.3)
Age	Age of the user	110,502,135 (97.1)
Province	Province (ie, location) of the user	103,786,537 (91.2)
Health care–related variables		
Diagnosis	Diagnosis as given by the physician	81,574,890 (71.7)
Department	Subentity of the online hospital, such as dermatology	113,805,518 (100)
Details about consultation process		
Dialogue rounds	Number of rounds of dialogue between the user and the physician	113,805,518 (100)
Consultation time	Time of day when the user described their chief complaints	113,805,518 (100)
Consultation evaluation		
Satisfaction score	Scores given by the user after the consultation and/or inquisition	6,879,529 (6.0)

[a]This value represents all consultations and inquisitions combined.

Data Selection

The number of consultations and inquisitions accumulated between August 20, 2019, and August 22, 2020, was around 113.8 million. The detailed process of data selection is illustrated in Figure 1. Four types of consultations and inquisitions were excluded before the analyses of departments and diseases. First, invalid consultations, where the consultations and inquisitions had not started, were removed. Second, consultations from the Traditional Chinese Medicine (TCM) department were excluded since the diagnosis system of TCM is distinct from that of the International Classification of Diseases, Tenth Revision (ICD-10), and is poorly standardized. Third, consultations and inquisitions from some temporal consultation channels, such as the special consulting channel during COVID-19, that would bias the distributions of departments and diseases were also omitted. Lastly, we removed consultations that were labeled as "medical guide," which was an automatic diagnosis label assigned under certain circumstances, such as the unexpected end of a consultation due to the extended nonresponse of the user. After the four data-filtering steps, there were 35.3 million consultations and inquisitions remaining. After standardization of the departments, those data were used for department- and disease-related analyses. To further analyze user satisfaction of the online consultations, data that had no satisfaction scores were filtered out.

Figure 1. The process of data selection and the data sets that were used for different analyses.



Data Preprocessing

The diagnoses from the inquisitions were mainly based on the ICD-10 coding system [31], while consultations were assigned self-defined diagnosis labels, such as *general consultations*, *specific consultations*, and *lifestyle-related questions*. Here, the term *general consultations* was a vague definition referring to asking questions about any nonspecific aspect of health problems, such as how one should take exercise if one has had a stroke. *Specific consultations* referred to asking questions about a specific disease, such as diabetes and cardiovascular diseases. *Lifestyle-related questions* referred to asking questions about how to stay healthy without mentioning specific health problems, such as what tea to drink to protect eyesight.

Generally, the diagnosis labels for inquisitions and consultations were standardized as one-by-one mapping to ICD-10 or self-defined codes.

It should be noted that for 20.2% (7,140,917/35,296,169) of the consultations and inquisitions, there existed multiple diagnoses. In this case, the diagnosis distribution was calculated by splitting the multiple diagnoses of one sample into multiple samples, each sample with a distinct diagnosis. All analyses in terms of diagnoses were based on this splitting approach of diagnosis statistics, such as user experience for each diagnosis.

Statistical Analysis

Data were extracted from the database using Structured Query Language and were preprocessed by the *pandas* library in Python 3.6 (Python Software Foundation).

Results

Overview of the Online Consultations and Inquisitions

The overall statistics of our data set are shown in Table 2. There were 35,296,169 valid consultations and inquisitions in total. During the study period, 12,446,838 users participated in consultations via Ping An Good Doctor. A total of 40.3% of the consultations and inquisitions were made by male users, and the mean age of the users was 27.3 years (SD 17.2). There were, on average, 7.3 (SD 5.3) rounds of conversation between users and physicians. After standardization, there were 76 departments and 1419 diseases covered in the cross section that we studied.

As shown in Figure 2, which was drawn using the *pyecharts* library in Python 3.6, the consultations and inquisitions came from all the provinces, municipalities, and autonomous regions of China. Shandong (19,067,141/113,805,518, 18.4%), Yunnan (16,192,209/113,805,518, 15.6%), Shaanxi (7,439,540/113,805,518, 7.2%), and Guangdong (5,672,425/113,805,518, 5.5%) were the four leading provinces and accounted for 46.6% (48,371,315/113,805,518) of the total consultations and inquisitions in our data set. A city-level rank was also calculated. Table 3 lists the top 10 cities in yearly visits to Ping An Good Doctor. It was interesting to note that big cities such as Beijing and Shanghai only ranked 9th and 5th, respectively, accounting for 34.4% (1,699,251/4,932,918) and 54.7% (2,700,223/4,932,918) of the total visits, relative to the top city, Kunming.

Longitudinally, the COVID-19 epidemic activated the daily visits to Ping An Good Doctor. As shown in Table 4 [32,33], the average daily visits increased by 23.2% during the outbreak of COVID-19 in China. It was encouraging that the average daily visits increased more even after the outbreak of COVID-19 in China.

Table 2. Characteristics of the current data set.

Characteristic	Value
Total consultations and inquisitions, N	35,296,169
Consultations and inquisitions by male users, n (%)	14,063,178/34,904,918 (40.3)
Users of the platform, n	12,446,858
Age of users who submitted consultations and inquisitions (years), mean (SD)	27.3 (17.2)
Rounds of dialogue between physicians and users, mean (SD)	7.3 (5.3)
Departments represented in the data set, n	76
Diseases represented in the data set, n	1419

Figure 2. The number of consultations and inquisitions from all over China out of the original 113.8 million data points. Red dots indicate the capital city of each province and the black dot indicates the city of Kunming. Kunming had the largest volume of consultations and inquisitions in Ping An Good Doctor at the city level.

Table 3. Top 10 cities by visits to Ping An Good Doctor from August 2019 to August 2020.

City	Province	Relative yearly visit amount[a]
Kunming	Yunnan	1
Xi'an	Shaanxi	0.915
Yantai	Shandong	0.835
Qingdao	Shandong	0.559
Shanghai	Shanghai	0.547
Chongqing	Chongqing	0.424
Linyi	Shandong	0.410
Honghe Hani and Yi Autonomous Prefecture	Yunnan	0.385
Beijing	Beijing	0.344
Zhengzhou	Henan	0.305

[a]Values are based on the original 113.8 million data points; relative yearly visits of each city are calculated by setting the yearly visit amount by Kunming as a reference.

Table 4. Average daily visits to Ping An Good Doctor before, during, and after the outbreak of COVID-19 in China.

Phase	Time range	Daily visits, mean (SD)
Before the COVID-19 outbreak	August 20, 2019, to January 22, 2020	74,893.92 (5843.34)
During the COVID-19 outbreak	January 23, 2020[a], to February 29, 2020[b]	92,246.74 (20,411.83)
After the COVID-19 outbreak	March 1, 2020, to August 22, 2020	114,898.50 (11,708.20)

[a]The city of Wuhan closed on January 23, 2020, indicating the start of the outbreak of COVID-19 in China [32].

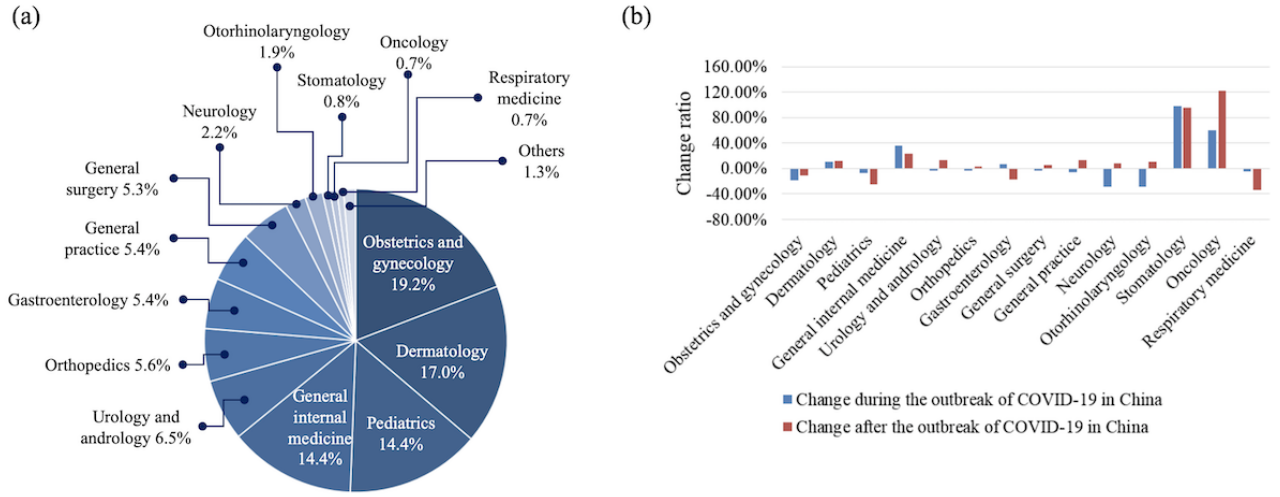
[b]As of February 29, 2020, the number of newly diagnosed COVID-19 cases per day in China was less than 1000 [33].

Characteristics of Departments and Diseases in Online Consultations

Distribution of Departments

Although there were 76 departments represented in our data set, about 14 (18%) departments accounted for more than 99% of the consultations and inquisitions. As shown in Figure 3 (a), the most popular department was gynecology and obstetrics, which was represented in 19.2% (5,868,172/30,579,297) of the consultations, followed by 17.0% (5,204,805/30,579,297) for the dermatology department and 14.4% (4,394,634/30,579,297) for pediatrics. We further analyzed the change ratios of distribution for the top 14 departments during and after the COVID-19 outbreak to investigate the effect of the COVID-19 epidemic on the distribution of departments. The result is shown in Figure 3 (b); for most of the departments, the change ratios were within ±10% during and after the COVID-19 outbreak. It should be noted that the fraction of consultations represented by dermatology increased by more than 10%, both during (from 15. 9% to 17.5%) and after (from 15.9% to 17.7%) the COVID-19 outbreak. In addition, the fraction of consultations represented by general internal medicine increased by 36.5% (from 11.7% to 16.0%) and 22.7% (from 11.7% to 14.4%) during and after the COVID-19 outbreak, respectively. Although the fractions of consultations represented by stomatology and oncology both had a sharp increase, their absolute fractions increased by 0.47% and 0.51% after the COVID-19 outbreak, respectively.

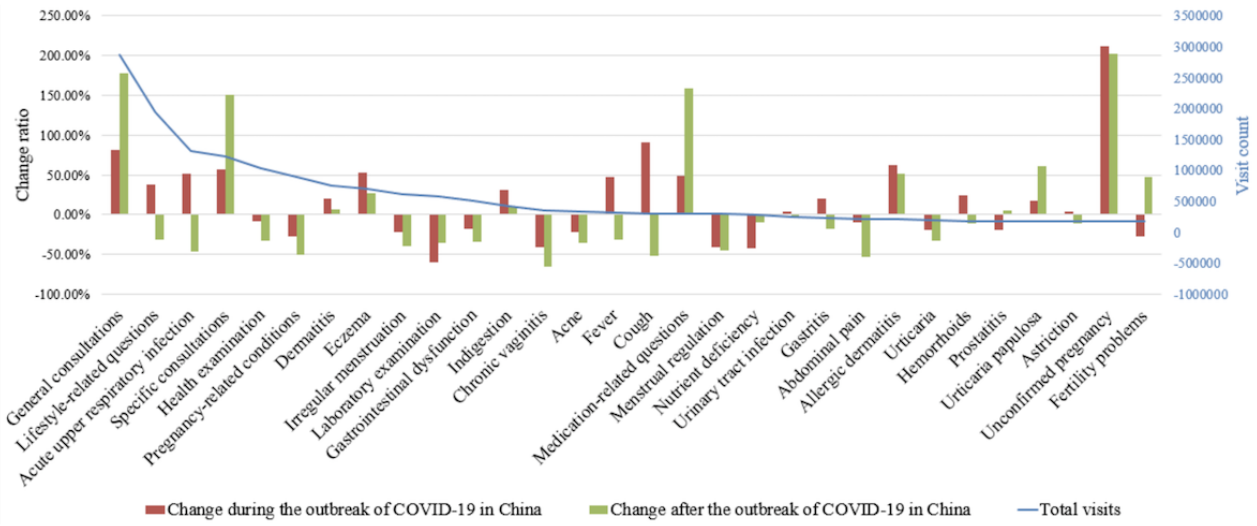
Figure 3. Departments represented in the consultations and inquisitions: (a) distribution of the departments; (b) change ratios of the distribution of the top 14 departments during and after the outbreak of COVID-19 in China.



Distribution of Diseases

There were 1419 diseases represented in our data set, and the top 30 diseases accounted for 53.6% (17,911,125/33,409,879) of the total consultations. As illustrated by the blue line in Figure 4, consultations were the key components in terms of yearly visits to Ping An Good Doctor; these included *general consultations*, *lifestyle-related questions*, *specific consultations*, *health examinations*, and *pregnancy-related conditions*. The change ratios of disease distribution during and after the COVID-19 outbreak were also analyzed. As shown by the red bars in Figure 4, diseases labeled as *unconfirmed pregnancy*, *cough*, *consultations*, and *specific consultations* increased the most during the COVID-19 outbreak. In addition, *eczema* and *allergic dermatitis* also had a significant increase. To summarize, diseases that increased in popularity during the COVID-19 outbreak, as represented in consultations and inquisitions, included those with respiratory or dermatology symptoms. After the COVID-19 outbreak, the inquisitions that included respiratory symptoms decreased, while the inquisitions that included characteristics of other diseases and consultations remained.

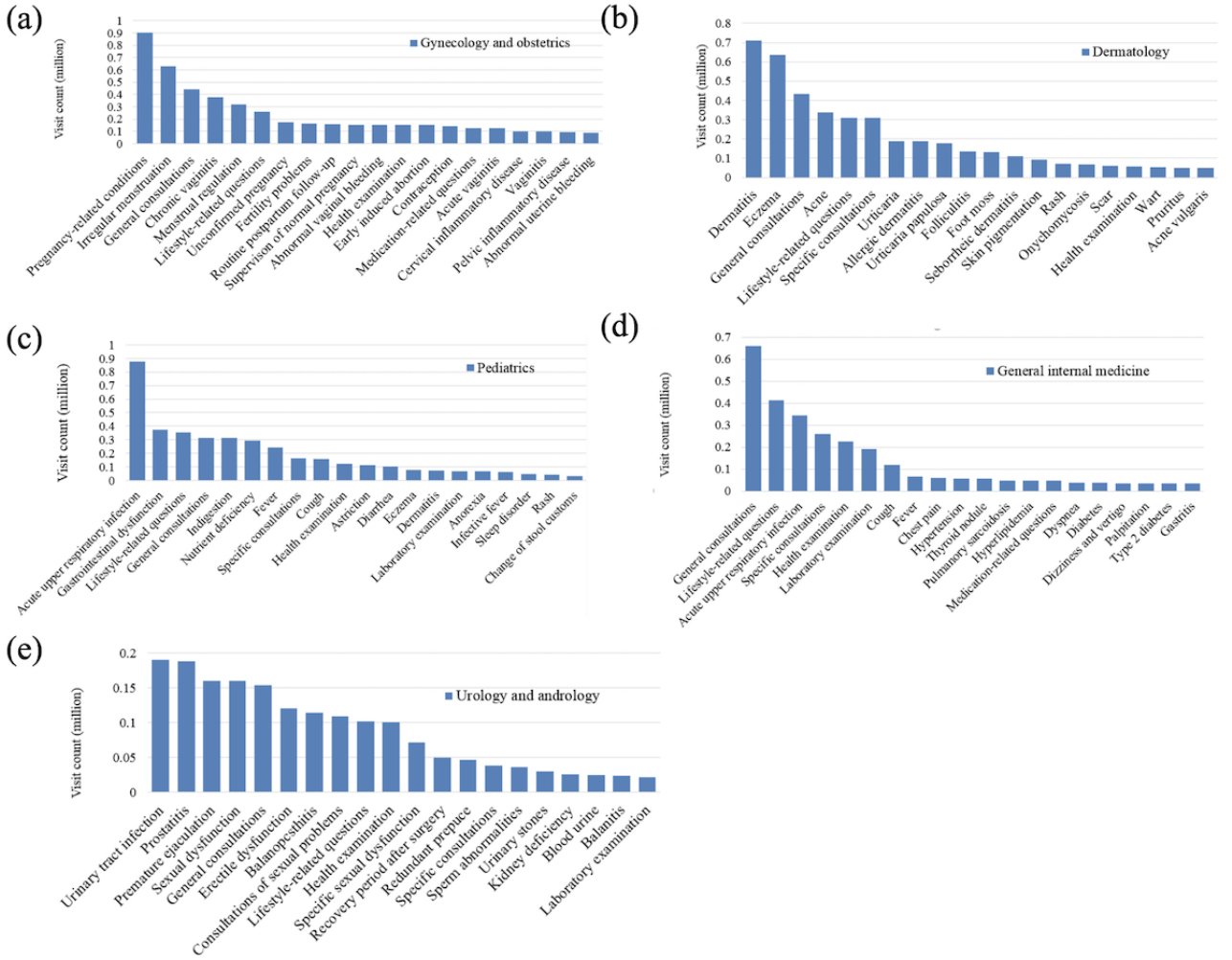
Figure 4. Distribution of the top 30 diseases represented in the yearly visits to Ping An Good Doctor; change ratios of the distribution of diseases during and after the outbreak of COVID-19 in China are also shown.



Disease Profiles for the Most Popular Departments

We further analyzed the disease profiles for the five most popular departments to see if major diseases existed in each department. The results are illustrated in Figure 5; for each department, the top 20 diseases are shown.

Figure 5. Disease profiles of the top five departments: (a) gynecology and obstetrics, (b) dermatology, (c) pediatrics, (d) general internal medicine, and (e) urology and andrology.



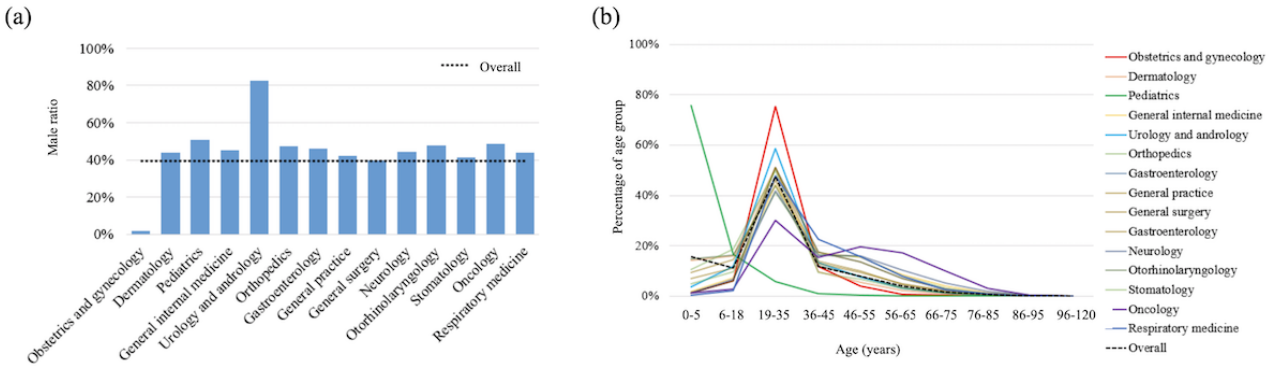
From Figure 5, pediatrics had the most obvious head effect regarding disease profiles, while acute upper respiratory infection (AURI) accounted for 16.5% (876,786/5,074,566) of all the consultations. In addition, urology and andrology had the least head effect regarding disease profiles, suggesting a relatively high degree of disease diversity.

User Profiles

Demographic Characteristics

To better understand the users of Ping An Good Doctor, we analyzed the user profiles for the top 14 departments. Male users accounted for 39.6% (12,098,998/30,579,308) of the consultations; this proportion varied little except for the reproductive-related departments, as shown in Figure 6 (a). For most of the departments, users aged 19 to 35 years constituted the largest proportion of users as compared with other age groups, as listed in Figure 6 (b). For pediatrics, there were many more consultations and inquisitions for children under 6 years of age than for children older than 6 years of age. For the respiratory medicine department , there was a relatively higher proportion of users aged 36 to 45 years as compared with the overall age groups. For oncology, the proportion of users in the senior age ranges increased.

Figure 6. User profiles for the top 14 departments: (a) percentiles of male users by department; (b) age distribution.

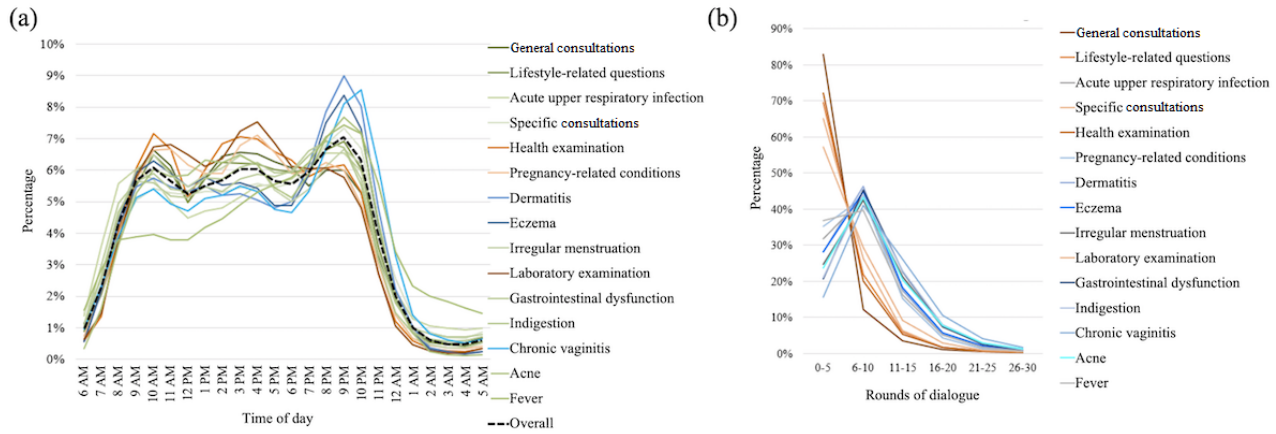


User Behaviors

The user profiles for the top diseases had concordant observations with those for the department analyses. We further investigated user consultation behaviors with respect to different diseases; namely, the times of day when users began to seek consultations and the rounds of dialogue between users and physicians.

As shown in Figure 7 (a), there were three peaks regarding times of day for consultations; these were 10 AM, 3 PM, and 9 PM. The diseases could be roughly divided into three groups according to the relative proportions for different time-of-day peaks as indicated in Figure 7 (a). Figure 7 (b) shows the distribution of rounds of dialogue between the users and physicians. For inquisitions, the number of rounds of dialogue peaked at around the *6 to 10* range; for consultations, at least half of the dialogues went through *0 to 5* rounds.

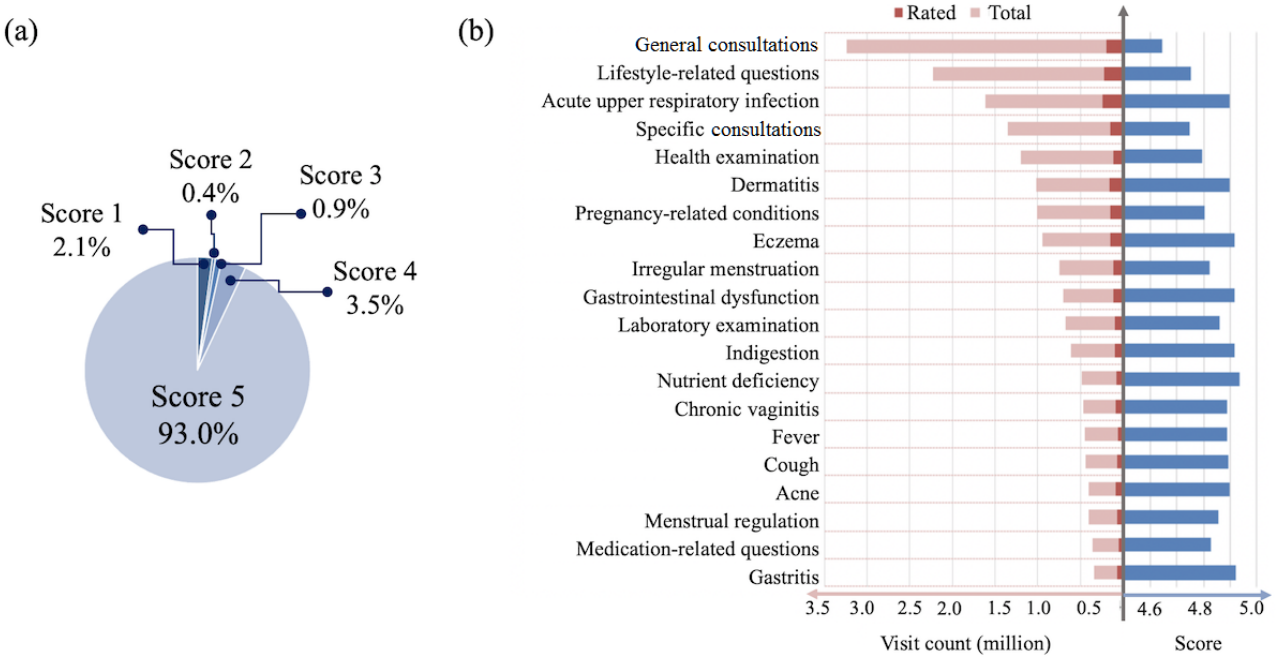
Figure 7. Characteristics of users' behaviors regarding the top 15 diseases: (a) distribution of times of day of consultations; (b) distribution of rounds of dialogue between users and physicians.



Users' Satisfaction With Online Consultations

Among all the 35.3 million consultations, there were 5.2 million that had satisfaction scores. A total of 93.0% (6,408,143/6,879,529) of these satisfaction scores were 5, suggesting that the vast majority of the users had a good experience when using Ping An Good Doctor, as seen in Figure 8 (a). We also analyzed user satisfaction in terms of different diseases. As shown in Figure 8 (b), consultations related to the top 20 diseases all had average satisfaction scores above 4.6. For some specific diseases, the average satisfaction scores were even higher, such as for AURI and dermatitis, which were also popular topics of consultation in our internet hospitals. For *general consultations* and *specific consultations*, the satisfaction scores were relatively low.

Figure 8. Users' satisfaction with the online consultations: (a) the distribution of satisfaction scores; (b) the proportion of users who scored the consultations and inquisitions and the corresponding scores related to the top 20 diseases. Scores range from 1 (least satisfaction) to 5 (most satisfaction).



Discussion

Principal Findings

Analyses in this study answered the three questions proposed in the Introduction section. First, the distributions of departments and diseases in Ping An Good Doctor differed from those of offline hospitals. The gynecology and obstetrics department and the dermatology department were the top two departments in Ping An Good Doctor in terms of number of consultations, which differed from offline hospitals. The popularity of gynecology and obstetrics came from about 22% of the users being female, aged 19 to 35 years, who were in the reproductive age group. The popularity of dermatology suggested that users found it suitable and more acceptable to seek consultations about dermatological diseases online. Consultations and diseases with moderate symptoms were commonly seen in terms of disease profiles. Second, we found that most of the users of Ping An Good Doctor were between 19 and 35 years of age, and females were slightly more highly represented compared to males. The gender distribution was in agreement with previous work by Qiu et al, though their results were based on one city with a much smaller population [34]. Last, the users had a good experience with Ping An Good Doctor, in general. We observed an average satisfaction score of 4.6 out of 5, suggesting that users were comfortable with Ping An Good Doctor as a health care service.

Our results provided the distribution profiles of departments and diseases. By examining those distributions, we could gain an increased understanding about which diseases can be diagnosed and treated online. The user profile analysis described the current user market of internet hospitals; a new strategy to explore the market should be based on this user profile analysis.

Open Attitudes Were the Key Factor for Developing Internet Hospitals

Based on the geographical distribution of the Ping An Good Doctor visits, we found that Shandong, Yunnan, Shaanxi, and Guangdong were the four leading provinces in terms of volume of visits; the reasons for this were four-fold. First, Shandong and Guangdong have large populations. Second, high-quality health care services are limited in these four leading provinces; Shandong, Yunnan, Shaanxi, and Guangdong ranked 17th, 24th, 20th, and 19th, respectively, in terms of number of tertiary hospitals per 100,000 persons, according to the China Health Statistics Yearbook 2020 [13]. Third, Guangdong and Shandong have a relatively high internet penetration rate [35]. Fourth, open attitudes toward internet technology were the key factor for developing internet hospitals. Guangdong was the pioneer in online health care services. The Guangdong government funded the first internet hospital in China [14]. Shandong has the most online physicians in a web-based medical consultation service in China, which suggests an open mind on the part of the physicians [36]. The Yunnan and Shaanxi governments also have an open attitude toward the internet-related industry and the public has a high acceptance of internet technology.

It was interesting to note that the usage of internet hospitals was not centered in big cities. For example, Beijing and Shanghai ranked 9th and 5th, respectively, among all the cities, although these two are megalopolitan cities with populations of more than 20 million. Kunming, the capital of Yunnan province with a population of around 7 million, had more yearly visits than Shanghai. The geographical distribution of yearly visits to Ping An Good Doctor suggested that internet hospitals provided accessible health care services to anyone in need and could relieve the uneven distribution of health care services across China.

Nondisease-Specific Consultations and Moderate Health Problems Were Popular in the Top Five Departments

The top five departments represented in Ping An Good Doctor were gynecology and obstetrics, dermatology, pediatrics, general internal medicine, and urology and andrology. As the characteristics varied across departments, we will discuss them one by one.

In gynecology and obstetrics, the most common consultations and inquisitions were focused on menstruation, vaginitis, and cyesis. These were mild health issues as compared with those in offline hospitals where cancers; cervical disease, such as hysteromyoma and ovarian cyst; and obstetric complications were seen in higher proportions. These profiles were consistent with the highest ratio of users who sought consultations in gynecology and obstetrics being aged 19 to 35 years, as seen in Figure 6 (b).

For dermatology-related concerns, the most common diseases were dermatitis and eczema, as seen in Figure 5 (b). The same pattern was also reported in other work [37]. It was interesting to note that disease inquisitions played a major role as compared with consultations. The most important reason for this was that the online health care service could replicate the offline clinical scenarios to the largest extent. Users were able to upload pictures of their skin, which played a vital role in diagnoses and treatments in the dermatology department. In addition, the efficiency and timeliness of online health services encouraged users to choose online inquisitions when they had dermatological problems. Online consultations also allowed patients to avoid the potential embarrassment of exposing themselves to physicians face-to-face.

In pediatrics, the most common inquisitions were about AURIs as well as growth and development, such as nutrient deficiency. Recent work on a pediatric map of Hangzhou also found that AURI was the top disease discussed in all the outpatient visits they considered. However, the next most popular category of diseases in their study was common symptoms, such as fever, abdominal pain, and vomiting, which was distinct from the pattern we found in the online health service [38]. Internet hospitals provide a way to efficiently and conveniently acquire knowledge about children's growth and development, relieving parents' anxiety. Educated with knowledge about their children's growth and development, parents can make proper and timely decisions when needed. Mild medical cases were able to be resolved online, which reduced unnecessary visits to offline hospitals and also relieved the stress experienced by them.

The diseases that were part of the general internal medicine department were mostly chronic, indicating that health management was in high demand. This was in agreement with previous findings from Di Tommaso et al [39]. Meanwhile, health management was vital for reducing fatality rates and disability rates. The proper management of chronic diseases could reduce the occurrences of complications and acute episodes. Traditionally, physicians mainly focus on chronic diseases at the time of patient in-person, offline visits; they have no time to conduct daily management for patients with chronic diseases. With developments in artificial intelligence and wearable devices, health care data are easy to collect and analyze. With well-educated patients and automatic tracking of physical measures, smart health care management has relieved physicians from tedious work and helped patients to maintain clinical conditions. Meanwhile, patients can benefit from advice on diet, exercise, and medication as given by smart health care management.

In the urology and andrology department, the consultations and inquisitions were focused on sexual function and urinary system infections. The online consultations were private and allowed users to avoid the embarrassment of face-to-face inquiries about sexual function problems. Urinary system infections were also common in both online and offline hospitals, but other frequent diseases seen in offline hospitals, such as cancer, urinary stones, and surgical trauma, were uncommon in online consultations.

To summarize, the common diseases seen in internet hospitals included moderate health problems. In this sense, internet hospitals play the role of general practitioner, which partially fills a gap that exists in the family doctor system in China.

Relationship Between the COVID-19 Epidemic and Internet Hospitals

The COVID-19 epidemic promoted widespread use of internet hospitals in China. From Table 4 and Figure 4, we observed that consultations and inquisitions did not decrease even after the first outbreak of COVID-19, from January 23, 2020, to February 29, 2020. COVID-19 provided people with an opportunity to change their behavior when seeking health care services. With the epidemic under control, the development of internet hospitals has normalized [40]. On the other hand, internet hospitals played a vital role in China's defense against COVID-19, from sharing knowledge on virus control and providing health consultations to follow-up treatment for chronic diseases and drug delivery. Ping An Good Doctor also played a positive role during the COVID-19 pandemic, suggesting that internet hospitals are a valuable resource in dealing with large-scale public health emergencies [41].

More Advantages of Internet Hospitals

In addition to the pros mentioned above, internet hospitals had many other advantages. First, internet hospitals had a positive effect on the physician-patient relationship [42]. Online health services made health information more accessible to patients, reducing the communication barrier between physicians and patients.

Second, for the physicians, providing health care services via internet hospitals was a way to not only increase their income but to widen and deepen their clinical experience of patients. The patients they met were not limited to their local regions. Meanwhile, physicians had the opportunity to inspect the whole disease process for the patients, from the onset to the prognosis.

Finally, internet hospitals could accelerate the reform of government health care insurance in China and possibly the world. Medical insurance costs increase year by year. The total cost reached 2085.4 billion yuan (US $299 billion) in 2019 in China, and had increased by 12.2% compared to 2018 [43]. Internet hospitals could help control four aspects of medical insurance costs. First, because the process of consultations and inquisitions was documented and all data were traced, it is feasible to conduct quality control and fraud detection. Second, equipped with artificial intelligence, standardized diagnosis and treatment procedures were simplified; for example, a robot helps to collect basic information before the consultations begin. Therefore, with increased efficiency, the cost of consultations in internet hospitals would decrease. Third, the whole process of health management in internet hospitals controls the rate of chronic diseases and, in the end, reduces medical insurance costs. China has a large population living with chronic diseases. Until 2019, there were 116.4 million people with diabetes in China [44]. The glucose control rate was 10.2% in China and was 49.9% in America [45]. A higher control rate is related to slower disease progression, lower comorbidity risk, lower disability rate, and lower costs. Efficient health management could help to increase the control rate and reduce medical costs. Fourth, internet hospitals could promote the separation of medical services and medications, which would prevent doctors from receiving rebates from drugs.

Limitations

We described the characteristics of internet hospitals in terms of the distributions of departments, diseases, user profiles, user behaviors, and user experiences. However, treatment—an important aspect of clinical practice—was not investigated and evaluated, but will be targeted in our future work. To deepen our understanding of the characteristics of internet hospitals, our future efforts will include additional comparisons between offline hospitals and internet hospitals.

Conclusions

As the largest entry point to online medical services in China, Ping An Good Doctor provides accessible health care services nationwide, breaking the geographical boundary for high-quality clinical resources. The COVID-19 epidemic accelerated the expansion of online health services. This also provided an opportunity to test the feasibility of such online modes of health care delivery. It should be emphasized that after the outbreak of COVID-19, the number of daily visits kept increasing, suggesting that online health services satisfied peoples' needs for convenient and contactless health services. Our analyses showed that nondisease-specific consultations and moderate health problems accounted for the majority of online visits to Ping An Good Doctor. This indicated that online health services played a similar role as family doctors. The positive feedback received from most users showed that online health services would be their preferred option in many circumstances. The changes in people's attitudes and behaviors, together with

support from the government, would further promote the development of online health services, leading to great benefits for both patients and health care providers.

Conflicts of Interest

None declared.

References

1. Notice on the issuance of three documents including the Administrative Measures for Internet Diagnosis and Treatment (trial). National Health Commission of the People's Republic of China and Pharmaceutical Administration. 2018 Jul 17. URL: http://www.gov.cn/zhengce/zhengceku/2018-12/31/content_5435436.htm [accessed 2021-04-06]
2. The National Telehealth Policy Resource Center. What is telehealth? The National Telehealth Policy Resource Center. URL: https://www.cchpca.org/about/about-telehealth [accessed 2020-12-24]
3. Center for Connected Health Policy. State Telehealth Laws & Reimbursement Policies. Oakland, CA: Public Health Institute; 2020. URL: https://www.cchpca.org/sites/default/files/2020-05/CCHP_%2050_STATE_REPORT_SPRING_2020_FINAL.pdf [accessed 2020-12-24]
4. Carrasqueiro S, Ramalho A, Esteves A, Pereira C, Martins D. Report on EU State of Play on Telemedicine Services and Uptake Recommendations.: Joint Action to Support the eHealth Network; 2017. URL: https://ec.europa.eu/health/sites/health/files/ehealth/docs/ev_20171128_co09_en.pdf [accessed 2020-12-20]
5. Edwards HB, Marques E, Hollingworth W, Horwood J, Farr M, Bernard E, et al. Use of a primary care online consultation system, by whom, when and why: Evaluation of a pilot observational study in 36 general practices in South West England. BMJ Open 2017 Nov 22;7(11):e016901 [FREE Full text] [doi: 10.1136/bmjopen-2017-016901] [Medline: 29167106]
6. Scott R, Mars M. Telehealth in the developing world: Current status and future prospects. Smart Homecare Technol Telehealth 2015 Feb 2:25-37 [FREE Full text] [doi: 10.2147/shtt.s75184]
7. Xie X, Zhou W, Lin L, Fan S, Lin F, Wang L, et al. Internet hospitals in China: Cross-sectional survey. J Med Internet Res 2017 Jul 04;19(7):e239 [FREE Full text] [doi: 10.2196/jmir.7854] [Medline: 28676472]
8. Han Y, Lie RK, Guo R. The internet hospital as a telehealth model in China: Systematic search and content analysis. J Med Internet Res 2020 Jul 29;22(7):e17995 [FREE Full text] [doi: 10.2196/17995] [Medline: 32723721]
9. Tu J, Wang C, Wu S. The internet hospital: An emerging innovation in China. Lancet Glob Health 2015 Aug;3(8):e445-e446. [doi: 10.1016/S2214-109X(15)00042-X]
10. China builds 158 internet hospitals amid digital tide. China Daily. 2019 May 08. URL: http://www.chinadaily.com.cn/a/201905/08/WS5cd29062a3104842260ba8f3.html [accessed 2020-10-24]
11. China embraces internet hospitals, with 900 now operational: Official. Xinhuanet. 2020 Oct 28. URL: http://www.xinhuanet.com/english/2020-10/28/c_139473917.htm [accessed 2020-12-18]
12. Yu M, He S, Wu D, Zhu H, Webster C. Examining the multi-scalar unevenness of high-quality healthcare resources distribution in China. Int J Environ Res Public Health 2019 Aug 07;16(16):2813 [FREE Full text] [doi: 10.3390/ijerph16162813] [Medline: 31394765]
13. National Health Commission of the People's Republic of China. China Health Statistics Yearbook 2020. Beijing, China: Peking Union Medical College Press; Oct 1, 2020.
14. Wu D, Hesketh T, Shu H, Lian W, Tang W, Tian J. Description of an online hospital platform, China. Bull World Health Organ 2019 Jun 03;97(8):578-579 [FREE Full text] [doi: 10.2471/blt.18.226936]
15. Feng XL, Martinez-Alvarez M, Zhong J, Xu J, Yuan B, Meng Q, et al. Extending access to essential services against constraints: The three-tier health service delivery system in rural China (1949-1980). Int J Equity Health 2017 May 23;16(1):49 [FREE Full text] [doi: 10.1186/s12939-017-0541-y] [Medline: 28532500]
16. Gong P, Liang S, Carlton EJ, Jiang Q, Wu J, Wang L, et al. Urbanisation and health in China. Lancet 2012 Mar;379(9818):843-852. [doi: 10.1016/s0140-6736(11)61878-3]
17. Li X, Krumholz HM, Yip W, Cheng KK, De Maeseneer J, Meng Q, et al. Quality of primary health care in China: Challenges and recommendations. Lancet 2020 Jun 6;395(10239):1802-1812 [FREE Full text] [doi: 10.1016/s0140-6736(20)30122-7]
18. Yuan S, Wang F, Li X, Jia M, Tian M. Facilitators and barriers to implement the family doctor contracting services in China: Findings from a qualitative study. BMJ Open 2019 Oct 08;9(10):e032444 [FREE Full text] [doi: 10.1136/bmjopen-2019-032444] [Medline: 31597653]
19. Shang X, Huang Y, Li B, Yang Q, Zhao Y, Wang W, et al. Residents' awareness of family doctor contract services, status of contract with a family doctor, and contract service needs in Zhejiang Province, China: A cross-sectional study. Int J Environ Res Public Health 2019 Sep 09;16(18):3312 [FREE Full text] [doi: 10.3390/ijerph16183312] [Medline: 31505783]
20. Healthy China 2030 (from vision to action). World Health Organization. 2016. URL: https://www.who.int/healthpromotion/conferences/9gchp/healthy-china/en/ [accessed 2020-10-30]

21. The National Health Commission strengthens internet diagnosis and treatment consulting services in epidemic prevention and control. National Health Commission of the People's Republic of China. 2020 Feb 08. URL: http://www.nhc.gov.cn/yzygj/s7652m/202002/32c3e98988894fa18280e4543d2710c7.shtml [accessed 2020-12-18]
22. Gong K, Xu Z, Cai Z, Chen Y, Wang Z. Internet hospitals help prevent and control the epidemic of COVID-19 in China: Multicenter user profiling study. J Med Internet Res 2020 Apr 14;22(4):e18908 [FREE Full text] [doi: 10.2196/18908] [Medline: 32250962]
23. Li L, Liu G, Xu W, Zhang Y, He M. Effects of internet hospital consultations on psychological burdens and disease knowledge during the early outbreak of COVID-19 in China: Cross-sectional survey study. J Med Internet Res 2020 Aug 04;22(8):e19551 [FREE Full text] [doi: 10.2196/19551] [Medline: 32687061]
24. Ding L, She Q, Chen F, Chen Z, Jiang M, Huang H, et al. The internet hospital plus drug delivery platform for health management during the COVID-19 pandemic: Observational study. J Med Internet Res 2020 Aug 06;22(8):e19678 [FREE Full text] [doi: 10.2196/19678] [Medline: 32716892]
25. Wang O. Ping An Good Doctor. 2015 Dec. URL: http://www.pingan.com/app_upload/images/info/upload/5789191e-3019-4370-b8e1-a3a81561d4e3.pdf [accessed 2020-10-01]
26. Ping An Good Doctor issues sustainable development report: Platform visits hit 1.11 billion during epidemic. Cision PR Newswire. 2020 Feb 17. URL: https://www.prnewswire.com/news-releases/ping-an-good-doctor-issues-2019-sustainable-development-report-platform-visits-hit-1-11-billion-during-epidemic-301005828.html [accessed 2020-10-01]
27. Internet medical industry: The big health battle, the "three kingdoms" of JD, Alibaba, and Ping An. Xiaoyi. 2020 Dec 20. URL: https://mp.weixin.qq.com/s/bFwo5qSjMTdA4noG5RZR2A [accessed 2020-12-20]
28. The 2020 China internet medical app evaluation report is released!. Southern Metropolis Daily. 2020 Dec 07. URL: https://m.mp.oeeee.com/a/BAAFRD000020201206387475.html [accessed 2020-12-20]
29. JD Health: On the first day of listing, the average number of consultations per day was only one-eighth of Ping An Good Doctor. Bread Finance. 2020 Dec 09. URL: https://finance.sina.com.cn/stock/relnews/hk/2020-12-09/doc-iiznctke5531516.shtml [accessed 2020-12-20]
30. Compared with Ali Health and Ping An Good Doctor, what are the advantages of JD Health? Futubull. 2020 Nov 26. URL: https://news.futunn.com/market/137820?src=3 [accessed 2020-12-20]
31. National Center for Health Statistics. International Classification of Diseases (ICD-10-CM/PCS) transition - Background. Centers for Disease Control and Prevention. 2015. URL: https://www.cdc.gov/nchs/icd/icd10cm_pcs_background.htm [accessed 2020-10-01]
32. Moritsugu K, Wang Y. China closes off large city of Wuhan to stop spread of deadly virus. The Times of Israel. 2020 Jan 23. URL: https://www.timesofisrael.com/china-closes-off-large-city-of-wuhan-to-stop-spread-of-deadly-virus/ [accessed 2020-12-20]
33. Xinhua News Agency. White paper on "China's action against the new coronary pneumonia epidemic". China International Development Cooperation Agency. 2020 Jun 07. URL: http://www.cidca.gov.cn/2020-06/07/c_1210650054.htm [accessed 2020-12-20]
34. Qiu Y, Liu Y, Ren W, Qiu Y, Ren J. Internet-based and mobile-based general practice: Cross-sectional survey. J Med Internet Res 2018 Sep 25;20(9):e266 [FREE Full text] [doi: 10.2196/jmir.8378] [Medline: 30257819]
35. The 46th China Statistical Report on Internet Development. Beijing, China: China Internet Network Information Center; 2020. URL: http://www.gov.cn/xinwen/2020-09/29/5548176/files/1c6b4a2ae06c4ffc8bccb49da353495e.pdf [accessed 2020-12-20]
36. Li Y, Yan X, Song X. Provision of paid web-based medical consultation in China: Cross-sectional analysis of data from a medical consultation website. J Med Internet Res 2019 Jun 03;21(6):e12126 [FREE Full text] [doi: 10.2196/12126] [Medline: 31162129]
37. Wang X, Yang H, Wang Y, Yuan J, Yu Q. Analysis of cases managed by dermatological service in the Chinese peacekeeping level 1+ medical treatment facility in Lebanon, 2018-2019. BMJ Mil Health 2020 Feb 27:1-4 [FREE Full text] [doi: 10.1136/jramc-2019-001315] [Medline: 32111673]
38. Li H, Yu G, Dong C, Jia Z, An J, Duan H, et al. PedMap: A pediatric diseases map generated from clinical big data from Hangzhou, China. Sci Rep 2019 Nov 28;9(1):17867 [FREE Full text] [doi: 10.1038/s41598-019-54439-w] [Medline: 31780760]
39. Di Tommaso MR, Spigarelli F, Barbieri E, Rubini L. On health production and demand: And why an effective health industry is vital for China (and the rest of the world). In: The Globalization of China's Health Industry: Industrial Policies, International Networks and Company Choices. Cham, Switzerland: Palgrave Macmillan; 2020:1-41.
40. Zhang X. 2020 Internet Hospital Report. Vcbeat. 2020 Aug 05. URL: https://vcbeat.top/MmIyMDA1ZWU2NDFiOTJhYzI1ZGZlZjJmMGZhYjFlN2Q= [accessed 2020-12-18]
41. The healthcare market in China. Daxue Consulting. 2020 May 19. URL: https://daxueconsulting.com/healthcare-market-china/ [accessed 2020-10-30]
42. Tan SS, Goonawardene N. Internet health information seeking and the patient-physician relationship: A systematic review. J Med Internet Res 2017 Jan 19;19(1):e9 [FREE Full text] [doi: 10.2196/jmir.5729] [Medline: 28104579]

43. 2019 Statistical Bulletin of the National Medical Security Development. Beijing, China: National Healthcare Security Administration; 2020 Jun 24. URL: http://www.nhsa.gov.cn/art/2020/6/24/art_7_3268.html [accessed 2020-12-20]
44. IDF Diabetes Atlas, Ninth edition. Brussels, Belgium: International Diabetes Federation; 2019. URL: https://www.diabetesatlas.org/upload/resources/material/20200302_133351_IDFATLAS9e-final-web.pdf [accessed 2020-12-20]
45. View Point: Analysis and Prospect of Chronic Disease Drug Market. Shanghai, China: IQVIA; 2019 Feb. URL: https://www.iqvia.com/-/media/iqvia/pdfs/china/thought-leadership/view-point/iqvia-viewpoint-issue-41-chronic-disease-management.pdf?_=1617705413229 [accessed 2021-04-06]

Abbreviations

AURI: acute upper respiratory infection
ICD-10: International Classification of Diseases, Tenth Revision
TCM: Traditional Chinese Medicine

Edited by G Eysenbach; submitted 17.11.20; peer-reviewed by R Guo, L Ding; comments to author 09.12.20; revised version received 30.12.20; accepted 01.03.21; published 15.04.21.

Please cite as:
Jiang X, Xie H, Tang R, Du Y, Li T, Gao J, Xu X, Jiang S, Zhao T, Zhao W, Sun X, Hu G, Wu D, Xie G
Characteristics of Online Health Care Services From China's Largest Online Medical Platform: Cross-sectional Survey Study
J Med Internet Res 2021;23(4):e25817
URL: https://www.jmir.org/2021/4/e25817
doi:10.2196/25817
PMID:33729985

Original Paper

Patients' Experiences of a Nurse-Led, Home-Based Heart Failure Self-management Program: Findings From a Qualitative Process Evaluation

Ying Jiang[1], PhD; Karen Wei Ling Koh[2], PhD; Hadassah Joann Ramachandran[1], BSc (Hons); Yee Kian Tay[3], MN; Vivien Xi Wu[1], PhD; Shefaly Shorey[1], PhD; Wenru Wang[1], PhD

[1]Alice Lee Centre for Nursing Studies, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore

[2]National University Heart Centre Singapore, National University Hospital, Singapore, Singapore

[3]Regional Health System, National University Health System, Singapore, Singapore

Corresponding Author:
Wenru Wang, PhD
Alice Lee Centre for Nursing Studies
Yong Loo Lin School of Medicine
National University of Singapore
Level 2, Blk MD11, 10 Medical Drive
Singapore, 117597
Singapore
Phone: 65 66011761
Email: nurww@nus.edu.sg

Abstract

Background: Heart failure (HF) is a major public health problem that places a significant disease burden on society. Self-care is important in the management of HF because it averts disease progression and reduces the number of hospitalizations. Effective nursing interventions promote HF self-care.

Objective: This study aims to explore participants' perspectives on a nurse-led, home-based heart failure self-management program (HOM-HEMP) in a randomized controlled trial conducted in Singapore to gain insight into the effectiveness of the study intervention.

Methods: A descriptive, qualitative approach was used. English- or Chinese-speaking participants from the intervention arms were recruited through a purposive sampling method from January 2019 to July 2019. Individual, face-to-face, semistructured interviews were conducted with 11 participants. All interviews were audio recorded and transcribed verbatim, with the participant identifiers omitted to ensure confidentiality. The thematic analysis approach was used to identify, analyze, and report patterns (themes) within the data.

Results: A total of six themes emerged from the process evaluation interviews and were categorized according to the Donabedian structure-process-outcome framework as *intervention structure*, *intervention process*, and *intervention outcome*. These six themes were manageability of the intervention, areas for improvement, benefits of visiting, personal accountability in self-care, empowered with knowledge and skills in self-care after the intervention, and increased self-efficacy in cardiac care.

Conclusions: The findings of the process evaluation provided additional information on participants' perceptions and experiences with the HOM-HEMP intervention. Although a home visit may be perceived as resource intensive, it remains to be the preferred way of engagement for most patients. Nurses play an important role in promoting HF self-care. The process of interaction with the patient can be an important process for empowering self-care behavior changes.

(J Med Internet Res 2021;23(4):e28216) doi:10.2196/28216

KEYWORDS

self-care; psychosocial educational; nurse-led; mHealth; self-management; heart failure; process evaluation; nursing; mobile phone

Introduction

Background

Heart failure (HF) is a major public health problem that places a significant disease burden on society. It is one of the most common causes of hospital admissions and readmissions, leading to an increased demand for hospital beds, staff resources, subsidies, and insurance coverage [1]. Southeast Asians are at particular risk of HF because of the high prevalence of risk factors such as hypertension and diabetes. They also have an earlier onset of HF than Americans and Europeans [2]. Furthermore, the incidence and prevalence of HF are closely related to age. As one gets older, the condition becomes more common [3]. Singapore has one of the fastest aging populations among developed countries worldwide. Therefore, it is expected that the number of people with HF is likely to surge in the near future.

To sustain the health care system, the Ministry of Health in Singapore recently announced three paradigm shifts: beyond hospital to community, beyond quality to value, and beyond health care to health [4]. With this change, more patients with chronic conditions will be managed in the community, and nurses will need to embrace new knowledge and skills that match the care needs of their patients. Simultaneously, as hospital stays become shorter and less frequent, there is a need for patients and families to acquire disease self-care skills to achieve the success of managing their conditions outside the hospital.

In HF, self-care encompasses a range of behaviors, including adhering to medication and treatment, avoiding excessive fluid and salt intake, monitoring daily weight, engaging in exercise, monitoring and identifying exacerbating symptoms, and taking appropriate steps to intervene if symptoms worsen [5]. Self-care is the cornerstone of HF management. Patients who practice constant self-care have been shown to avert 30% of hospital admissions and more than half of the readmissions [6]. Despite its importance, several studies have reported gaps in patients' knowledge and skills regarding HF self-care, with many patients finding HF self-care challenging, especially when transitioning from hospital care to home and community care [7,8]. Traditionally, hospital nurses have provided disease education with information on the value of medication adherence and lifestyle risk modification. Most of these education programs are done as part of discharge instructions and/or as part of inpatient education. However, the severity of symptoms during stressful hospital stays (eg, breathlessness and fatigue) precludes patients from being physically and psychologically ready to absorb any information. As a result, many patients feel underinformed when they return home [8].

We recently conducted a 3-arm stratified randomized controlled trial to evaluate the effectiveness of a newly developed, nurse-led, home-based psychosocial and educational intervention (home-based heart failure self-management program [HOM-HEMP]) for patients with HF in Singapore (ClinicalTrials.gov NCT03108235) [Jiang Y, unpublished data, 2021]. The development of the intervention and its detailed description have been published previously [9]. In brief, the HOM-HEMP intervention is a 6-week, multicomponent self-management program designed to promote HF self-care. It consisted of an HF patient education and self-care toolkit and three biweekly home visits by the research nurse. The toolkit included an HF education manual, a drinking cup marked with fluid volume, picture cards of common local foods and their associated sodium content, a weight scale, a weight monitoring calendar, and a pillbox with an alarm clock that reminded patients to take their medications (Multimedia Appendix 1). The toolkit provided cues (eg, a visual reminder of fluid intake and alarm reminder for medication adherence) of self-care activities in the context of daily life, whereas biweekly home visits provided an opportunity for the research nurse to interact with patients face-to-face, during which the research nurse also employed motivational interviewing to promote patient self-care engagement. In addition, a supplementary smartphone app was developed based on the study intervention (Multimedia Appendix 2), which was only available to participants in the experimental group B (one of the intervention arms). The smartphone app offered features such as educational content; customized scheduled medication and medical appointment reminders; weight, blood pressure, and symptom logs; and a chat room to communicate with the research nurse via text message. The research nurse reviewed the patient's symptoms on the back end through a secured web portal.

The aims of the research project are to evaluate the effectiveness of the HOM-HEMP intervention in improving patients' HF self-care, cardiac self-efficacy, psychological well-being, health-related quality of life, social support, and clinical outcomes. Consenting participants were randomly assigned to a control group, an experimental group A, and an experimental group B. All study participants received usual care provided by the study hospital. Participants in the experimental group A and experimental group B received usual care and the HOM-HEMP intervention. In addition, participants in the experimental group B received a supplemental smartphone app. The purpose of having 2 experimental groups was to explore the effects of the mobile health (mHealth) component on the study outcomes. The findings of this trial are reported elsewhere [Jiang Y, unpublished data, 2021].

Objectives

This study details the process evaluation of the study intervention, which is part of the HOM-HEMP research project [9]. Process evaluations are recommended by the UK Medical Research Council guidance to provide additional information to understand how an intervention might or might not work in a specific context or whether the outcomes of the trial can be reproduced [10]. It is an important step in the evaluation of a complex intervention, as it allows researchers and practitioners to understand whether improvements in outcomes have resulted from the intended intervention process, whether the intervention causes poor outcomes, or even sometimes whether good outcomes are produced by the intervention that is actually less satisfactory [10].

To better understand participants' views and experiences regarding the HOM-HEMP intervention provided to them, the Donabedian structure-process-outcome framework was

employed in the process evaluation [11]. In his model, Donabedian proposed *structure*, *process*, and *outcome* of care as the key aspects of quality of care [12]. The framework recognizes that outcomes are associated with the relationships between the context and actions of patients and care providers and that understanding the implementation of an intervention requires not only what is delivered but also the mechanism through which the intervention is delivered [11]. In this study, *structure* refers to physical components of the intervention and the organization of the intervention, such as the duration and frequency of the intervention; *process* refers to the participants' experiences of interactions and encounters during the study intervention; and *outcome* refers to changes in participants as a result of the intervention and/or interactions with the research nurse during the intervention.

The aim of this process evaluation is to explore participants' views and experiences with regard to the HOM-HEMP intervention and to understand the mechanism by which the intervention had an impact.

Methods

Study Design and Participants

A descriptive, qualitative approach was used. One-to-one semistructured interviews were conducted to elicit participants' responses to their overall experiences with the HOM-HEMP intervention. An interview guide was developed to guide the one-to-one semistructured interviews and to ensure uniform data collection across all participants. The questions in the interview guide were reviewed by the members of the researcher's team, all of whom had experience in conducting qualitative research. The semistructured interview offers some flexibility that allows participants to voice out their views in a naturalistic narrative fashion while ensuring that all question areas were discussed with each participant [13]. Compared with group interviews, the one-to-one interview format provides a less threatening environment for the participants to share their personal views and/or describe their experiences [13].

Participants were recruited through a purposive sampling method from January 2019 to July 2019 at a tertiary public hospital in Singapore. Participants from the intervention arms with either positive or negative differences in any of the study outcomes were included in the process evaluation if they were willing to be audio recorded. The inclusion criteria for the HOM-HEMP research project were patients who were aged ≥21 years, had been clinically diagnosed with HF, were able to read and understand English or Chinese, owned and used a smartphone in their everyday lives, and were able to be followed up at home after discharge from the hospital. The exclusion criteria were patients who had unstable angina, resting tachycardia, or severe arterial hypertension; had a terminal illness other than HF that affects their self-care ability and/or completion of the study, for example, end-stage cancer; had a psychiatric condition or impaired cognitive functioning that affects their understanding of the study intervention; were bed- or wheelchair-bound affecting their mobility and self-care ability; and had no internet access at home. To ensure a diverse sample of participants, participants' ethnicity, age, and socioeconomic status were taken into consideration when choosing the cases in this study to identify important patterns across variations [13].

Sample Size

There is no established rule for sample size in qualitative evaluations. In this study, the sample size was determined based on information needs and the concept of data saturation [13]. Recruitment for the process evaluation interview stopped when no new information emerged and redundancy was achieved.

Data Collection

At the end of the HOM-HEMP study, participants from either experimental group A or experimental group B were approached based on the purposive sampling criteria. Details regarding the interview process and objectives were explained to the participants. In addition, they were informed that the session would be audiotaped. Written informed consent was obtained from the participants before the interviews commenced. Interviews were arranged at a place convenient for the participant.

Qualitative data were obtained from individual, face-to-face, semistructured interviews. All interviews were conducted by a research nurse who delivered the study intervention. Interviews were conducted in either English or Chinese, depending on the participants' language preference. Before the start of the interview, the research nurse reassured the participants that all the information they provided would be kept confidential and that the interview would be like a conversation with no right or wrong answers to the questions asked. Rapport was established by one or two warm-up questions before putting forward the topics on the interview guide. Open-ended questions were used to encourage participants to talk freely on the questions asked. If the participants' family members were present during the process evaluation interview, their role was limited to helping to relay the point the participant was trying to make when the participant had difficulties in explaining his or her point of view clearly. All interviews were audio recorded and transcribed verbatim. The participant identifiers were omitted from the transcripts to maintain confidentiality. The transcripts were cross-checked by a second researcher to ensure accuracy.

Ethical Considerations

Ethical approval was obtained from National Healthcare Group-Domain Specific Review Board (NHG DSRB ref: 2017/00249) before the commencement of the HOM-HEMP research project. Voluntary participation, confidentiality, the right to withdraw, and potential risks and benefits were explained. Informed consent was obtained from each participant before the study commenced. Participants were also reassured that declining to participate in the study would not lead to any penalties or any differences in treatment or care.

Data Analysis

Descriptive data analysis was performed to profile the sample characteristics based on participants' sociodemographic and clinical data. Frequencies and percentages were used to describe the categorical variables, such as gender, race, marital status, highest education level, employment status, types of housing,

smoking status, the New York Heart Association functional classification, and the presence of comorbidities.

The thematic analysis approach was used to identify, analyze, and report the patterns (themes) within the data, as they are not reliant on any theoretical framework and provide sufficient flexibility for data analysis [14]. Specifically, the thematic analysis process involved five different phases, including familiarizing the data, generating initial codes, collating codes into potential themes, reviewing themes, and defining and naming those themes [14]. The transcripts were coded independently by the research nurse and a second researcher. The results were cross-checked for consistency and revised as required. The primary coding and collation of codes into themes were manually performed. To ensure that themes were formed coherently and accurately, the initial themes were reviewed and refined based on discussions with team members. Illustrative quotations were translated into English if they were in Chinese. Repeated words and grammatical errors were edited for ease of understanding, but no substantive changes were made. The translations were cross-checked by two researchers proficient in both English and Chinese.

Trustworthiness

The trustworthiness of the qualitative approach in the process evaluation was established by ensuring the feasibility and adequacy of the semistructured interview guide in eliciting appropriate data to achieve the evaluation objectives. The questions in the interview guide were reviewed by members of the research team and the author's supervisor, all of whom had experience in conducting qualitative research. In addition, to ensure the accuracy and authenticity of the data, the interviews were audio recorded and transcribed verbatim. The transcribed data and data analysis were cross-checked by two researchers, and thematic agreement was reached through discussion. These steps provided a complete and accurate description of the participants' responses, thus ensuring the credibility of the process evaluation. The confirmability was to ensure that the themes were a true representation of the participants' experiences and perceptions [13]. It was established by recording the participants' quotations to support the results. Auditability was established by keeping records of the data analysis process, such as records of the evolution of findings and emergence of themes. Finally, transferability was established by providing a detailed research context to allow readers to relate to the conclusions drawn from this study.

Results

Sociodemographic Data and Clinical Characteristics of the Participants

A total of 11 participants were approached and invited to participate in the process evaluation interview, of which 6 were from experimental group A and 5 were from experimental group B. The average duration of the interviews was 9.24 minutes, ranging from 5 to 17 minutes. Of 11 participants, 7 (82%) were Chinese speaking. The descriptive statistics of the participants' baseline sociodemographic data and clinical characteristics are presented in Table 1. Of the 11 participants, 7 (64%) were male and 4 (36%) were female. Their age ranged from 44 to 85 years (mean 64.45, SD 12.58 years). Of the 11 participants, 6 (55%) participants were still working full time or on a part-time basis. None of the participants drank alcohol, but 9% (1/11) of the participants had smoked in the past 4 weeks. From the 11 participants, 9 (82%) had coronary heart disease, 7 (64%) had diabetes, 6 (55%) had hypertension, and 7 (64%) had high cholesterol.

Table 1. Sociodemographic data and clinical characteristics of the participants (N=11).

Variables	Values
Age (years), mean (SD)	64.45 (12.58)
Gender, n (%)	
Male	7 (64)
Female	4 (36)
Race, n (%)	
Chinese	10 (91)
Indian	1 (9)
Marital status, n (%)	
Married	9 (82)
Unmarried, widowed, divorced, or separated	2 (18)
Highest education level, n (%)	
Primary school or no formal education	7 (64)
Tertiary (university, ITE[a], polytechnic, or junior college)	4 (36)
Employment status, n (%)	
Working part time or full time	6 (55)
Not working or retired	5 (45)
Monthly household income, SGD[b] (US $), n (%)	
<1000 (745)	6 (55)
1000-3000 (745-2236)	1 (9)
3001-5000 (2237-3727)	2 (18)
>5000 (3727)	2 (18)
Main caregiver, n (%)	
Family	8 (73)
Other (maid, friend, or alone)	3 (27)
Type of housing, n (%)	
2-3 rooms	3 (27)
4-5 rooms	6 (55)
Executive maisonette, private, or landed	2 (18)
Ex-smoker, n (%)	2 (18)
Smoking (in the past 4 weeks), n (%)	1 (9)
Alcohol drinking, n (%)	0 (0)
The New York Heart Association Functional Classification, n (%)	
I-II (mild to moderate)	1 (9)
III-IV (severe)	10 (91)
Coronary heart disease, n (%)	9 (82)
Type 2 diabetes, n (%)	7 (64)
Hypertension, n (%)	6 (55)
High cholesterol, n (%)	7 (64)

[a]ITE: institute of technical education.

[b]SGD: Singapore dollars.

Qualitative Findings

The findings are organized and presented based on the Donabedian framework. A total of six underlying themes representing participants' overall experience and perceptions of the study intervention were identified. They were grouped under *intervention structure*, *intervention process*, and *intervention outcome* (Table 2). The emerging themes were manageability of the intervention, areas for improvement, benefits of visiting, personal accountability in self-care, empowered with knowledge and skills in self-care after the intervention, and increased self-efficacy in cardiac care.

Table 2. Themes and subthemes.

Themes	Subthemes
Intervention structure	
1. Manageability of the intervention	• Ease of learning • Essential topics were covered • Convenient to participate • Different views on the use of technology
2. Areas for improvement	• Participants' different preferences for frequency and duration of intervention • Intervention format can be more flexible • Some of the self-care tools did not match the participants' usage habits • To include topics related to chronic diseases
Intervention process	
3. Benefits of visiting	• Allow patients to better express themselves • Preferred personalized advice over information from other resources • Felt nurses can better understand patients' problems • Reassurance for family members • Worries and burden are being addressed
4. Personal accountability in self-care	• You have to help yourself first • Self-discipline is required
Intervention outcomes	
5. Empowered with knowledge and skills in self-care after the intervention	• Monitoring inculcated in daily life • Able to sustain effort after intervention
6. Increase self-efficacy in cardiac care	• Increased initiative • Improved self-awareness of health status • Determination to change a certain lifestyle

Intervention Structure

Theme 1: Manageability of the Intervention

Most participants reported that overall, the intervention was easy to manage with no added burden. It covered all the essential topics that participants wanted to know with an adequate amount of information. The most helpful aspects of the intervention included coaching on weight monitoring and controlling fluid and salt intake. One participant mentioned that it was important that the intervention be as simple as possible so that it would help them apply what they had learned easily:

> *...[the programme] is really very good, because I have learnt a lot of things that I did not know last time, like how to read the sodium amount in the food label...* [P2, aged 46 years]

> *...[the most helpful aspects] are weight monitoring, coz it tells me if I have water retention...also controlling my water intake and diet...I try to keep my water intake to 1L/day, and eat less oily and salty food...* [P6, aged 76 years]

> *I think the best is simple and useful...yes...because we need to work, if it is too complex, we feel very troublesome. So, I think the simpler the better...And also, we are not that young now, sometimes we cannot remember so many things.* [P9, aged 59 years]

In addition, home visits reduced the travel needs of the participants. Convenience of participation through home visits encouraged participants to participate in the study. For some participants, the flexibility of scheduled home visits allowed them to plan ahead between different commitments:

> *Look, instead of you asking patients come over [to the hospital], you are visiting them now. So it is more convenient for the patients as well. Rather than you ask them to come to the hospital, you are meeting them halfway at their premises. It is a good thing and I think a lot of patients will like it. It is easier. If you ask me to go down to the hospital I might be like "ah, cannot today, tomorrow cannot." I may want to push the appointment for the later dates.* [P8, aged 58 years]

> *I think it is good because you will call me to confirm home visit appointment before you come, and if I cannot make it, I can also give you a call. This is very good...* [P6, aged 76 years]

In terms of the mHealth component of the study intervention, participants had different views on the usability and acceptability of the technology. Some participants reported that employing the smartphone app in health monitoring increased their awareness of their personal health condition and found it convenient to record and monitor their condition via the app. Importantly, having the nurse and other health care professionals accessing and monitoring the participants' conditions at the backend was especially useful in motivating them to use the app. However, older patients needed to rely on their helper to monitor their health data because of difficulties in using a smartphone:

> *...initially, it was not easy to use, there were some errors in [Chinese] translation...subsequently, this problem has been fixed, we can understand [the meaning], and it becomes very easy to use...app is preferred over paper and pen, because you just need to press a few buttons, if [you] record on paper, you need to first look for a pen, then write it down...* [P6, aged 76 years]

> *...I prefer the app...I really appreciated that. It gives you a purpose that you want to do it every morning to key in whatever information. So that it can be shared all over in the hospital. They are also looking at it. And it's so easy. I think the app is user friendly. No problem.* [P8, aged 58 years]

> *We don't know how to use smartphone, it is difficult for our age, so I have to ask my helper to key in the record.* [P3, aged 83 years]

Theme 2: Areas for Improvement

Participants also made suggestions on intervention frequency, duration, and content coverage for future improvement. A male participant believed that the interval (ie, every 2 weeks) between home visits was too short and the duration of the intervention program (ie, 6 weeks) was too long. In contrast, female participants felt that the frequency and duration of home visits could be increased. The participants also preferred that the disease education and self-care content provided by the research nurse could cover other topics related to chronic disease management and not just HF. Furthermore, one participant wanted the research nurse to relay his information to the doctors so that he did not need to repeat himself during the doctor's consultation, as consultations are usually very short:

> *Personally, I feel that the home visit interval is too close, may be one or two months will be ok...yes, both are ok, if home visit interval increased, give a telephone call in between home visits is also fine...* [P9, aged 59 years]

> *I hope that your programme duration won't be so short, it will be good if you can visit the patient two to three times a month. Or perhaps, visit more frequently at the start, and if patient is stable, can slowly [reduce the frequency]...but will continue to see them...* [P2, aged 46 years]

> *..., but now you can only teach us about heart problems...Sometimes it's inevitable that there will be other diseases, although I hope I won't be so unlucky as to have other diseases after this heart problem...But things like high blood pressure, high cholesterol, and high blood sugar are common and can be discussed...* [P10, aged 58 years]

> *I hope that whatever you saw during home visit, [you] can let my doctor know too...my problems...because doctor's consultation is very short, only have 10 minutes...but during home visit, I can tell you more...* [P3, aged 83 years]

Finally, not all participants used all the tools in the HF self-care toolkit provided by the research team for self-care. Participants felt that some of these tools did not fit their usage habits. The size of the pillbox in the self-care toolkit was too small, and the weight monitoring calendar was too complex to use. Participants preferred to record their weight in their own way, rather than following it on weekly pages, as instructed:

> *...among all the things you gave me, I did not use the table [referring to the weight monitoring calendar], and the pillbox is too small, not big enough to put all my medication in, you should have given me a bigger one...* [P2, aged 46 years]

> *the pillbox in the self-care toolkit is less useful, because if I put my medication inside, I cannot see the medication instructions and the name of the medication, also the size of the box is too small...* [P6, aged 76 years]

> *Basically, everything is ok, but...except for the weight record calendar, it is too complicated, I only need to keep my own record...* [P9, aged 59 years]

Intervention Process

Theme 3: Benefits of Visiting

Participants reported that they preferred a nurse to visit them for a variety of reasons, including better clarity and understanding with face-to-face interaction for both the nurse and participant, enabling increased depth of discussion and the family being reassured with an understanding of their conditions, and that they were well taken care of. Participants also felt that their thoughts and existing problems could be better understood during home visits. Older adult patients stated receiving too much information from the news and internet nowadays, which can confuse them sometimes. However, through face-to-face interactions, the nurse can probe and discover minor problems encountered by the patient that might otherwise go unnoticed:

> *I find it hard to communicate clearly through phone, it is better to talk face-to-face, it is much clearer in this way...[P6, aged 76 years]*

> *I preferred face-to-face interaction, because sometimes, we wanted to clarify something...I mean when you are trying to describe some problems, it is very hard to talk over the phone. Therefore, I think*

> *home visit and face-to-face interaction are much better.* [P9, aged 59 years]
>
> *I think it is very important to talk face-to-face, because you received too much conflicting information from computer...the newspaper said coffee is good for the heart yesterday but now it said [coffee] is not good [for the heart], so you don't know which information is correct...and become very confused...* [P3, aged 83 years]

In addition, some participants, especially those who were staying alone or only with their spouse who is an older adult, identified that the relational component in the intervention process made a difference to them. Perceived caring presence can be as simple as individual attention, understanding of their concerns, or comfortable interactions. During face-to-face home visits, a participant felt that the ability to voice out his problems to a nurse who is knowledgeable about HF was helpful in unloading his burden:

> *...I would love to see you coming...otherwise the house will be so quiet...* [P4, aged 58 years]
>
> *Yes, yes I do. And I really appreciate that because you took your time to come to my home, be there. It wasn't a very formal meeting. It's like talking to friends. So, it is good...we can just talk, not just about medical advice. We can sit down and talk, and have a good conversation.* [P8, aged 58 years]
>
> *When you are talking to somebody, you are putting whatever your problem or whatever elements, you are just voicing out. Just like some help to unload the burden that you've been carrying along. And talking to you, whoever the interviewer is, they know about the subject. So, it helps, you know they might understand. But talking to family member is just the basic, I don't know. They may not understand. So, it is very much helpful.* [P8, aged 58 years]

Theme 4: Personal Accountability in Self-Care

During the intervention process, participants recognized their own personal responsibility in HF self-care and felt that self-discipline was required to benefit from the study intervention:

> *actually, it's very helpful. At the same time when people try to help you, you yourself must help you. You have to be disciplined as well. That means whatever the doctor or the nurse advises you, you have to follow it. But at the same time, you have to gauge all these yourself, to see whether the conditions suit you or not. But most important thing is to discipline yourself. If you don't discipline yourself, nobody can help you.* [P11, aged 57 years]

Intervention Outcomes

Theme 5: Empowered With Knowledge and Skills in Self-Care After the Intervention

A couple of participants mentioned that they will continue to use the intervention materials, such as the measuring cup and weighing scale, to monitor HF even after the intervention ended. They expressed sustained confidence in incorporating changes after the intervention, as regular weight monitoring and control of diet have been inculcated in daily life:

> *I feel my physical health and mood have improved...I used to be less energetic, but now I feel my energy level goes up...and I feel more confident [to manage my condition] after someone taught me these methods...yes, yes, I will definitely continue to maintain healthy diet, control my water intake...will follow the instructions...* [P5, aged 69 years]
>
> *last time, I did not know soup can be salty, so I keep drinking soup and I thought that was good...but now I cut it down totally...when I go to supermarket, I will also look at the food label and pay attention to the sodium content...in the past, I only look at place of production and nothing else...* [P2, aged 46 years]
>
> *...I have been following the instructions on the HF manual...I have done that...see I used the measuring cup you gave me to make tea, with no sugar...I like to use the [measuring] cup, I have been using it and I also keep track of my weight every day, it was 65 point something in the past, now is 62.5, 63.8 and then no further increase...I lost some weights and [feel] less tired...I prefer to keep my current weight, and I am happy about it.* [P4, aged 58 years]

Theme 6: Increased Self-Efficacy in Cardiac Care

Participants reported that one of the biggest changes after the study intervention was the ease of control of symptoms due to a better understanding of the HF condition. They felt that they were gaining more control in managing their HF symptoms. A few of them showed increased initiative to control their symptoms. For example, a participant would control his activity level to reduce dyspnea. Another participant reported that he kept track of his weight to better understand his current health and adjust his diet accordingly. When necessary, they would take the initiative to act quickly:

> *I can control my condition more easily because I have more knowledge about how my condition is right now...I am quite confident...* [P11, aged 57 years]
>
> *I walk, but not very fast...and avoid strenuous exercises. I still practise tai Chi for a while, but not too long...* [P3, aged 83 years]
>
> *...I think after quitting smoking, finally I am getting on my way. And there is a way...* [P8, aged 58 years]
>
> *it is better to record down as it helps me to better understand my current health condition...and I started to pay more attention to my diet, so as not to aggravate my condition...I started to control, so I cook myself...* [P10, aged 58 years]

Discussion

Principal Findings

A total of six themes (manageability of intervention, areas for improvement, benefits of visiting, personal accountability in self-care, empowered with knowledge and skills in self-care

after the intervention, and increased self-efficacy in cardiac care) emerged from the process evaluation and summarized participants' experiences and perspectives after receiving the HOM-HEMP study intervention. Overall, participants expressed more advantages than disadvantages of the study intervention. In addition, participants appeared to be more attuned to the process of the intervention and the experience of interacting with the research nurse than to the content and form of the intervention. Thus, the strengths they expressed were more focused on the intervention process than on the intervention components. This was also reflected in the fact that although the HF self-care toolkit provided by the research team was not perfect, this did not prevent them from using other methods for self-care.

In this study, home visits remained to be the preferred method of engagement for most patients. Although home visits may be perceived as resource intensive, they have intangible values. Face-to-face interactions allowed for more authentic and in-depth exchanges between the participants and the research nurse. Participants felt that they were being listened to, rather than being rushed to end the conversation, and gave the nurse an opportunity to probe and discover problems that might otherwise have gone unnoticed. In addition, human touch and care remain to be important in the healing process. Personal encounters help foster therapeutic nurse-patient relationships that will further enhance their recovery process [15]. This was further demonstrated when some participants reported that the relational component and the perceived caring presence in the intervention process had an impact on them. This finding is consistent with the results of a previous review on the effective mechanisms of disease management interventions in HF [16]. In their review, Clark et al [16] found that studies with effective health professional supports were those with *sufficient* consultation time, who incorporated patient goals into care, providing rapid feedback, and who had a consolidated patient-professional relationship. Conversely, interventions are less effective when health care providers focus too much on mere information giving or prioritizing treatment goals over the patient's own goals [16].

In this study, participants reported that they became more confident in self-care as a result of increased knowledge and understanding of symptom monitoring and interpretation. Effective intervention promoted the understanding of many complex links between HF symptoms and self-care tasks [17]. Although a plethora of accessible health information exists on the internet, in newspapers, and on instant messaging tools such as WhatsApp, a considerable proportion of patients are left with either an incomplete or an incorrect understanding of their condition, as they are not equipped to discern this vast amount of information. Even when they have the right information, guidance in translating their knowledge and skills into pragmatic actions is needed. As such, the consultative role of the nurse has evolved from simply providing health information to helping patients sift through and ascertain the credibility and authority of the informational sources, identify knowledge gaps, dispel confusion, and provide personalized advice. Participants seemed to particularly value this kind of support from health care professionals. This can be seen in the participants who expressed the wish that the research nurse could have also included education topics on other chronic disease management besides HF. Individualized interventions that are responsive to individual learning needs and preferences, with personalized reinforcement of specific salient information, have been shown to be effective in promoting self-care behavior change [17-19].

Although mHealth and telemedicine have great potential to overcome many barriers to health care delivery [19], especially in the context of the current COVID-19 pandemic [20], participants had different views on the use of technology in the study. Some believed that employing a smartphone app in health monitoring was not only convenient but also increased their awareness of personal health. One of the main motivations for participants to use the app was that they knew that their recorded data on the app would be monitored backend by the research nurse. Athilingam et al [21] reported that mHealth apps have the potential to improve self-care by providing patients with real-time access to health care providers. Similar observations were found in an unpublished local study in which postmyocardial infarction patients highlighted that the greatest advantage of telemedicine interventions was the ability to have direct access to their health care providers [22]. Although participants commended the many benefits of using a smartphone app in their health monitoring, older patients needed to rely on their caregivers to monitor their health data because of the difficulties in using a smartphone. Past research has reported that the efficacy of mobile phone–based telemonitoring in improving HF outcomes may depend on the characteristics of the patient population [23,24]. In this regard, further trials are needed to explore the effectiveness of mHealth apps in defined patient subgroups.

In this process evaluation, participants provided a lot of valuable feedback to maximize the study intervention in future studies, particularly on the structure of the intervention. First, interventions should be kept simple and manageable so as not to overwhelm or intimidate the user. Second, there should be some flexibility in the implementation of study interventions. Depending on the patient's progress, the interval between home visits can be extended or shortened. Long intervals between home visits can be augmented by telephone calls to check on the patient. In addition, one interviewee suggested that the intervention duration be extended as needed. Finally, although the HF self-care toolkit was designed to help patients establish a systematic self-care process and provide some tangible help in integrating self-care into their daily lives, not all of the items in the toolkit were user-friendly. Consequently, some participants did not like to follow up on tracking their weight trends and managing their medication in the way they were told.

Limitations

It is inevitable that this study had limitations that may affect the transferability of the research findings. First, as the research nurse for this study delivered both the study intervention and conducted the interviews for the process evaluation, this may have led to bias in the collection of participant feedback because of an established rapport between the participants and the research nurse. However, the development of rapport is not only a limitation but can also be seen as a strength. When participants

are familiar with the interviewer, they will be more open to speaking and sharing their genuine opinions. Moreover, as the interviews were semistructured, in addition to discussing the parts of the intervention that were found to be helpful, we made a point of exploring with participants where the research interventions were least helpful and where future improvements were needed. These measures facilitated a relatively balanced discussion with participants. Second, some participants were very reticent during the interview process; even after much probing, they just answered *everything is fine* or repeated what they had already mentioned. As a result, some of the interviews were very short, which may have affected the authenticity of the data.

Conclusions

Despite these limitations, the findings of the process evaluation provided additional information about participants' perceptions and experiences with the HOM-HEMP intervention. Although a home visit may be perceived as resource intensive, it remains to be the preferred way of engagement for most patients. During the intervention, the nurse played an active and essential role in the individual's healing process, further demonstrating that not only does the value of nursing come from the content of the intervention but that interactions with the patient can be an important empowering process for self-care behavior changes. These findings shed light on why and how the intervention worked and areas for improvement in future studies. It allows researchers and practitioners to better understand the mechanism by which an intervention generates impact. Therefore, future intervention programs can leverage these mechanisms and adapt them to different contexts for better intervention effectiveness.

Acknowledgments

This study was funded by a grant from the Health Services Research Grant of the National Medical Research Council in Singapore (grant NMRC/HSRG/0064/2016). The study funder was not involved in the study design, data collection, data analysis, or writing for publication.

Authors' Contributions

All authors have agreed on the final version and meet at least one of the following criteria: substantial contributions to conception and design, acquisition of data, or analysis and interpretation of data and drafting the paper or revising it critically for important intellectual content.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The heart failure education and self-care toolkit.
[PNG File , 388 KB - jmir_v23i4e28216_app1.png]

Multimedia Appendix 2
Heart failure supplementary smartphone app.
[PNG File , 512 KB - jmir_v23i4e28216_app2.png]

References

1. Santhanakrishnan R, Ng TP, Cameron VA, Gamble GD, Ling LH, Sim D, et al. The Singapore Heart Failure Outcomes and Phenotypes (SHOP) study and Prospective Evaluation of Outcome in Patients with Heart Failure with Preserved Left Ventricular Ejection Fraction (PEOPLE) study: rationale and design. J Card Fail 2013 Mar;19(3):156-162. [doi: 10.1016/j.cardfail.2013.01.007] [Medline: 23482075]
2. Lam CS. Heart failure in Southeast Asia: facts and numbers. ESC Heart Fail 2015 Jun;2(2):46-49 [FREE Full text] [doi: 10.1002/ehf2.12036] [Medline: 28834655]
3. Dickstein K, Cohen-Solal A, Filippatos G, McMurray JJ, Ponikowski P, Poole-Wilson PA, ESC Committee for Practice Guidelines (CPG). ESC guidelines for the diagnosis and treatment of acute and chronic heart failure 2008: the Task Force for the Diagnosis and Treatment of Acute and Chronic Heart Failure 2008 of the European Society of Cardiology. Developed in collaboration with the Heart Failure Association of the ESC (HFA) and endorsed by the European Society of Intensive Care Medicine (ESICM). Eur Heart J 2008 Oct;29(19):2388-2442. [doi: 10.1093/eurheartj/ehn309] [Medline: 18799522]
4. Primary care model. Agency for Integrated Care. 2019. URL: https://www.primarycarepages.sg/practice-management/primary-care-model/primary-care-network-(pcn) [accessed 2021-04-09]
5. Lainscak M, Blue L, Clark AL, Dahlström U, Dickstein K, Ekman I, et al. Self-care management of heart failure: practical recommendations from the Patient Care Committee of the Heart Failure Association of the European Society of Cardiology. Eur J Heart Fail 2011 Feb;13(2):115-126 [FREE Full text] [doi: 10.1093/eurjhf/hfq219] [Medline: 21148593]

6. Ma H, Lum CM, Woo J, Fhkam M, Edin F. Readmission of patients with congestive heart failure: the need for focused care. Asian J Gerontol Geriatr. 2006. URL: https://tinyurl.com/2tkf894z [accessed 2021-04-09]
7. Currie K, Strachan PH, Spaling M, Harkness K, Barber D, Clark AM. The importance of interactions between patients and healthcare professionals for heart failure self-care: a systematic review of qualitative research into patient perspectives. Eur J Cardiovasc Nurs 2015 Dec;14(6):525-535. [doi: 10.1177/1474515114547648] [Medline: 25139468]
8. Dickson VV, Lee C, Riegel B. How do cognitive function and knowledge affect heart failure self-care? J Mix Methods Res 2011 Mar 23;5(2):167-189. [doi: 10.1177/1558689811402355]
9. Jiang Y, Shorey S, Nguyen HD, Wu VX, Lee CY, Yang LF, et al. The development and pilot study of a nurse-led HOMe-based HEart failure self-Management Programme (the HOM-HEMP) for patients with chronic heart failure, following Medical Research Council guidelines. Eur J Cardiovasc Nurs 2020 Mar 05;19(3):212-222. [doi: 10.1177/1474515119872853] [Medline: 31486332]
10. Moore GF, Audrey S, Barker M, Bond L, Bonell C, Hardeman W, et al. Process evaluation of complex interventions: Medical Research Council guidance. Br Med J 2015 Mar 19;350 [FREE Full text] [doi: 10.1136/bmj.h1258] [Medline: 25791983]
11. Donabedian A, Attwood JC. An evaluation of administrative controls in medical-care programs. N Engl J Med 1963 Aug 15;269(7):347-354. [doi: 10.1056/nejm196308152690706]
12. Ayanian JZ, Markel H. Donabedian's lasting framework for health care quality. N Engl J Med 2016 Jul 21;375(3):205-207. [doi: 10.1056/NEJMp1605101] [Medline: 27468057]
13. Polit DF, Beck CT. Essentials of Nursing Research : Appraising Evidence for Nursing Practice. Philadelphia: Wolters Kluwer; 2017.
14. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
15. Pullen Jr RL, Mathias T. Fostering therapeutic nurse-patient relationships. In: Nursing Made Incredibly Easy!. Netherlands: Wolters Kluwer; 2010:4.
16. Clark AM, Wiens KS, Banner D, Kryworuchko J, Thirsk L, McLean L, et al. A systematic review of the main mechanisms of heart failure disease management interventions. Heart 2016 May 23;102(9):707-711. [doi: 10.1136/heartjnl-2015-308551] [Medline: 26908100]
17. Riegel B, Dickson VV, Garcia LE, Creber RM, Streur M. Mechanisms of change in self-care in adults with heart failure receiving a tailored, motivational interviewing intervention. Patient Educ Couns 2017 Feb;100(2):283-288 [FREE Full text] [doi: 10.1016/j.pec.2016.08.030] [Medline: 27599712]
18. Riegel B, Dunbar SB, Fitzsimons D, Freedland KE, Lee CS, Middleton S, et al. Self-care research: where are we now? Where are we going? Int J Nurs Stud 2019 Aug 23 [FREE Full text] [doi: 10.1016/j.ijnurstu.2019.103402] [Medline: 31630807]
19. Indraratna P, Tardo D, Yu J, Delbaere K, Brodie M, Lovell N, et al. Mobile phone technologies in the management of ischemic heart disease, heart failure, and hypertension: systematic review and meta-analysis. JMIR Mhealth Uhealth 2020 Jul 06;8(7) [FREE Full text] [doi: 10.2196/16695] [Medline: 32628615]
20. Hutchings OR, Dearing C, Jagers D, Shaw MJ, Raffan F, Jones A, et al. Virtual health care for community management of patients with COVID-19 in Australia: observational cohort study. J Med Internet Res 2021 Mar 09;23(3) [FREE Full text] [doi: 10.2196/21064] [Medline: 33687341]
21. Athilingam P, Osorio RE, Kaplan H, Oliver D, O'neachtain T, Rogal PJ. Embedding patient education in mobile platform for patients with heart failure: theory-based development and beta testing. Comput Inform Nurs 2016 Feb;34(2):92-98. [doi: 10.1097/CIN.0000000000000216] [Medline: 26765655]
22. Koh KW, Wang W, Richards AM, Chan MY, Cheng KK. Effectiveness of advanced practice nurse-led telehealth on readmissions and health-related outcomes among patients with post-acute myocardial infarction: ALTRA Study Protocol. J Adv Nurs 2016 Jun;72(6):1357-1367. [doi: 10.1111/jan.12933] [Medline: 26915719]
23. Seto E, Leonard KJ, Cafazzo JA, Barnsley J, Masino C, Ross HJ. Mobile phone-based telemonitoring for heart failure management: a randomized controlled trial. J Med Internet Res 2012 Feb 16;14(1):e31 [FREE Full text] [doi: 10.2196/jmir.1909] [Medline: 22356799]
24. Ware P, Dorai M, Ross HJ, Cafazzo JA, Laporte A, Boodoo C, et al. Patient adherence to a mobile phone-based heart failure telemonitoring program: a longitudinal mixed-methods study. JMIR Mhealth Uhealth 2019 Feb 26;7(2) [FREE Full text] [doi: 10.2196/13259] [Medline: 30806625]

Abbreviations

HF: heart failure
HOM-HEMP: home-based heart failure self-management program
mHealth: mobile health

Edited by G Eysenbach; submitted 25.02.21; peer-reviewed by W Jiang, K Liu, MI Cajita; comments to author 18.03.21; revised version received 23.03.21; accepted 03.04.21; published 27.04.21.

Please cite as:
Jiang Y, Koh KWL, Ramachandran HJ, Tay YK, Wu VX, Shorey S, Wang W
Patients' Experiences of a Nurse-Led, Home-Based Heart Failure Self-management Program: Findings From a Qualitative Process Evaluation
J Med Internet Res 2021;23(4):e28216
URL: https://www.jmir.org/2021/4/e28216
doi:10.2196/28216
PMID:33904823

Original Paper

Gender Disparity in the Authorship of Biomedical Research Publications During the COVID-19 Pandemic: Retrospective Observational Study

Goran Muric[1], PhD; Kristina Lerman[1,2], PhD; Emilio Ferrara[1,2,3], PhD

[1]Information Sciences Institute, University of Southern California, Los Angeles, CA, United States

[2]Department of Computer Science, University of Southern California, Los Angeles, CA, United States

[3]Annenberg School for Communication and Journalism, University of Southern California, Los Angeles, CA, United States

Corresponding Author:
Emilio Ferrara, PhD
Annenberg School for Communication and Journalism
University of Southern California
3630 Watt Way
Suite 402
Los Angeles, CA, 90089
United States
Phone: 1 3104488661
Email: emiliofe@usc.edu

Abstract

Background: Gender imbalances in academia have been evident historically and persist today. For the past 60 years, we have witnessed the increase of participation of women in biomedical disciplines, showing that the gender gap is shrinking. However, preliminary evidence suggests that women, including female researchers, are disproportionately affected by the COVID-19 pandemic in terms of unequal distribution of childcare, elderly care, and other kinds of domestic and emotional labor. Sudden lockdowns and abrupt shifts in daily routines have had disproportionate consequences on their productivity, which is reflected by a sudden drop in research output in biomedical research, consequently affecting the number of female authors of scientific publications.

Objective: The objective of this study is to test the hypothesis that the COVID-19 pandemic has had a disproportionate adverse effect on the productivity of female researchers in the biomedical field in terms of authorship of scientific publications.

Methods: This is a retrospective observational bibliometric study. We investigated the proportion of male and female researchers who published scientific papers during the COVID-19 pandemic, using bibliometric data from biomedical preprint servers and selected Springer-Nature journals. We used the ordinary least squares regression model to estimate the expected proportions over time by correcting for temporal trends. We also used a set of statistical methods, such as the Kolmogorov-Smirnov test and regression discontinuity design, to test the validity of the results.

Results: A total of 78,950 papers from the bioRxiv and medRxiv repositories and from 62 selected Springer-Nature journals by 346,354 unique authors were analyzed. The acquired data set consisted of papers that were published between January 1, 2019, and August 2, 2020. The proportion of female first authors publishing in the biomedical field during the pandemic dropped by 9.1%, on average, across disciplines (expected arithmetic mean y_{est}=0.39; observed arithmetic mean y=0.35; standard error of the estimate, S_{est}=0.007; standard error of the observation, σ_x=0.004). The impact was particularly pronounced for papers related to COVID-19 research, where the proportion of female scientists in the first author position dropped by 28% (y_{est}=0.39; y=0.28; S_{est}=0.007; σ_x=0.007). When looking at the last authors, the proportion of women dropped by 7.9%, on average (y_{est}=0.25; y=0.23; S_{est}=0.005; σ_x=0.003), while the proportion of women writing about COVID-19 as the last author decreased by 18.8% (y_{est}=0.25; y=0.21; S_{est}=0.005; σ_x=0.007). Further, by geocoding authors' affiliations, we showed that the gender disparities became even more apparent when disaggregated by country, up to 35% in some cases.

Conclusions: Our findings document a decrease in the number of publications by female authors in the biomedical field during the global pandemic. This effect was particularly pronounced for papers related to COVID-19, indicating that women are producing fewer publications related to COVID-19 research. This sudden increase in the gender gap was persistent across the 10 countries

with the highest number of researchers. These results should be used to inform the scientific community of this worrying trend in COVID-19 research and the disproportionate effect that the pandemic has had on female academics.

(J Med Internet Res 2021;23(4):e25379) doi:10.2196/25379

KEYWORDS

science of science; gender disparities; research evaluation; COVID-19

Introduction

As of the date of this writing, the COVID-19 pandemic has claimed hundreds of thousands of lives worldwide and disrupted almost all aspects of human society. The socioeconomic impacts of the pandemic are yet to be assessed and the impending economic crisis and recession are becoming evident [1-3]. During recessions, men are more likely to lose their jobs, as men work in industries that are heavily affected by the slowdown in economic activity, such as manufacturing and construction. Compared to previous economic crises, the current crisis has disproportionately affected female workers [4-11]. One of the reasons for such disparity is women's overrepresentation in occupations in industries that are most affected by the closures and movement restrictions imposed by public health policies, such as restaurants and hospitality. Another large part of the gender disparity is related to the unequal division of labor in the household, as women are traditionally expected to continue to devote more time to childcare and domestic chores than their partners [4]. In the case of dual-earner, heterosexual married couples with children, the partners have unequally adjusted their work time during the pandemic. Mothers with young children have reduced their work hours 4 to 5 times more than fathers, which contributes to the increased gender gap in earnings [5]. Working mothers affected by the unequal distribution of working hours and the additional burden of domestic chores have reported lower work productivity and job satisfaction than men [6].

Stay-at-home orders, lockdowns, and school closures have affected scientists as well, especially those caring for children or other family members [12,13]. Female scientists reported that their ability to devote time to their research has been substantially affected, and the impact is most pronounced for female scientists with young dependents [14]. The sudden shift in daily activities makes it hard to balance between increasing professional requirements and childcare.

As a result, the research productivity of female scientists appears to have decreased [15-18]. Early evidence suggests that the proportion of publications with female authors is lower during the pandemic with the evident gendered authorship disparities in journal submissions [19,20]. Reports from journal editors in the fields of international studies, political science, economics, medicine, and philosophy indicate that the proportion of submissions authored by women has dropped in most cases [21]. Even though female academics are still submitting manuscripts for publication during the crisis, they are submitting less of their own work than men [22].

A similar effect has been observed with publications on preprint servers. The proportion of female authors publishing on the most popular economics preprint servers is lower than expected [23,24], with only 14.6% of female authors; comparably, they usually make up about 20% of the authors in these databases. Similarly, women publish less in other disciplines, such as physics, earth science, and sociology [25]. In regard to medical and related sciences, on top of the exacerbated gender disparity in publishing during the pandemic, the proportion of female scientists publishing research specifically about COVID-19 is much lower than expected, by almost 23% [25-27].

Motivated by ongoing research efforts, we expand on the previous research by analyzing a large bibliographic data set in the biomedical field; we also employ different modeling techniques that can further improve our understanding of this phenomenon. The aim of this study is to quantify how the COVID-19 crisis exacerbates the gender gap in scientific publishing in the biomedical field.

Methods

Data

Bibliometric data on published papers were collected from three separate sources:

1. The bioRxiv repository contains 51,171 papers and 225,110 authors; Rxivist is the application programming interface (API) provider for bioRxiv publications [28].
2. The medRxiv repository contains 8845 papers and 52,364 authors; data are scraped directly from medrxiv.org.
3. The Springer-Nature repository contains 19,525 papers and 91,257 authors; data from 62 journals are collected using the Springer-Nature OpenAccess API. Springer-Nature data include high-impact journals, such as Nature Genetics, Nature Medicine, and Nature Immunology, as well as multiple BMC journals, such as BMC Bioinformatics and BMC Genomics.

We included the data from all journals in the biomedical field for which Springer-Nature provides data. A complete list of journals used in the analysis is available in Table S1 of Multimedia Appendix 1. All the papers from the data set were published between January 1, 2019, and August 2, 2020. The earliest publication on medRxiv is from June 25, 2019.

For each source, we collected the relevant metadata. For each paper in bioRxiv and medRxiv, we kept the *date* of publishing. For Springer-Nature journals, we kept the date of manuscript submission, which is the most comparable date to publication dates in bioRxiv and medRxiv. We additionally stored the *title* and the *abstract* of each paper as well as the scientific discipline of each paper. There were a total of 112 scientific disciplines represented in the data, and each paper belonged to a single discipline. The complete list of all disciplines is provided in

Table S2 in Multimedia Appendix 1. For each author, we preserved the *name*, *affiliation*, and the *authorship order*. We removed the papers with group authors, such as scientific consortiums and projects (~0.1% of all papers), as they do not represent individuals. We used socioeconomic data on countries, including their respective gross domestic product (GDP) per capita provided by Our World in Data [29].

Identifying Authors' Genders

To infer each author's gender from their name, we used a state-of-the-art tool, namely the genderize.io API [30]. Given an input name, the model returns a gender and a confidence score between 0.5 and 1. The uncertainty is greater for Asian names, which often are not gender specific [31]. We filtered out all authors for which the confidence scores were lower than 0.8. Overall, 19% of names yielded a score below this threshold, with Chinese and Korean names topping the ranking at 54% and 41%, respectively.

In our data set, we identified the most likely gender of 466,836 authors in total. Out of these, the gender of 348,506 (74.7%) unique authors (214,095 male [61.4%] and 134,411 female [38.6%]) could be inferred with high accuracy, with confidence scores from genderize.io higher than 0.8.

Identifying Authors' Countries

To identify each author's country in the bioRxiv and medRxiv data sets, we first located a toponym in each author's affiliation and assigned to it the most likely country code. If there was no toponym, we queried the Global Research Identifier Database, found the institution with the most similar name, and assigned the institution's country to that author. Additionally, we manually checked the location of the most common affiliation names from the data set that covered most of the authors. The countries of approximately 80% of all authors were determined using this method. The countries of the authors in the Springer-Nature data set were already provided by the API.

Identifying COVID-19 Papers

The papers that dealt specifically with COVID-19 and similar topics were identified by the set of keywords that appeared in their titles or abstracts.

Calculating the Differences Between the Expected and Observed Proportions

To measure the discrepancy between the expected and observed proportions of female authors, we first established baselines, which were the expected proportions of female researchers that appeared as authors of publications. The expected proportions were calculated using the ordinary least squares (OLS) model and historical data from January 2019 to March 2020 (see the Model section). We then calculated the true observed proportions of female authors who published during the COVID-19 pandemic in 2020 and compared it to the expected baselines. The error for the predicted value was the mean standard error of the prediction. The error of the observed value was calculated as the standard error of the mean: $SE = \sigma/\sqrt{n}$. The percentage change is calculated as $diff = (f^{exp} - f^{obs})/f^{exp}$, where f^{exp} is the expected proportion and f^{obs} is the observed proportion. The errors for the percentage change were calculated as the total sum of the errors of predicted and observed values.

Model

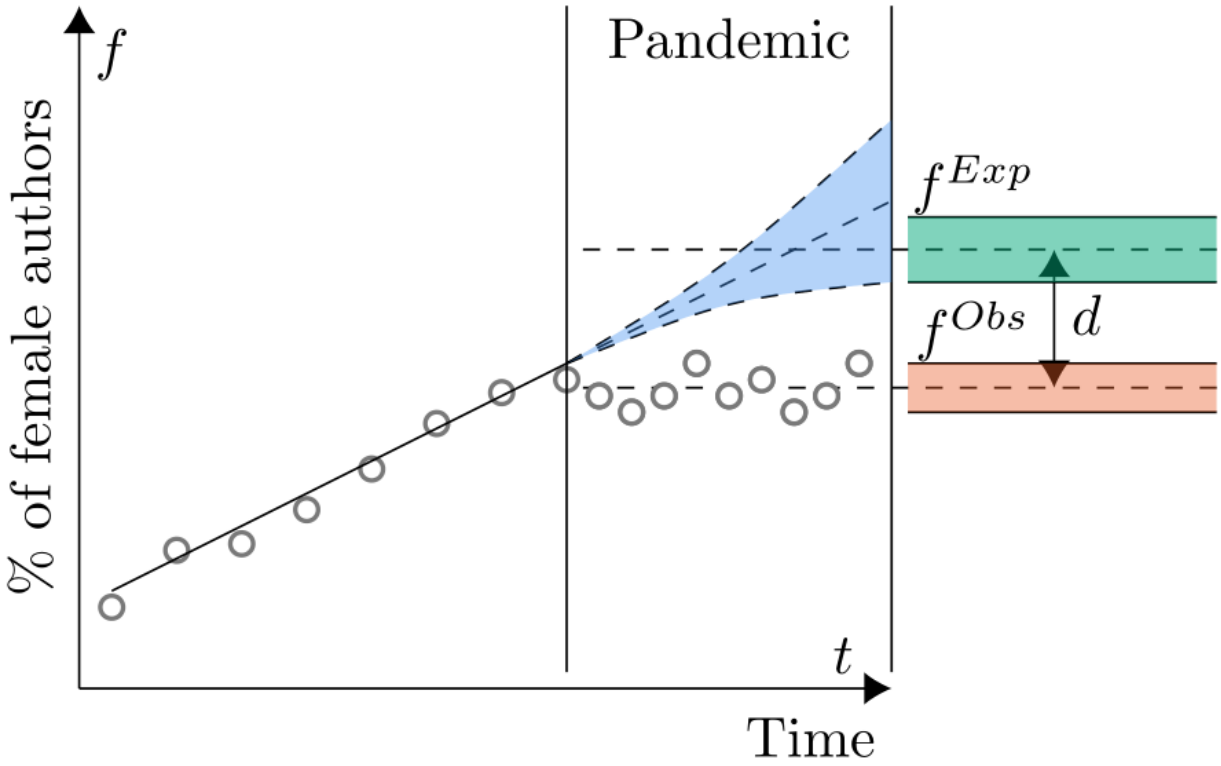
Using historical data from before March 15, 2020, we calculated the proportion of female authors who published each week. We fit an OLS regression model, $f = \beta t + c$, where f is the proportion of female authors, which serves as a response variable; t is the predictor variable—time of publication/submission (to the nearest week); and β and c are the slope and the intercept, respectively. We fit the separate models depending on the level of disaggregation (country, publisher, etc). The model is illustrated in Figure 1. From the model, we derived the expected fraction, $f^{exp} = \sum f/n$, which is the mean fraction of all predicted values for the observed period and $f^{obs} = \sum f_{true}/n$. To estimate the expected number of papers and authors, we used a similar approach, where the response variables were the numbers of papers and authors rather than the proportion of female authors. We used the *statsmodels* [32] package in Python 3.6 (Python Software Foundation) for this purpose.

The OLS model tends to weight all data points equally, regardless of the number of samples. To guarantee the validity of the statistical analysis, we established the conditions under which the data points would be evaluated. The number of data points used to fit the OLS model before March 2020 and the number of data points after March 2020 were at least 10 each. This way, we limited the impact of small-sample observations that could skew the estimate.

We additionally evaluated the model by applying the generalized linear model with binomial errors and a logit link function, as the OLS model could overestimate the proportions in binary variables. Both models performed similarly, and the OLS model did not provide any out-of-norm estimates. For the sake of better interpretability and consistency with modeling the nominal number of authors and papers, we decided to use the OLS model.

To better capture the productivity of the population, we counted each publication from each author separately, effectively modeling the proportion of papers authored by the population of female authors. Considering that multiple authorships in the observed period were relatively rare (<5% of all first authors and <10% of all last authors had >1 paper), we considered each authorship independently.

Figure 1. Statistical model. Schematic illustration of the ordinary least squares model used to calculate the expected numbers and proportions. *d*: difference; f^{exp}: expected proportion; f^{obs}: observed proportion.



Regression Discontinuity Design

To estimate the potential causal effects of the pandemic on the proportion of female researchers, we devised a typical nonparametric regression discontinuity design (RDD) with a local linear regression in time, with the following general form:

$$Y = \alpha + \tau D + \beta_1(X - c) + \beta_2 D(X - c) + \varepsilon$$

where c is the treatment cutoff and D is a binary variable equal to 1 if $X \geq c$. In our case, we assumed that the date c of a policy change was mid-March 2020. For all dates $t > c$, the unit was treated, and for all dates $t < c$, the unit was not. This regression discontinuity setup used time-series data and, in this case, weekly observations, both globally and on a country level. By comparing observations lying closely on either side of the temporal threshold, we estimated the average treatment effect. We made sure to focus on observations not too far in time from the threshold, avoiding potential bias from unobservable confounders [33].

The falsification, or placebo, tests were performed by using fake cutoffs before and after mid-March 2020 and comparing the treatment effect. We identified the optimal cutoff point c_0 as the point in time when the treatment effect was the most prominent, $c_0 = c|\max(|\tau|)$. The RDD was implemented using the *rdd* package in Python [34].

Data Availability and Reproducibility

The data and source code for reproducing the results are available at GitHub [35].

Results

The Gender Gap in Research During the COVID-19 Pandemic

Overall, during the pandemic, scientists posted papers on preprint servers at an increasing rate. On average, we observed 31.2% more papers than expected and a 41.6% increase in the number of authors (39.2% increase for females and 42.9% increase for males). Despite the absolute increase in the numbers of papers and authors across publishers (see Figure S1 and Tables S3-S5 in Multimedia Appendix 1), the proportion of female authors was lower than expected.

In biology, medicine, and related disciplines, the most active contributors are usually listed first. The author listed last is the most senior author, typically the head of the lab. To address the high variability of the number of authors on the publications (μ=7.4, σ=9.2), we analyzed the proportion of women, separately, who appeared as the first author, the last author, an author regardless of authorship order, and the solo author. Additionally, we performed a separate analysis on the papers with topics that were directly related to COVID-19 (see Table 1).

Table 1. The expected and observed proportions of female authors disaggregated by the order of authorship and the topic.

Author order and paper topic	Expected proportion		Observed proportion		Drop, %
	y_{est}[a]	S_{est}[b]	y[c]	σ_x[d]	
First					
All	0.389	0.007	0.353	0.004	9.142
COVID-19	0.389	0.007	0.280	0.007	28.031
Non-COVID-19	0.389	0.007	0.380	0.004	2.372
Last					
All	0.257	0.005	0.236	0.003	7.961
COVID-19	0.257	0.005	0.209	0.007	18.812
Non-COVID-19	0.257	0.005	0.246	0.003	4.416
Any					
All	0.354	0.003	0.348	0.002	1.578
COVID-19	0.354	0.003	0.341	0.009	3.530
Non-COVID-19	0.354	0.003	0.351	0.002	0.934
Solo					
All	0.210	0.030	0.137	0.008	34.586
COVID-19	0.210	0.030	0.137	0.023	34.514
Non-COVID-19	0.210	0.030	0.168	0.013	19.802

[a]y_{est}: the arithmetic mean of the estimate.

[b]S_{est}: the mean standard error of the estimate.

[c]y: the arithmetic mean of the observation.

[d]σ_x: the standard error of the mean of the observation.

The aggregate results suggest that the proportion of female authors publishing on all topics as the first author decreased by 9.1% (expected arithmetic mean y_{est}=0.38; observed arithmetic mean y=0.35; standard error of the estimate, S_{est}=0.007; standard error of the observation, σ_x=0.004). The percentage drop became unusually prominent when we analyzed the papers about COVID-19. The proportion of female scientists who wrote on COVID-19-related topics as the first author was lower by 28% (y_{est}=0.38; y=0.27; S_{est}=0.007; σ_x=0.007). When considering the last authors, the proportion of women writing about COVID-19 decreased by 18.8% (y_{est}=0.25; y=0.2; S_{est}=0.005; σ_x=0.007). However, when we focused only on papers that did not deal with COVID-19, we saw a smaller change both for the first author (2.3%) and the last author (4.4%). The proportion of women publishing papers on topics other than COVID-19 on medRxiv increased by 14%, on average. The results are shown in Table 1 and are illustrated in Figure 2. The expected proportions are plotted as the green bars, and the true proportions are plotted in orange. The standard errors were relatively small (see Methods section for details about the error calculation).

Additionally, we focused our analysis on the papers with a single author and discovered an even greater disparity. We observed 34.5% (y_{est}=0.21; y=0.13; S_{est}=0.03; σ_x=0.008) fewer female solo authors during the pandemic who published on all topics across the platforms (see Figure S2 in Multimedia Appendix). A similar disparity appeared in case of the solo authors publishing papers about COVID-19. Note that only 3.2% (2551/79,528) of all papers were authored by a single author, hence, the relatively large standard errors of both the estimate and the observation, especially for papers published on medRxiv. The effect still exists, although much less prominently, when we observed all authors regardless of the order of authorship. More detailed information is provided in Table S3 in Multimedia Appendix 1.

The results suggest that the aggregate gender disparity in academia during the pandemic was due to the increased publication rate of papers about COVID-19 authored by men. To further explore this possibility, we tracked the individual publication records and calculated the probability that the author would publish work about COVID-19. Around 3.7% of men who had publication records in our data set would publish at least one paper about COVID-19, compared to ~2.2% of women. Men who already had a publication before the pandemic were 37% more likely to publish a paper about COVID-19. This suggest that women are getting excluded from critical research about COVID-19.

Figure 2. Comparison of the expected and observed proportions of female authors that published during the COVID-19 pandemic. Green bars represent the expected proportion of female authors, estimated by the ordinary least squares model from the historical data from 2019. Orange bars represent the observed proportion of female authors that published during the COVID-19 pandemic. The standard errors of the aggregate analyses are represented as the vertical lines on top of the bars. The papers are divided by topic into three groups: (1) all papers from the data set, (2) papers that deal directly with COVID-19 and related topics, and (3) papers that are not about COVID-19 or related topics. The first row shows the results from all publishers combined. The following rows represent the results for each publisher separately.

When disaggregated by publisher, the relative drop in the proportion of female first authors for COVID-19-related research was 12.6%, 23.2%, and 2.1% for bioRxiv, medRxiv, and Springer-Nature journals, respectively (see Figure 2). A similar disparity was observed for last authors, with a relative drop of 20.1%, 30.8%, and 23.6%, and for authors regardless of the authorship order, with a relative drop of 2.2%, 10.7%, and 16.1% for bioRxiv, medRxiv, and Springer-Nature journals, respectively. In the case of solo papers, the average drop across the platforms was 34.5%. The proportion of females publishing on topics other than COVID-19 remained within the standard error of the estimate, without strong evidence of decrease. Note the large standard errors in the estimated proportion of women publishing COVID-19-related papers in Springer-Nature journals due to the lack of data. Only published papers have metadata available through the Springer-Nature API, and many papers submitted during the pandemic that will ultimately be published have not yet been accepted and published (see Methods section).

Additionally, we checked whether there was a significant change in the proportion of women authors that occurred in mid-March 2020. To test the hypothesis, we performed an RDD analysis in time (see Methods section). We estimated a vertical discontinuity of the proportion of women over time by the coefficient τ at the cutoff point c_0 = March 15, 2020. For all the papers, regardless of the topic, we obtained τ=–0.008 with P=.03. However, when considering only the papers about COVID-19, the discontinuity became clearer with τ=–0.049 and P<.001. To assert the robustness of our model, we performed a placebo test (see Methods section) to confirm that the discontinuity was likely aligned with the start of the pandemic, and that it happened at or around March 2020 and

not during any time in 2019. When RDD analysis was performed at the country level, we confirmed that, for most countries, the cutoff threshold fell between mid-March and mid-April 2020 (see Table S6 in Multimedia Appendix 1). The RDD analysis suggests that there was a drop in the proportion of female authors at the beginning of April 2020 that was more significant than any other fluctuation that occurred in 2019 or after April 2020.

Further, we checked whether we could confidently use the proportion of women who published before the pandemic as the reference to estimate the proportion of women who published papers specifically about COVID-19. A hypothesis is that before the pandemic, women were less likely to be represented in the scientific disciplines that would produce COVID-19 research. To check this hypothesis, we first performed a chi-square test on the distribution of disciplines involved in COVID-19 research. We discovered that some disciplines, such as infectious diseases, epidemiology, public health, and global health, were overrepresented (P<.001). Then, we tested whether the proportions of women in COVID-19 disciplines were significantly different from non-COVID-19 disciplines. By performing the Kolmogorov-Smirnov test, we compared the distributions of the proportion of women across two groups of disciplines and we obtained P=.84. We conclude that the two groups were sampled from populations with the same distributions, and we can be confident that we can use the data on the proportion of women from before the pandemic to model the proportion of women that publish about COVID-19.

Some Trends During the Pandemic

To assess the temporal trends during the pandemic, we built the linear model $f(t) = \alpha + \beta t + \varepsilon$, where $f(t)$ is the proportion of female scientists, and t is the time in weeks after mid-March 2020. The regression coefficient β is used to quantify the trend. We did not identify a significant change in the proportion of female first authors (see Table 2). However, we observed a small but significant increase in the proportion of female scientists appearing as the last author (β=.001, P=.002) and as an author regardless of authorship order (β=.001, P<.001). An even stronger positive trend was observed for the COVID-19-related research for the last author (β=.003, P<.001) and for an author regardless of authorship order (β=.005, P<.001) (see Table S7 in Multimedia Appendix 1).

Table 2. Parameters of the linear model of the proportion of female authors over time during the pandemic.

Paper topic	First author		Last author		All authors		Solo author	
	β[a]	P value	β	P value	β	P value	β	P value
All	–.002	.08	.002	.002	.001	<.001	.000	>.99
COVID-19	.001	.28	.003	<.001	.005	<.001	–.005	.12
Non-COVID-19	–.002	.02	.001	.03	.000	.17	.004	.13

[a]β is a regression coefficient.

Country-Level Analysis

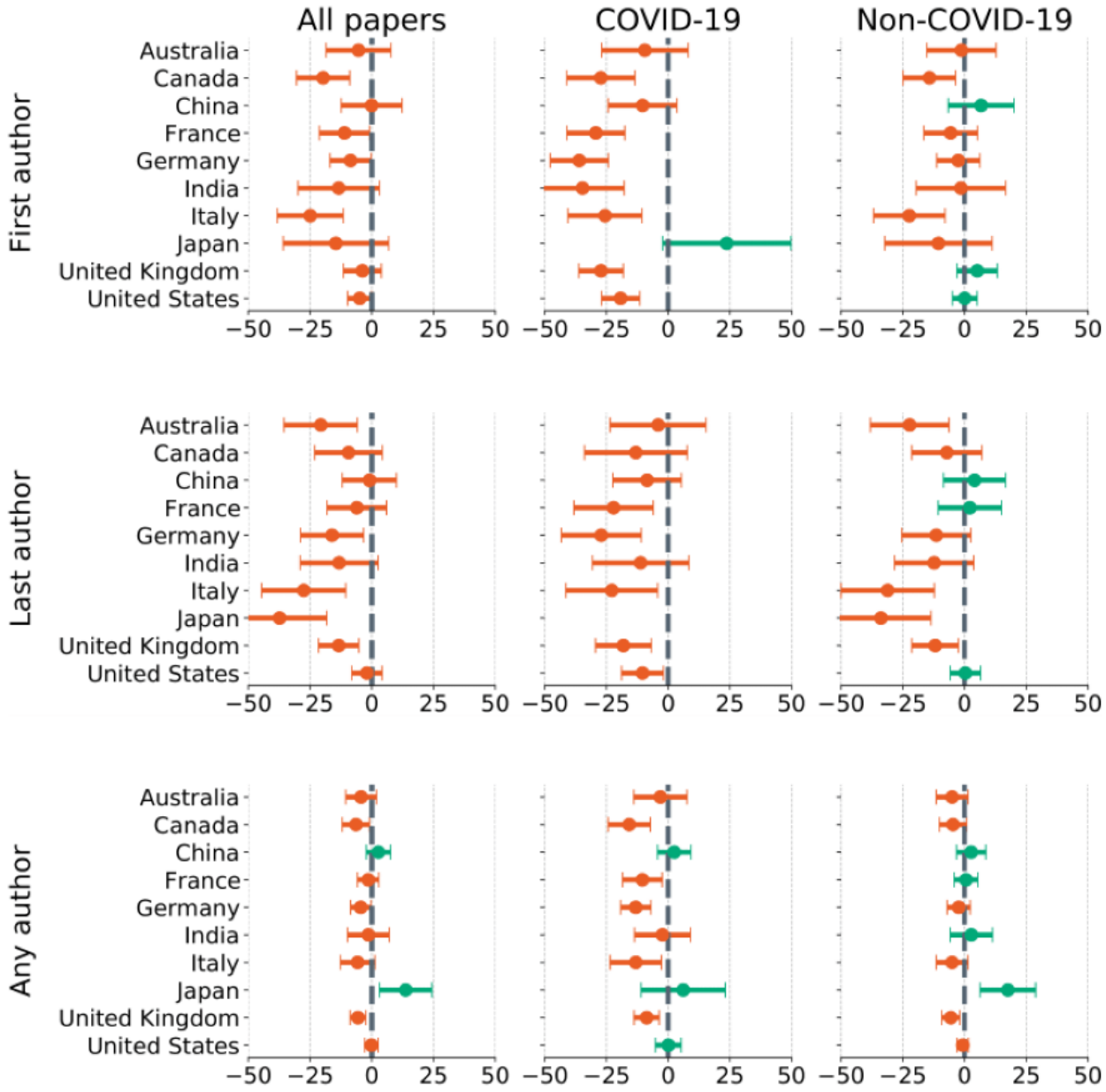
We identified the most likely country of the authors based on their affiliations (see Methods section) and measured the difference between the expected and observed proportions of female authors during the pandemic. Figure 3 shows the pandemic-related gender gap across the countries with the largest share of authors. The values represent percentage differences between the expected and observed fractions of female authors publishing in bioRxiv, medRxiv, and selected Springer-Nature journals between March and August 2020. Points to the left of the midline (orange) represent countries with less than expected fractions of female authors, and points to the right of the midline (in green) represent countries an increase in the fractions of female authors. The left-hand plots are for all papers regardless of topic, the middle plots are only for COVID-19-related papers, and the right-hand plots are only for papers that are not related to COVID-19. More detailed information is provided in Tables S8-S10 in Multimedia Appendix 1.

A significant drop in the proportion of female first authors was consistent across the countries. Regardless of the topic, we observed a 24.9% drop in Italy (y_{est}=0.526; y=0.395; S_{est}=0.046; σ_x=0.024), followed by Canada (19.7%), Sweden (15.9%), Japan (14.5%), India (13.4%), and France (11.1%). For research dealing explicitly with the topic of COVID-19 (see Figure 3, middle panels), we observed a greater gender gap than with papers on other research topics (see Figure 3, right-hand panels). In Germany, for example, the relative drop in the proportion of female first authors was 36% (y_{est}=0.39; y=0.25; S_{est}=0.02; σ_x=0.027), indicating that male scientists affiliated with German institutions are publishing disproportionately more than their female colleagues about COVID-19. Similar considerations applied to India, France, Italy, Great Britain, Canada, and the United States. The opposite was true for Japan, where the proportion of women publishing about COVID-19 as the first author increased by 23.7%. A similar disparity applied to the last authors (see Figure 3, second row). Missing points indicate that there were not enough data from the pandemic period to calculate the observed mean.

When we observed the proportion of female authors regardless of the authorship order, the drop became less prominent but still consistent across the countries. For example, in Canada, a drop in the proportion of female authors for COVID-19-related papers was 15.7% (y_{est}=0.387; y=0.318; S_{est}=0.014; σ_x=0.018), with similar-sized drops for Germany, Italy, Great Britain, and France (see Figure 3, third row). On the other hand, there was an increase of 6% in the proportion of female authors writing about COVID-19 in Japan (y_{est}=0.155; y=0.164; S_{est}=0.010; σ_x=0.017). The increase in the proportion of female authors in China (2.4%) was within the margin of error.

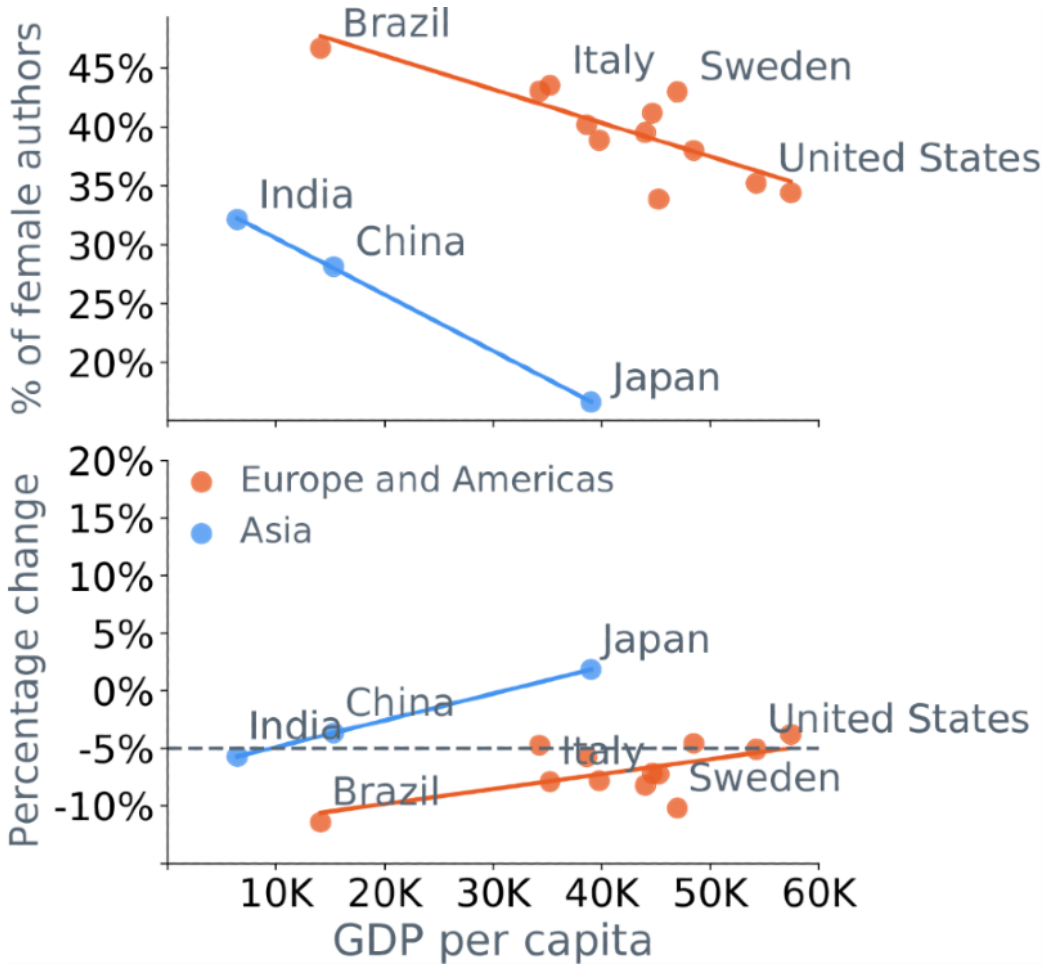
The gender gap for non-COVID-19-related research (see Figure 3, right-hand panels) was found to exist during the pandemic, but it is smaller than for COVID-19 research. Again, we observed stark differences between countries, with the proportion of female first authors publishing during the pandemic significantly decreasing in Italy, Canada, Japan, and France. Note that the plots for the single-author papers are missing, as the samples became too small when disaggregated by country.

Figure 3. Percentage drop in proportion of female authors during the pandemic across countries. Orange points mark the percentage decrease in proportion of female authors; green points mark the increase. Horizontal lines represent standard errors. The analysis is divided by topic into three groups: (1) all papers from the data set, (2) papers that deal directly with COVID-19 and related topics, and (3) papers that are not about COVID-19 or related topics. Missing points indicate insufficient sample size.



Further, we explored whether there were any commonalities among the countries with respect to the participation of women in research. Figure 4 illustrates the proportion of female authors (upper panel) and the percentage change in the proportion of female authors (lower panel) as a function of GDP. When disaggregated by region, we observed that the wealthier countries—those with higher per capita GDP—had proportionally fewer female researchers, with Asian countries exhibiting the most pronounced gender disparity. However, the countries with a higher GDP per capita demonstrated a smaller drop in the proportion of women publishing during the pandemic.

Figure 4. Gender disparity in research and gross domestic product (GDP). The proportion of women active in research is higher in countries with lower per capita GDP (upper). The proportion of female authors of research articles decreased more than expected in countries with lower per capita GDP (lower).



Discussion

Principal Findings

We analyzed bibliographical data from biomedical preprint servers and Springer-Nature journals and showed that the fraction of women publishing during the COVID-19 pandemic dropped significantly across disciplines and research topics. Since the announcement of the global pandemic and the start of lockdowns, we observed a drop of 9.1% in the number of women publishing biomedical scientific papers as the first author. Women were significantly excluded from COVID-19-related research, as we measured a 28% drop in female first authors in that area of research. This confirms some earlier suggestions that female first authors contributed less to COVID-19 studies than to research in other areas [25]. Women remain underrepresented, even though we observed an increased publishing rate for both genders during the pandemic. A similar disparity can be observed for last authors as well as solo-authored papers. The increased gender gap in publishing is persistent across the 10 countries with the highest number of researchers.

For papers on topics other than COVID-19, we did not observe this high discrepancy, and, in the case of medRxiv, we observed more women than projected by the model. The overall gender disparity in research during the pandemic was mostly driven by the higher publication rate of papers on COVID-19 and related topics. It seems that such research is conducted disproportionately by men, as male authors are more likely to appear in first author positions on papers posted on preprint servers and published in peer-reviewed journals.

It appears that the most significant drop in proportion of female authors happened early in the pandemic. The proportion of women has been increasing gradually for some authorship categories. Note that the observed gradual increase is statistically significant but is very slow. One can think that a possible explanation for such a sudden drop and a subsequent gradual increase is that most of the COVID-19 papers published early during the pandemic were various epidemic models focusing on cases and death counts. Many of the authors' affiliations were departments of engineering, mathematics, and physics, which might have a different proportion of women than the population of scientists in biology and medicine. Since research in the biomedical field usually takes longer to conduct and publish, it could lead to a shift in the gender distribution later. However, this argument does not explain the phenomenon entirely, as the base gender gap in science, technology, engineering, and mathematics fields is not higher than in biology [31]; therefore, future publications from biologists are not expected to narrow the gender gap. On the contrary, they might even increase it.

Another likely explanation of a sudden drop in the proportion of female authors is that caregiving demands have exploded during the pandemic, and these have mostly fallen on women [12,36,37]. These include childcare demands [38], elderly care, and other kinds of domestic and emotional labor. Sudden lockdowns and other preventive measures unevenly increased the burden on certain populations, causing the productivity of female scientists to decrease. As the world started fighting off the pandemic, people got used to the "new normal" and scientists started returning to their routines. That can partially explain the gradual increase in the proportion of female authors. Nevertheless, further research and more time is needed to investigate the reasons for such a sudden drop and gradual revival of the proportion of papers published by female scientists during the pandemic.

The global pandemic has touched almost every nation on the planet. Countries, however, responded differently in containing the spread of the disease. The variability of the measures and their timing, combined with differences in cultural norms and outbreak severity, have had a variable impact on researchers across the world. Country-level analysis better reveals global trends, as the aggregate data can be skewed by countries with a disproportionately large number of publications, such as the United States, which represents almost 29% of all authors in the data set (see Table S11 in Multimedia Appendix 1). Additionally, our analysis can reveal regional, political, and cultural differences between the nations. It is known that gender disparities in research are strongly associated with a country's wealth [39]. The wealthier countries—those with higher per capita GDP—have proportionally fewer female researchers, with Asian countries exhibiting the most pronounced gender disparity. However, the countries with higher GDP per capita were more resilient to the effects of COVID-19 on gender imbalance. In addition, wealthier countries showed a smaller pandemic-related drop in women's participation in research than poorer countries, with wealthier Asian countries experiencing an increase in the proportion of active female researchers. This suggests that women experience bigger life disruptions in poorer countries, which affects their productivity. Additionally, women are more excluded from COVID-19 research in poorer nations. This certainly should not imply any purpose or deliberate action, but rather the disproportionate variations in the social environments across nations, caused by the various expectations for the female members of households.

Implications

Gender imbalances in academia have been evident historically and still persist today. Various measures of research output, including the proportion of authors, fractionalized authorships [40], tenure decisions, and number of research grants [41], indicate the significant gender gap that is observed worldwide. For the past 60 years, we have witnessed an increase in participation by women in science across scientific disciplines [31] and lower levels of discrimination [42], demonstrating that the gender gap is shrinking over time [43]. Thus, a sudden drop in women's research output in biomedical research about COVID-19 appears as a surprising reverse trend.

The factors that led to such extreme and consistent differences in the proportion of female scientists can be numerous. The already existing barriers for female participation in science vary across countries. In some nations, men are more favorably placed than women [44-47] and can be more likely to receive quick funding for COVID-19-related research. Additionally, traditional gender norms differ and can affect the genders differently. Caregiving demands have mostly fallen on women. At the same time, new challenges bring new opportunities, and men who are likely to engage in more aggressive self-promotion [48,49] and pursue careers more forcefully [43] can be motivated to push for faster publication. Identifying the exact reasons for an increased gender gap can be an important topic for future studies.

The global pandemic caused this unforeseen crisis that will most certainly affect academia. All the difficulties female scientists faced previously may possibly be exacerbated by the extended lockdowns and sudden shift in work-life dynamics. It is important to understand the impact of such an extraordinary circumstance on the scientific community that will disproportionately affect research outputs as well as prospects for tenure and promotions [21]. Future research evaluation practices should be informed by our findings to account for and mitigate the penalizing effects that COVID-19 is having on female researchers.

Strengths and Limitations of the Study

The strengths of our study include the use of a relatively large and diverse data set from three different publishing platforms. The focus on preprint papers allows for the assessment of the observed effects in a timely manner. We focused on a structured and rigorous statistical analysis, making sure that the results are significant. The data and the code to reproduce the results are available.

Potential limitations warrant consideration. First, the gender of a publication's author can be wrongly identified. Even though we excluded the results that had a low confidence, a small fraction of the authors could have been misgendered. Additionally, we acknowledge that automated gender classifiers do not recognize the various nonbinary gender identities [50], and we assigned the gender label based on the historical distribution of typical male and female names. As awareness of the nature of gender and identity shift, so may the number of researchers who do not identify within the binary categories of male and female. Such researchers face additional layers of discrimination that our study does not consider. While we understand that binary gender can be an oversimplification that can introduce some amount of bias and inaccuracy, the problem that we highlight hopefully can bring some attention to the multifaceted issue of gender, identity, and discrimination. Second, the algorithm that identified the authors' countries relies on recognizing the names of the toponyms in the names of the authors' affiliated institutions. Even though we made sure that the most popular institutions were properly localized and we optimized the localization resolution, some errors are possible. Third, throughout the paper, the word "productivity" was used to refer to the rate of publication output of scientists in terms of publications per week and it did not capture changes to

scientists' other inputs. For example, female scientists appearing less productive in terms of publications per week may simply reflect that they were not able to spend as much time on their research (ie, hours worked were not captured). We are aware that there are other preprint servers and journals that publish papers in the biomedical field. By analyzing the data from the two largest preprint servers and the largest publisher of peer-reviewed papers, we aimed to cover a representative sample of papers and authors from the field. Finally, our analysis was focused on the first 6 months of the pandemic and might not accurately evaluate the effects that can be observed later in the pandemic.

Conclusions

Our findings documented a decrease in the proportion of female authors in the biomedical field who published research papers during the global pandemic. This effect was particularly pronounced for papers related to COVID-19, indicating that women are producing fewer publications related to COVID-19 research. A sudden increase in this gender gap was persistent across the 10 countries with the highest number of researchers. The results should be used to inform the scientific community of this worrying trend in COVID-19 research and the disproportionate effect the pandemic has had on female academics' research outputs.

Acknowledgments

The authors are grateful to the Defense Advanced Research Projects Agency (DARPA) (contract W911NF192027) for their support.

Authors' Contributions

All authors conceived and designed the study. GM collected and analyzed the data. All authors wrote and revised the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary materials.
[PDF File (Adobe PDF File), 719 KB - jmir_v23i4e25379_app1.pdf]

References

1. Nicola M, Alsafi Z, Sohrabi C, Kerwan A, Al-Jabir A, Iosifidis C, et al. The socio-economic implications of the coronavirus pandemic (COVID-19): A review. Int J Surg 2020 Jun;78:185-193 [FREE Full text] [doi: 10.1016/j.ijsu.2020.04.018] [Medline: 32305533]
2. Atkeson A. What Will Be the Economic Impact of COVID-19 in the US? Rough Estimates of Disease Scenarios. Cambridge, MA: National Bureau of Economic Research; 2020 Mar. URL: https://doi.org/10.3386/w26867 [accessed 2020-06-27]
3. Fernandes N. Economic effects of coronavirus outbreak (COVID-19) on the world economy. IESE Business School working paper No. WP-1240-E. SSRN 2020 Mar 22:1 [FREE Full text] [doi: 10.2139/ssrn.3557504]
4. Alon T, Doepke M, Olmstead-Rumsey J, Tertilt M. The impact of COVID-19 on gender inequality. Covid Economics, Vetted And Real-Time Papers. Issue 4. London, UK: CEPR Press; 2020 Apr 14. URL: https://cepr.org/file/9014/download?token=S-8mCQli [accessed 2021-04-07]
5. Collins C, Landivar L, Ruppanner L, Scarborough WJ. COVID-19 and the gender gap in work hours. Gend Work Organ 2020 Jul 02:1 [FREE Full text] [doi: 10.1111/gwao.12506] [Medline: 32837019]
6. Feng Z, Savani K. Covid-19 created a gender gap in perceived work productivity and job satisfaction: Implications for dual-career parents working from home. Gend Manag 2020 Sep 07;35(7/8):719-736 [FREE Full text] [doi: 10.1108/gm-07-2020-0202]
7. Kalyanam J, Quezada M, Poblete B, Lanckriet G. Prediction and characterization of high-activity events in social media triggered by real-world news. PLoS One 2016;11(12):e0166694 [FREE Full text] [doi: 10.1371/journal.pone.0166694] [Medline: 27992437]
8. Qian Y, Fuller S. COVID-19 and the gender employment gap among parents of young children. Can Public Policy 2020 Aug 01;46(S2):S89-S101 [FREE Full text] [doi: 10.3138/cpp.2020-077]
9. Fodor E, Gregor A, Koltai J, Kováts E. The impact of COVID-19 on the gender division of childcare work in Hungary. Eur Societies 2020:95-110 [FREE Full text] [doi: 10.1080/14616696.2020.1817522]
10. Czymara CS, Langenkamp A, Cano T. Cause for concerns: Gender inequality in experiencing the COVID-19 lockdown in Germany. Eur Societies 2020:S68-S81 [FREE Full text] [doi: 10.1080/14616696.2020.1808692]
11. Kristal T, Yaish M. Does the coronavirus pandemic level the gender inequality curve? (It doesn't). Res Soc Stratif Mobil 2020 Aug;68:100520 [FREE Full text] [doi: 10.1016/j.rssm.2020.100520]

12. Kowal M, Coll-Martín T, Ikizer G, Rasmussen J, Eichel K, Studzińska A, et al. Who is the most stressed during the COVID-19 pandemic? Data from 26 countries and areas. Appl Psychol Health Well Being 2020 Dec;12(4):946-966 [FREE Full text] [doi: 10.1111/aphw.12234] [Medline: 32996217]
13. Craig L, Churchill B. Dual-earner parent couples' work and care during COVID-19. Gend Work Organ 2020 Jun 25:1 [FREE Full text] [doi: 10.1111/gwao.12497] [Medline: 32837023]
14. Myers KR, Tham W, Yin Y, Cohodes N, Thursby J, Thursby M, et al. Unequal effects of the COVID-19 pandemic on scientists. Nat Hum Behav 2020 Sep;4(9):880-883. [doi: 10.1038/s41562-020-0921-y] [Medline: 32669671]
15. Minello A. The pandemic and the female academic. Nature 2020 Apr 17:1 [FREE Full text] [doi: 10.1038/d41586-020-01135-9] [Medline: 32303729]
16. Viglione G. Are women publishing less during the pandemic? Here's what the data say. Nature 2020 May;581(7809):365-366 [FREE Full text] [doi: 10.1038/d41586-020-01294-9] [Medline: 32433639]
17. Amano-Patiño N, Faraglia E, Giannitsarou C, Hasna Z. The Unequal Effects of COVID-19 on Economists' Research Productivity. Cambridge Working Papers in Economics: 2038. Cambridge, UK: University of Cambridge; 2020 May 07. URL: https://www.repository.cam.ac.uk/bitstream/handle/1810/310888/cwpe2038.pdf?sequence=1&isAllowed=y [accessed 2020-04-07]
18. Cui R, Ding H, Zhu F. Gender inequality in research productivity during the COVID-19 pandemic. SSRN 2020 Jun 09:1 [FREE Full text] [doi: 10.2139/ssrn.3623492]
19. Fuchs-Schuendeln N. Gender structure of paper submissions at the Review of Economic Studies during COVID-19: First evidence. Rev Econ Stud 2020:1-6 [FREE Full text]
20. Fazackerley A. Women's research plummets during lockdown - But articles from men increase. The Guardian. 2020 May 12. URL: https://www.theguardian.com/education/2020/may/12/womens-research-plummets-during-lockdown-but-articles-from-men-increase [accessed 2020-06-07]
21. Oleschuk M. Gender equity considerations for tenure and promotion during COVID-19. Can Rev Sociol 2020 Aug;57(3):502-515 [FREE Full text] [doi: 10.1111/cars.12295] [Medline: 32779307]
22. Dolan K, Lawless JL. It takes a submission: Gendered patterns in the pages of AJPS. Am J Pol Sci 2020 Apr 20:1 [FREE Full text]
23. Shurchkov O. Is COVID-19 turning back the clock on gender equality in academia? Medium. 2020 Apr 23. URL: https://medium.com/@olga.shurchkov/is-covid-19-turning-back-the-clock-on-gender-equality-in-academia-70c00d6b8ba1 [accessed 2021-02-06]
24. Amano-Patiño N, Faraglia E, Giannitsarou C, Hasna Z. Who is doing new research in the time of COVID-19? Not the female economists. In: Galliani S, Panizza U, editors. Publishing and Measuring Success in Economics. London, UK: CEPR Press; 2020:137-142.
25. Vincent-Lamarre P, Sugimoto CR, Larivière V. The decline of women's research production during the coronavirus pandemic. Nature Index. 2020 May 19. URL: https://www.natureindex.com/news-blog/decline-women-scientist-research-publishing-production-coronavirus-pandemic [accessed 2020-06-07]
26. Andersen JP, Nielsen M, Simone N, Lewiss R, Jagsi R. COVID-19 medical papers have fewer women first authors than expected. Elife 2020 Jun 15;9:e58807 [FREE Full text] [doi: 10.7554/eLife.58807] [Medline: 32538780]
27. Pinho-Gomes AC, Peters S, Thompson K, Hockham C, Ripullone K, Woodward M, et al. Where are the women? Gender inequalities in COVID-19 research authorship. BMJ Glob Health 2020 Jul;5(7):e002922 [FREE Full text] [doi: 10.1136/bmjgh-2020-002922] [Medline: 32527733]
28. Abdill RJ, Blekhman R. Tracking the popularity and outcomes of all bioRxiv preprints. Elife 2019 Apr 24;8:e45133 [FREE Full text] [doi: 10.7554/eLife.45133] [Medline: 31017570]
29. Our World in Data. 2020. URL: https://ourworldindata.org/ [accessed 2021-03-30]
30. Santamaría L, Mihaljević H. Comparison and benchmark of name-to-gender inference services. PeerJ Comput Sci 2018;4:e156 [FREE Full text] [doi: 10.7717/peerj-cs.156] [Medline: 33816809]
31. Huang J, Gates AJ, Sinatra R, Barabási AL. Historical comparison of gender inequality in scientific careers across countries and disciplines. Proc Natl Acad Sci U S A 2020 Mar 03;117(9):4609-4616 [FREE Full text] [doi: 10.1073/pnas.1914221117] [Medline: 32071248]
32. Seabold S, Perktold J. Statsmodels: Econometric and statistical modeling with Python. In: Proceedings of the 9th Python Science Conference. 2010 Presented at: 9th Python Science Conference; June 28-July 3, 2010; Austin, TX p. 92 URL: https://conference.scipy.org/proceedings/scipy2010/pdfs/seabold.pdf [doi: 10.25080/majora-92bf1922-011]
33. Hausman C, Rapson D. Regression discontinuity in time: Considerations for empirical applications. Annu Rev Resour Economics 2018 Oct 05;10(1):533-552. [doi: 10.1146/annurev-resource-121517-033306]
34. rdd 0.0.3. Python Package Index. URL: https://pypi.org/project/rdd/ [accessed 2020-09-05]
35. Gender disparity in the authorship of biomedical research publications during the COVID-19 pandemic. GitHub. 2020. URL: https://github.com/gmuric/CovidGenderGap [accessed 2021-03-30]
36. Myers K, Tham WY, Yin Y, Cohodes N, Thursby JG, Thursby M, et al. Quantifying the immediate effects of the COVID-19 pandemic on scientists. SSRN 2020:1. [doi: 10.2139/ssrn.3608302]

37. Malisch JL, Harris BN, Sherrer SM, Lewis KA, Shepherd SL, McCarthy PC, et al. Opinion: In the wake of COVID-19, academia needs new solutions to ensure gender equity. Proc Natl Acad Sci U S A 2020 Jul 07;117(27):15378-15381 [FREE Full text] [doi: 10.1073/pnas.2010636117] [Medline: 32554503]
38. Minello A, Martucci S, Manzo L. The pandemic and the academic mothers: Present hardships and future perspectives. Eur Societies 2020 Aug 28;23(sup1):S82-S94 [FREE Full text] [doi: 10.1080/14616696.2020.1809690]
39. Holman L, Stuart-Fox D, Hauser CE. The gender gap in science: How long until women are equally represented? PLoS Biol 2018 Apr;16(4):e2004956 [FREE Full text] [doi: 10.1371/journal.pbio.2004956] [Medline: 29672508]
40. Larivière V, Ni C, Gingras Y, Cronin B, Sugimoto C. Bibliometrics: Global gender disparities in science. Nature 2013 Dec 12;504(7479):211-213. [doi: 10.1038/504211a] [Medline: 24350369]
41. Shen H. Inequality quantified: Mind the gender gap. Nature 2013 Mar 07;495(7439):22-24. [doi: 10.1038/495022a] [Medline: 23467149]
42. Ceci SJ, Williams WM. Understanding current causes of women's underrepresentation in science. Proc Natl Acad Sci U S A 2011 Feb 22;108(8):3157-3162 [FREE Full text] [doi: 10.1073/pnas.1014871108] [Medline: 21300892]
43. Lerchenmueller MJ, Sorenson O, Jena AB. Gender differences in how scientists present the importance of their research: Observational study. BMJ 2019 Dec 16;367:l6573 [FREE Full text] [doi: 10.1136/bmj.l6573] [Medline: 31843745]
44. Ranga M, Gupta N, Etzkowitz H. Gender Effects in Research Funding. Bonn, Germany: Deutsche Forschungsgemeinschaft (DFG); 2012 Mar. URL: https://www.dfg.de/download/pdf/dfg_im_profil/geschaeftsstelle/publikationen/studien/studie_gender_effects.pdf [accessed 2021-04-07]
45. Witteman HO, Hendricks M, Straus S, Tannenbaum C. Are gender gaps due to evaluations of the applicant or the science? A natural experiment at a national funding agency. Lancet 2019 Feb 09;393(10171):531-540. [doi: 10.1016/S0140-6736(18)32611-4] [Medline: 30739688]
46. Waisbren SE, Bowles H, Hasan T, Zou K, Emans S, Goldberg C, et al. Gender differences in research grant applications and funding outcomes for medical school faculty. J Womens Health (Larchmt) 2008 Mar;17(2):207-214. [doi: 10.1089/jwh.2007.0412] [Medline: 18321172]
47. Oliveira DFM, Ma Y, Woodruff T, Uzzi B. Comparison of National Institutes of Health grant amounts to first-time male and female principal investigators. JAMA 2019 Mar 05;321(9):898-900 [FREE Full text] [doi: 10.1001/jama.2018.21944] [Medline: 30835300]
48. Moss-Racusin CA, Rudman L. Disruptions in women's self-promotion: The backlash avoidance model. Psychol Women Q 2010 Jun;34(2):186-202. [doi: 10.1111/j.1471-6402.2010.01561.x]
49. Scharff C. Blowing your own trumpet: Exploring the gendered dynamics of self-promotion in the classical music profession. Sociol Rev 2015 May;63(1_suppl):97-112. [doi: 10.1111/1467-954x.12243]
50. Keyes O. The misgendering machines. Proc ACM Hum Comput Interact 2018 Nov;2(CSCW):1-22. [doi: 10.1145/3274357]

Abbreviations

API: application programming interface
DARPA: Defense Advanced Research Projects Agency
GDP: gross domestic product
OLS: ordinary least squares
RDD: regression discontinuity design

Edited by C Basch; submitted 30.10.20; peer-reviewed by J Zhang, E Oghogho, SGS Shah; comments to author 18.11.20; revised version received 09.12.20; accepted 14.03.21; published 12.04.21.

Please cite as:
Muric G, Lerman K, Ferrara E
Gender Disparity in the Authorship of Biomedical Research Publications During the COVID-19 Pandemic: Retrospective Observational Study
J Med Internet Res 2021;23(4):e25379
URL: https://www.jmir.org/2021/4/e25379
doi:10.2196/25379
PMID:33735097

Original Paper

An Automatic Ontology-Based Approach to Support Logical Representation of Observable and Measurable Data for Healthy Lifestyle Management: Proof-of-Concept Study

Ayan Chatterjee[1], MEng; Andreas Prinz[1], PhD; Martin Gerdes[1], PhD; Santiago Martinez[2], PhD

[1]Department of Information and Communication Technologies, Centre for e-Health, University of Agder, Grimstad, Norway

[2]Department of Health and Nursing Science, Centre for e-Health, University of Agder, Grimstad, Norway

Corresponding Author:
Ayan Chatterjee, MEng
Department of Information and Communication Technologies
Centre for e-Health
University of Agder
Jon Lilletuns Vei 9
Grimstad, 4879
Norway
Phone: 47 38141000
Email: ayan.chatterjee@uia.no

Abstract

Background: Lifestyle diseases, because of adverse health behavior, are the foremost cause of death worldwide. An eCoach system may encourage individuals to lead a healthy lifestyle with early health risk prediction, personalized recommendation generation, and goal evaluation. Such an eCoach system needs to collect and transform distributed heterogenous health and wellness data into meaningful information to train an artificially intelligent health risk prediction model. However, it may produce a data compatibility dilemma. Our proposed eHealth ontology can increase interoperability between different heterogeneous networks, provide situation awareness, help in data integration, and discover inferred knowledge. This "proof-of-concept" study will help sensor, questionnaire, and interview data to be more organized for health risk prediction and personalized recommendation generation targeting obesity as a study case.

Objective: The aim of this study is to develop an OWL-based ontology (UiA eHealth Ontology/UiAeHo) model to annotate personal, physiological, behavioral, and contextual data from heterogeneous sources (sensor, questionnaire, and interview), followed by structuring and standardizing of diverse descriptions to generate meaningful, practical, personalized, and contextual lifestyle recommendations based on the defined rules.

Methods: We have developed a simulator to collect dummy personal, physiological, behavioral, and contextual data related to artificial participants involved in health monitoring. We have integrated the concepts of "Semantic Sensor Network Ontology" and "Systematized Nomenclature of Medicine—Clinical Terms" to develop our proposed eHealth ontology. The ontology has been created using Protégé (version 5.x). We have used the Java-based "Jena Framework" (version 3.16) for building a semantic web application that includes resource description framework (RDF) application programming interface (API), OWL API, native tuple store (tuple database), and the SPARQL (Simple Protocol and RDF Query Language) query engine. The logical and structural consistency of the proposed ontology has been evaluated with the "HermiT 1.4.3.x" ontology reasoner available in Protégé 5.x.

Results: The proposed ontology has been implemented for the study case "obesity." However, it can be extended further to other lifestyle diseases. "UiA eHealth Ontology" has been constructed using logical axioms, declaration axioms, classes, object properties, and data properties. The ontology can be visualized with "Owl Viz," and the formal representation has been used to infer a participant's health status using the "HermiT" reasoner. We have also developed a module for ontology verification that behaves like a rule-based decision support system to predict the probability for health risk, based on the evaluation of the results obtained from SPARQL queries. Furthermore, we discussed the potential lifestyle recommendation generation plan against adverse behavioral risks.

Conclusions: This study has led to the creation of a meaningful, context-specific ontology to model massive, unintuitive, raw, unstructured observations for health and wellness data (eg, sensors, interviews, questionnaires) and to annotate them with semantic metadata to create a compact, intelligible abstraction for health risk predictions for individualized recommendation generation.

(J Med Internet Res 2021;23(4):e24656) doi:10.2196/24656

KEYWORDS

activity; nutrition; sensor; questionnaire; SSN; ontology; SNOMED CT; eCoach; personalized; recommendation; automated; CDSS; healthy lifestyle; interoperability; eHealth; goal setting; semantics; simulation; proposition

Introduction

Overview

Lifestyle diseases are an economic burden to an individual, household, employer, and government, and lead to financial and productivity risks for poor and rich countries alike [1-3]. The key risk factors behind lifestyle diseases are the excessive use of alcohol, inappropriate food plan, physical inactivity, excessive salt intake, saturated fat consumption, and tobacco use [1-3]. These result in excess weight gain, elevated blood glucose, high blood pressure (BP), elevated total cholesterol in the blood, and social isolation. Obesity is one of the foremost lifestyle diseases that lead to other noncommunicable diseases such as cardiovascular diseases, chronic obstructive pulmonary disease, cancer, diabetes type II, hypertension, and depression [1-3]. eHealth monitoring has become increasingly popular, providing information and communications technology (ICT)–based remote, timely care support to patients and health care providers [1-3]. An eHealth virtual coaching recommendation system can guide people and convey the appropriate recommendations in context with enough time to prevent and improve living with lifestyle diseases. It requires capturing physiological (vital signs such as BP, pulse, lipid profile, glycemic response, BMI), behavioral (sleep, diet, exercise), and contextual data (position, and weather) from secure wearable sensors, manual interactions, feedback, and customized questionnaires over time, to train an artificial intelligence (AI) model for behavior analysis and early prediction of wellness trends and risks [4-6]. However, data collection from heterogenous sources may lead to data interoperability, annotation, and semantization problem.

Background and Problem Description

Health and wellness data collected from heterogeneous sources (eg, multimodal sensors, interviews, questionnaires) are of different format and lead to well-known problems in health informatics, which are related to logical data representation, aggregation, data analysis, data standardization, and data interoperability [7,8]. Targeted personal, habitual, physiological, activity, and nutrition data are generally collected via secure wearable sensors, manual interactions, interviews, web-based interactions, smartphone apps, customized questionnaires, and feedback forms over time. Weather application programming interfaces (APIs) and external weather sensors are useful for the collection of contextual weather data over time. The wearable activity monitors need to connect to a personal smartphone via Bluetooth nearfield communication technology (Bluetooth low energy [BLE]) [9,10]. The device can seamlessly measure and transfer high-resolution raw acceleration data and multiple activity parameters to a secure storage to process the data further with a machine intelligence module [11]. High-end, time-dependent activity data collection with wearable BLE devices has become accessible and feasible for ubiquitous monitoring. Some of the activity data, such as nonwear time or intensive activity details, are questionnaire dependent.

Physiological data are collected either invasively (eg, glycemic response, cholesterol level) or noninvasively (eg, weight, BP, heart rate, body assessment data). The questionnaire-dependent nutrition data are collected either daily or on an alternate day or on a weekly basis. The assessment of nutrition data helps to determine the type of food, amount of food, conceptual information (temporal/spatial), dietary pattern, and intake of alcohol or energy drinks. Some baseline data (medical history, habit, preference, personal details, initial weight and height, initial BP, and initial body assessment data) are collected during the initial recruitment of the participant or every month for either demographic statistics or population clustering or individual goal assessment. Each data have their unit and range following a standard guideline based on the context and domain (eg, data on temperature are applicable for both health and environment domain with a different range, meaning, and context). Therefore, each measurement process owns separate challenges related to logical or semantic data representation, proper usage of data, and improving data reusability. The data usability involves the transformation of data into an understandable computer format. It creates a challenge to systematically and syntactically analyze health and wellness data in aggregation with other clinical data. Incorporation of physical activity, diet as a care procedure, or investigating how it afflicts healthy outcomes involves a more detailed and diverse representation of participant's behavioral level and physiological condition [7,8,12,13].

Furthermore, the challenges of reusing the existing physiological and behavioral data of a participant within the electronic health record remain and include concerns related to opacity and semantic inconsistency [7,8]. Besides, these health and wellness data are still mostly hidden in clinical narratives with highly variable forms of expression. In this regard, ontology can provide a framework to allow the mentioned heterogeneous health and wellness data to be organized, compact, structured, consistent, machine understandable, and queried through high-level specifications. Ontology helps to annotate diverse health and wellness data with semantic metadata to increase interoperability among heterogeneous networks, data integration, discovery, and situation awareness. An eHealth ontology can reuse the concept of existing, proven, well-accepted ontologies (eg, semantic sensor network [SSN] ontology [14], Systematized Nomenclature of Medicine—Clinical Terms [SNOMED CT] ontology [15]) to enhance its vocabularies and better semantic representation.

A rule-based decision support system (DSS) can use such an eHealth ontology model to measure and predict health risks, and to generate useful personalized recommendations following proven clinical rules. If the collected health and wellness data are not annotated accurately with semantic metadata in the

medical domain, then the DSS may fail to deliver accurate decisions to both physicians and patients or participants in the form of incorrect recommendation plan, goal setting, and goal evaluation. DSS decision inaccuracy may appear primarily due to the following effects—improper design of knowledge base (KB), the inadequacy of tools or technologies applied in the execution of DSS, problems related to the ontology reasoning engine, and issues associated with inferring new knowledge.

Aim of the Study

After studying existing ontology models, we found that many ontologies and regulated terminologies cover aspects of obesity and related chronic illness domains, but concept analysis remains incomplete. After reviewing relevant ontologies, we proposed a freshly created OWL-based ontology to deal with different data inputs (internet of things [IoT] sensors, interviews, and questionnaires) and annotate them with semantic data. The proposed ontology will support data interoperability, logical representation of collected health and wellness data in context, and to build a rule-based DSS for health risk prediction related to obesity and afterward generation of lifestyle recommendations for a healthy lifestyle.

We have not evaluated the impact of the suggested recommendations on participants as we executed the complete scenario under a simulated environment. Still, we evaluated the performance of the proposed ontology model. In the proposed ontology, we annotated every participant's data with semantic web language rules and stored the generated OWL file in a triple-store format for better readability (Multimedia Appendix 1). The proposed ontology model allows automatic inferencing, efficient knowledge representation, balancing a trade-off between complexity and eloquence, and reasoning about formal knowledge. The entire study is divided into the following 2 segments: (1) ontology design and development and (2) its verification. This study addresses the following identified research questions:

(RQ-1) How to annotate distributed, heterogenous health and wellness data received from sensors, questionnaires, and interviews into meaningful information to build a future machine learning model for health risk prediction for obesity?

(RQ-2) How to integrate existing IoT and medical ontologies to design and develop proposed eHealth ontology for obesity study case?

(RQ-3) How to verify the proposed ontology with rule-based behavioral recommendation generation?

For this set of semantic data, which will be considered as asserted true facts, the primary goal of the paper is to trigger logical rules of the shape (A IMPLIES B) or trigger rules in a logically equivalent way, that is, (NOT(A) OR B). If some specific variables are inferred to be true, then some recommendations shall be provided to the user from whom the semantic data are originating.

Related Work

This section offers existing background knowledge applicable for this research. It includes (1) a discussion of existing, relevant eHealth ontology models for chronic illness, health monitoring, and ontology-based DSS, (2) ontologies in the IoT domain for modeling sensor data, and (3) ontologies in the medical domain.

Existing eHealth Ontology Models

Different research groups have conducted different studies on eHealth ontology modeling for chronic illness, health monitoring, and ontology-based clinical decision support system (CDSS). For example, Kim et al [16] developed an ontology model for obesity management with the nursing process in the mobile device domain for spontaneous participant engagement and continuous weight monitoring. The scope of the obesity management included behavioral interventions, dietary recommendations, and physical activity, and for this purpose, the study included assessment data (BMI, sex, and hip-to-waist circumference), inferred data for representing diagnosis results, evaluations (cause of obesity, success, or failure of behavioral modifications), and implementation (education, suggestion, intervention). Sojic et al [17] modeled an obesity domain–specific ontology with OWL to design inference patterns to individualize health condition assessment as age and gender specific. The ontology helped classify personal profiles based on the changes of personal behavior or feature over time and infer personal health status automatically, which are important for obesity evaluation and prevention. The ontology rules were written in semantic web rule language (SWRL). Kim et al [18] proposed an ontology model for physical activity (PACO) to support physical activity data interoperability. The ontology was developed in Protégé (version 4.x), and the FaCT++ reasoner verified its structural consistency. Lasierra et al [19] developed an automatic ontology–based approach to manage information in home-based scenarios for telemonitoring services based on the automatic computing paradigm, namely, MAPE (monitor, analyze, plan, and execute). They proposed another 3-stage ontology-driven solution [20] (stage 1: ontology design and implementation; stage 2: ontology application to study personalization issues; and stage 3: software prototype implementation) for giving personalized care to chronic patients at home. The proposed ontology was designed in OWL DL language in Protégé-OWL version 4.0.2 ontology editor and was verified using FACT++ reasoner. The ontology development involved data from heterogeneous sources, such as clinical knowledge, data from medical devices, and patient's contextual data. Yao and Kumar [21] proposed a novel CONFlexFlow (Clinical cONtext based Flexible workFlow) approach using ontology modeling for incorporating flexible and adaptive clinical pathways into CDSS. They developed 18 SWRL rules for practical explanation of heart failure. The model was verified with the Pellet Reasoner Plug-in for Protégé version 3.4. Additionally, they developed a "proof-of-concept" prototype of the proposed approach using the Drools framework. Chi et al [22] constructed a chronic disease dietary consultation system using web ontology language (OWL) and SWRL. The KB involved heterogeneous sources of data and interaction of factors, such as the illness stage, the physical condition of the patient, the activity level, the quantity of food intake, and the critical nutrient constraints. Rhayem et al [23] proposed an ontology-based system (HealthIoT) for patient monitoring with sensors, radiofrequency identification devices, and actuators. They claimed that data obtained from medically connected

devices are enormous, and thereby lack repressibility and understandability, and are manipulated by other systems and devices. Therefore, they proposed an ontology model to represent both the connected medical devices and their data based on a semantic rule, followed by model evaluation with the proposed IoT Medicare system that supports decision making after analyzing the vital signs of the patients. Galopin et al [24] proposed an ontology-based prototype CDSS to manage patients with multiple chronic disorders following clinical practice guidelines. The KB decision rules were based on the "if–then" rules following clinical practice guidelines and patient observation data. Sherimon et al [25] proposed an ontology system (OntoDiabetic) using OWL2 language to support a CDSS for patients with cardiovascular disease, diabetic nephropathy, and hypertension following clinical guidelines and "if–then" decision rules. Hristoskova et al [26] proposed another ontology-driven ambient intelligence framework to support personalized medical detection and alert generation based on the analysis of vital signs collected from the patients diagnosed with congestive heart failure. The DSS system can classify personalized congestive heart failure risk stages, and thereby, notify patients through ambient intelligence's inference engine. Riaño et al [27] proposed an ontology-based CDSS for monitoring and intervening chronically ill patients to prevent critical conditions, such as incorrect diagnoses, undetected comorbidities, missing information, and unobserved related diseases. Jin and Kim [7] designed and implemented an eHealth system using the IETF YANG ontology based on the SSN concept. The approach assisted in the autoconfiguration of eHealth sensors (responsible for collecting body temperature, BP, electromyography, and galvanic skin response) with the help of internet and communication technologies and querying the sensor network with semantic interoperability support for the proposed eHealth system. The proposed eHealth system consisted of 3 main components: SSN (eHealth sensors, patient, unified resource identifier [URI]), internet (eHealth server, KB), and eHealth clients (patient, and professionals). The proposed semantic model used a "YANG to JSON translator" to convert YANG semantic model data to JSON semantic model data for semantic interoperability before storing them in the database (KB). Ganguly et al [28] proposed an ontology-based model to manage semantic interoperability problems in eHealth in the context of diet management for diabetes. The development of the framework included rules of dialogue games, DSS with KB (rule base and database), a dialogue model based on decision mechanism, the syntax of dialogue game, decision mechanism, and translational rules.

Ontologies in the Internet of Things Domain

Ontology [29] provides a framework for describing sensors. SSN-XG (W3C Semantic Sensor Network Incubator Group) developed the SSN ontology to model sensor devices, systems, processes, and observations. SSN annotates sensor data with semantic metadata (semantic sensor web) to increase interoperability among diverse networks, data integration, discovery, and situation awareness. The Sensor Model Language (SensorML) was developed by the Open Geospatial Consortium (OGC), which provides syntactic descriptions using XML to describe sensors, observations, and measurements. While SensorML provides an XML schema for defining sensors, it lacks the repressibility provided by ontology languages such as OWL [30-32]. Semantic sensor web, a combination of sensor and semantic web technologies, helps to annotate spatial, temporal, and thematic semantic metadata for the more artistic representation of sensor data, advanced access, formal analysis of sensor resources, and data standardization. SSN ontology is used to describe sensor devices; sensing; sensor measurement capabilities; and sensor observations, process, and systems [30-32]. SSN allows its network, sensor devices, and data to be installed, structured, managed, queried, and controlled through high-level specifications. Sensors Annotation and Semantic Mapping Language offers XML schema to transfer sensor data and sources into the instances of SSN ontology based on a predefined XML-based document (resource description framework [RDF]), which is achieved automatically with sensor data to RDF mapping algorithm [33]. "M3 Ontology" (machine-to-machine) was developed based on the "SenML" protocol (designed for simple sensor measurement), which is an extension of SSN, to enable the interoperable design of domain-specific or cross-domain-specific applications which are termed as Semantic Web of Things [13]. AeroDAML, KIM, M3 Semantic Annotator, MnM, and SemTag are different available semantic annotators for sensor observations for their corresponding semantic models (DAML, KIMO, M3, Kmi, and TAP) [34]. Like SSN, there are other IoT-based contextual ontologies, such as IoT-Ontology, IoT-Lite, and IoT-O [35]. SCUPA, CoBrA-Ont, CoDAMoS, PalSPOT, the delivery context ontology, and Fuzzy-Onto are different IoT-based ontologies for activity recognition [34]. URI, HTTP, HTML5, REST, SOAP, Web Socket, Web feed, MQTT, CoAP, and AMQP are some standard IoT protocols applicable to Web of Things [14,34,36,37]. In this study, we integrated the concept of SSN ontology to model sensor observations.

Ontologies in Medical Domain

SNOMED CT, 11th edition of the International Classification of Diseases (ICD-11), Unified Medical Lexicon System (UMLS semantic network), Foundational Model of Anatomy, OpenEHR, Gene Ontology, DOLCE, Basic Formal Ontology, Cyc's upper ontology, Sowa's top-level ontology, the top level of GALEN, and Logical Observation Identifiers Names and Codes (LOINC) are biomedical ontologies introduced to deliver semantic interoperability and complete knowledge related to the specific biological and medical domains [38]. Most laboratory and clinical systems send out data using the HL7 (version 2) protocol and in an HL7 message, the LOINC codes represent the "question" for a laboratory test or experiment and the SNOMED CT code represents the "answer." In this study, we have reused the SNOMED CT ontology for modeling the health condition based on health and wellness data, and recommendation generation [8]. SNOMED CT was designed in 1965 as a controlled medical vocabulary licensed and supported by the International Health Terminology SDO. It is an organized list of a wide variety of clinical terminology defined with unique codes (ICD). It covers a wide range of medical terminologies for disorders and findings (what were observed!), procedures (what was done!), events (what happened!), substance/medication (what was consumed or administered!),

and anything related to medical data. It offers a shared language that enables a reliable way of indexing, storing, reclaiming, and accumulating clinical data across fields and care sites. It is a complete, multilingual clinical terminology that gives clinical content and clarity for clinical documentation and reporting [8,38,39].

As described above, most studies have developed ontologies using OWL to solve the data interoperability problem. Still, integration among the electronic health data, semantic rules, semantic annotation, clinical guidelines, health risk prediction, and personalized recommendation generation remains an issue in eHealth. This study addresses it and proposes a prototype ontology model for obesity as a case study, to integrate data from heterogeneous sources (eg, sensor, questionnaire, and interview) in order to enable data interoperability, information search and recovery, and automatic interference. We integrated SSN and SNOMED CT ontologies into our proposed eHealth ontology because of their vast vocabularies, appropriateness, and semantic capabilities as discussed above [40-43].

Methods

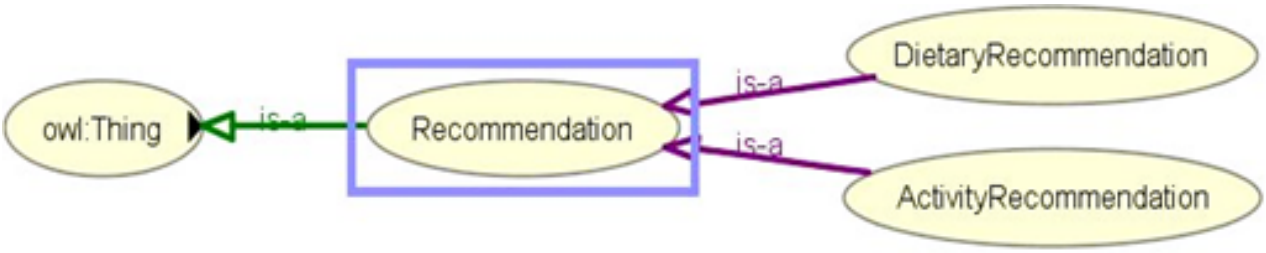
Basics of Ontology

Ontology commenced as a philosophical discipline studying the existence and being and expanded into information technologies. Ontology is a formalized model for specific domains with the following essential elements: individuals/objects, classes, attributes, relations, and axioms. A class diagram of a program written in object-oriented programming [44] is a visual representation of an ontology. Ontology is a philosophy that has been around for thousands of years, and it allows for design flexibility by reusing existing ontologies [45]. It follows the open world assumption knowledge representation style using OWL, RDF, and RDF schema (RDFS) syntaxes. It can be optimized with ontology patterns, and its logical and structural consistency is verified with ontology reasoners.

Overview

The proposed eHealth ontology encompasses the following steps: (1) ontology design approaches and used vocabularies; (2) ontology modeling in Protégé; (3) defining the scope; (4) integrating existing IoT and medical ontologies in the proposed ontology to annotate sensor and clinical observations; (5) ontology implementation (mapping the concepts to the proposed ontology classes and their properties in Protégé); and (6) rule expression (rule base) and basic SPARQL queries as a part of ontology verification. We further discuss how rule-based lifestyle recommendation messages (regarding activity and nutrition) could be delivered to the participants following an asserted hierarchy in the proposed eHealth ontology model, as depicted in Figure 1.

Figure 1. Asserted hierarchy for lifestyle recommendation for obesity management.



Ontology Design Approaches and Used Terminologies

Ontology design approaches can be classified into the following 5 categories: inspirational, inductive, deductive, synthetic, and collaborative [46]. We adopted a combination of inspirational and deductive approaches in our ontology design and development. The inspirational approach helped us identify the need for the ontology (what to design?) and obtain expert views to create the ontology (how to design?). The deductive approach helped us to adopt and adapt general principles to create the intended ontology tailored toward obesity as a study case. It includes the general notions being filtered and refined to be personalized to a specific domain subset (obesity). The overall approaches are divided into 5 phases as follows: in phase 1, we performed a systematic literature review to understand the need for an ontology to support the logical representation of observable and measurable data for healthy lifestyle management targeting obesity as a case study. In phase 2, we consulted experts with a research background in ICT, eHealth, nursing, and nutrition for designing the ontology. In phase 3, we developed the ontology to model and annotate health and wellness data observations with semantic metadata to create a lightweight, intelligible abstraction for health risk predictions for the personalized generation of recommendations based on rule-based decision making. In phase 4, we created rules for SPARQL queries and personalized recommendation generation (rule-based deduction). In phase 5, we verified the ontology with simulated data based on rule-based decision support.

The semantic web is W3C recommended, and it allows the specification of metadata that permit automatic reasoning [47,48]. The W3C-maintained specifications related to this study are XML, URI, RDF, turtle, RDFS, ontology web language (OWL), SPARQL Protocol and RDF Query Language (SPARQL), and SWRL. The following terminologies are relevant for our eHealth ontology representation and processing: propositional variable (an atomic name of a truth value that may change from one model to another), constant (the unique propositional variables TRUE and FALSE such that their truth value cannot be changed), and operators (the set of logical connectors in each logic). Besides, in this case, we use the operators (NOT, AND, OR, IMPLIES, and EQUIV); quantifiers (the set of logical quantifiers in a given logic; FORALL for the universal quantifier and EXISTS for the existential quantifier); quantified clause (a set of propositional variables linked together

by operators and quantifiers); clause (a quantified clause without any quantifiers); formula (a collection of clauses and quantified clauses related together by logical operators); and model of the procedure (a group of assignments for each propositional variable, such that when simplified, it leads the procedure to the constant TRUE).

Protégé, TopBraid Composer ($), NeOn Toolkit, FOAF editor, WebOnto, OntoEdit, Ontolingua Server, Ontosaurus, and WebODE are some popular ontology editors [49]. These ontology editors are open-source ontology development tools with OWL support. A reasoner is a crucial component for working with OWL ontologies. It derives new truths about the concepts that are being modeled with OWL ontology. Practically all querying of an OWL ontology (and its import closure) can be done using a reasoner [50,51]. That is why knowledge in an ontology might not be explicit, and a reasoner is required to deduce implicit knowledge so that the correct query results are obtained. The OWL API includes various interfaces for accessing OWL reasoners. For accessing reasoner via the API, a reasoner implementation is necessary. Reasoners can be classified into the following groups: OWL DL (Pellet 2.0*, HermiT, FaCT++, RacerPro), OWL EL [CEL, SHER, snorocket ($), ELLY], OWL RL [OWLIM, Jena, Oracle OWL Reasoner ($)], and OWL QL (Owlgres, QuOnto, Quill) [50-57]. In this study, we utilized Protégé ontology editor and HermiT reasoner to create and validate the structure of the ontology.

Apache Jena is a Java-based framework used for building semantic web applications. It provides an API to extract data from and write to RDF graphs. A Jena framework includes the following: (1) RDF API to parse, create, and search RDF models in XML, N-triple, N3, and Turtle formats. Triples can be stored in memory or database; (2) ARQ Engine/SPARQL API, which is a query engine for querying and updating RDF models using the SPARQL standards; (3) tuple database engine as a high-performance RDF store on a single machine; (4) ontology API for handling OWL and RDFS ontologies; and (5) Apache Jena Fuseki, which is the SPARQL server for supporting query and update. It is tightly integrated with tuple database to deliver a robust, transactional persistent storage layer. The framework has internal reasoners and an OWL API [58,59]. In this study, we used Apache Jena Fuseki for SPARQL processing with triple database.

Knowledge representation in computer-understandable form is well accepted among AI communities. Knowledge representation with symbols facilitates inferencing and the creation of new elements of knowledge. By contrast, the KB is a database for knowledge management. It provides a means for information to be collected, organized, shared, queried, and utilized for inferring new information. Knowledge engineering helps to obtain specific knowledge about some subject and represents it in a quantifiable form. KB consists of terminology models or TBox (atomic and complex) and assertions model instance or ABox (asserted and inferred). Inferred statements come as a logical outcome of the asserted statements and logical rules [35,60,61]. A KB is a pair (T, A) where T is a TBox and A is an ABox. The idea behind this paper is that the TBox concepts and relations are coming from the freshly created ontology and the ABox is a list of clauses assigning truth values to some variables. The TBox is coming from integration with the SSN Ontology and the SNOMED CT ontology plus additional concepts specific to the recommendation test case considered. The ABox is the semantic data, coming from the different data inputs (IoT sensors, interviews, and questionnaires). The satisfiability of the KB, and thus the model output, is obtained by using the hyper-tableau-based [62] reasoning solver HermiT [55]. The whole approach has been tested with 4 generated test cases to ensure that the whole mechanism can indeed set the propositional variables to true and thus send the corresponding recommendation message when needed.

Ontology Modeling

An ontology can be modeled with the following 2 ways in Protégé: frame based and OWL based. The Protégé frame editor ensures ontology development following the Open Knowledge Base Connectivity Protocol with the help of classes, properties, relationships, and instances of classes (objects). By contrast, the Protégé OWL editor (applied in this study) enables ontology development for the Semantic Web with the help of classes, properties, instances, and reasoning. We have used the steps detailed in Textbox 1 to model our proposed OWL-based eHealth ontology using the Protégé OWL editor.

Textbox 1. Steps to model the proposed OWL-based eHealth ontology.

Step 1

Create a new empty OWL project in Protégé and save it as a local file with “owl” or “ttl” extension (“ttl” signifies the turtle resource description framework [RDF] format).

Step 2

Create named classes under the “owl:Thing” super class following consistency

- Create a group of meaningful and required classes
- Define disjoint classes
- Define subclasses and disjoint subclasses

Step 3

Create OWL properties

- Object properties (associates object to object)
- Data properties (relates object to XML schema datatype or rdf:literal)
- Annotation properties (to add annotation information to classes, individuals, and properties)

Step 4

Define object properties if they are subproperties, inverse properties, functional properties, inverse functional properties, transitive properties, symmetric properties, and reflexive properties.

Step 5

Define property domain and ranges for both object and data properties (it is used as axioms in reasoning).

Step 6

Define property restrictions as follows:

- Quantifier restrictions (existential and universal)
- Cardinality restrictions (one or many)
- hasValue restrictions (eg,, string/integer/double)

Step 7

Ontology processing with a reasoner to check consistency in OWL DL, and to compute the inferred ontology class hierarchy.

- Blue color class in the inferred hierarchy signifies that the class has been reclassified.
- Red color class in the inferred hierarchy signifies an inconsistent class.

Step 8

Remove inconsistencies before importing the ontology file in Apache Jena for further processing, querying (Simple Protocol and RDF Query Language [SPARQL]), and storing it into tuple database for persistence. Tuple database supports the full range of Jena application programming interfaces. It can be used as a high-performance RDF store on a single machine.

Scope of the Proposed Ontology

We have planned to integrate the proposed eHealth ontology into a simulated eCoach system used for automatic rule-based recommendation generation to inspire individuals to manage healthy lifestyles with early health risk predictions. The planned system will have 2 main modules, as depicted in Figure 2: a data collection module and a data annotation module. The data collection module will collect identified fabricated set of habit, baseline, nutrition, personal, contextual, activity, and physiological data over time via a simulator, as depicted in Figure 3.

Figure 2. Proposed eCoach system architecture for data semantization.

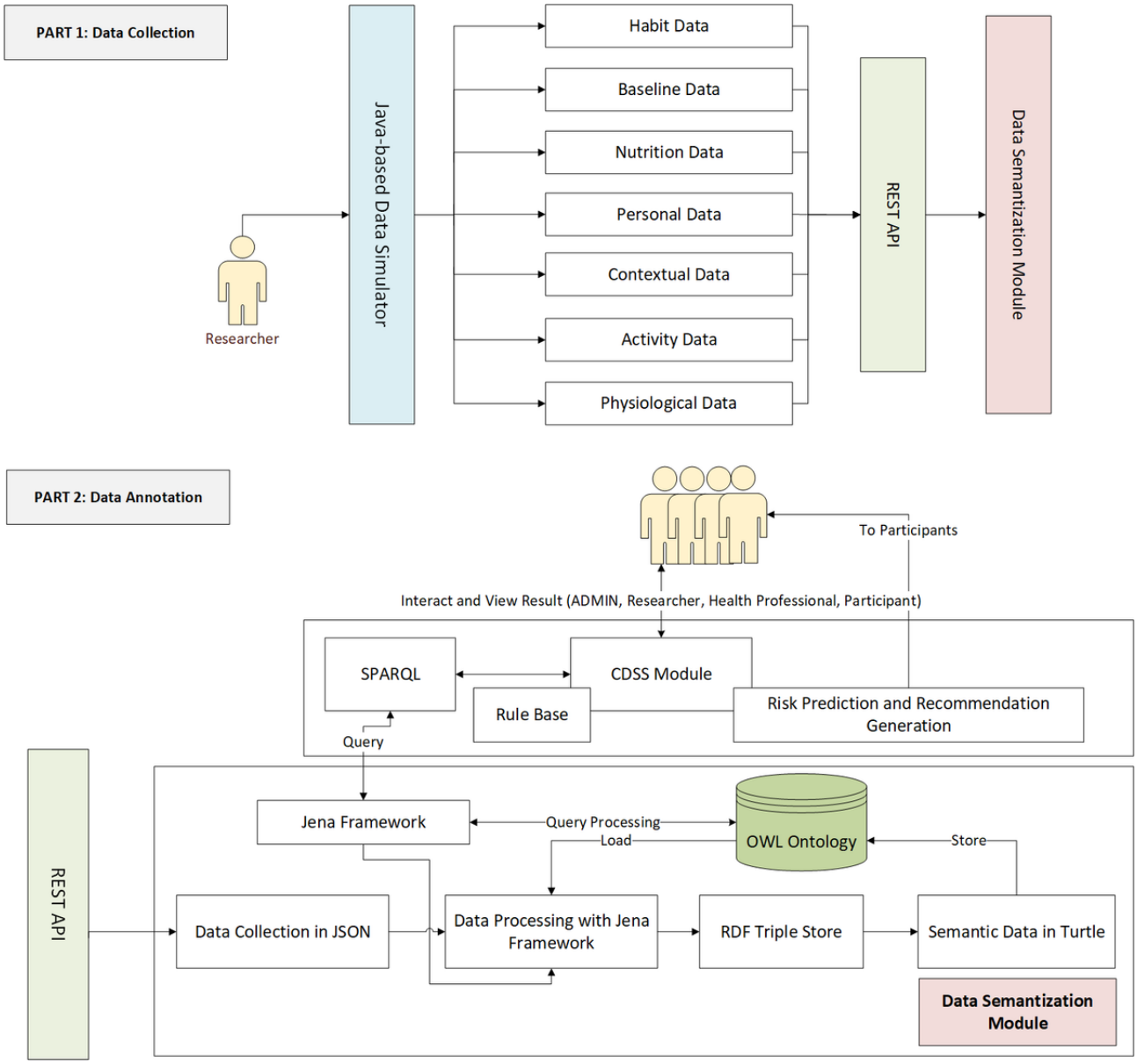
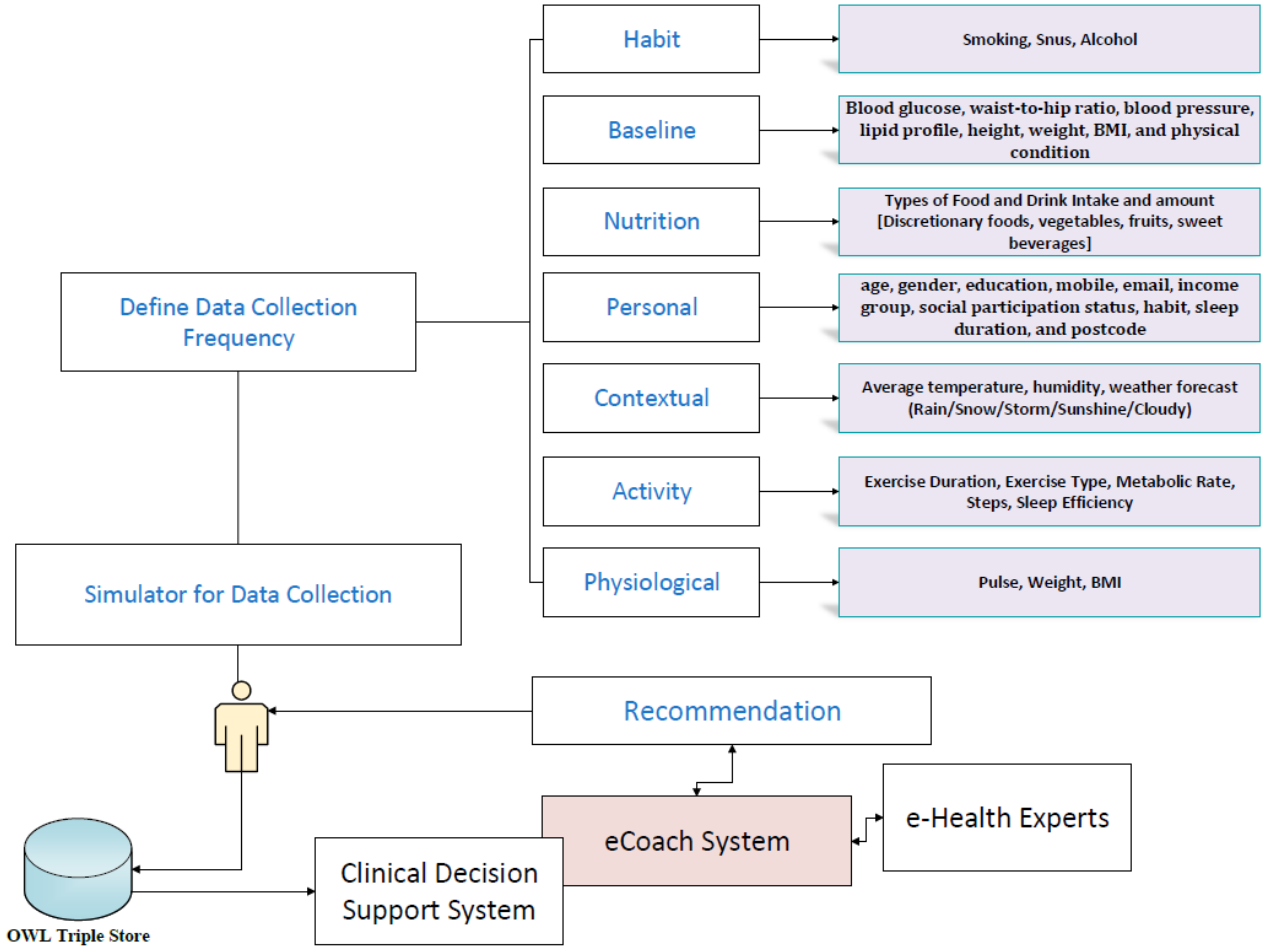
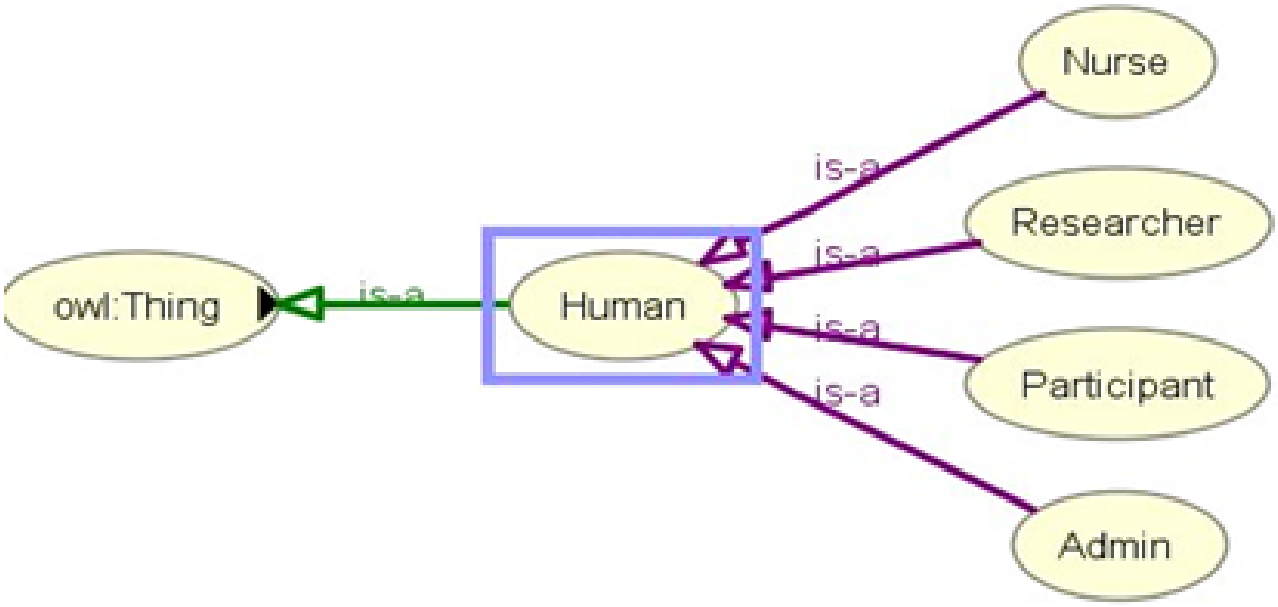


Figure 3. Types of data to be collected from participants.



The accumulated data were annotated with semantic metadata (RDF triple store graph) and stored in tuple database in turtle format. The DSS, rule base, SPARQL, risk prediction, and recommendation generation modules are not the core, and they are used for ontology verification as a test engine. The scopes of DSS are as follows: (1) periodic querying of the ontology with Jena framework using preset SPARQL queries [63-65] to assess the health condition; and (2) mapping the query result to preset clinical rules in "rule base" to generate lifestyle recommendations. This study involves 4 different user types: administrator, researcher, participants, and health professionals (eg, nurses; Figure 4). The ontology is protected from personal identity disclosure as no unique identifiers (eg, national identifiers) of participants were collected and stored in the simulated environment in accordance with the Norwegian Centre for Research Data guidelines [66]. Core eCoach and DSS concepts, AI integration for health and wellness data (activity and nutrition) analysis, real-world data collection from actual participants through web applications/mobile apps, real-life personalized recommendation generation, goal evaluation, pregnancy, genetics, child obesity, and obesity in older adults are beyond the scope of this study. This study's primary focus is to design and develop an eHealth ontology for the obesity case and to verify it with artificial data and behavioral recommendation generation with a rule-based DSS. Defined rules for test setup may vary with change in the context and is not the key focus of this paper.

Figure 4. Different types of users involved in the eCoach System.



We simulated habit, nutrition, contextual, activity, and physiological data for 4 dummy participants (2 healthy weight [N] and 2 overweight [O] participants aged between 18 and 40) for the very first day (day-n; n>0); see Multimedia Appendix 2. We assumed all the dummy participants are from the same region, so the contextual information is the same. Rule-based recommendations based on data analysis on "day-n" will be carried out by targeted participants on "day-(n+1)." Recommendations inform individual participants about their daily activity (sedentary or not), dietary intake, and activity/dietary plans. For dietary assessment, we have relied on the daily self-reported questionnaire, rather than on direct calorie calculation for basal metabolic rate. Baseline data help to compare (at the end of each month until the process ends) whether any improvement or deterioration occurred as a result of behavior change based on lifestyle recommendations. For example, reduction in BMI and BP for a person who is obese/overweight, and maintaining safe BMI and BP for a person with healthy weight upon following the behavioral recommendations is a good indication of maintaining a healthy lifestyle. We consulted with 5 experts with a research background in ICT, eHealth, nursing, and nutrition for simulating activity and nutrition data. Obesity-related information and guidelines were obtained from the World Health Organization (WHO) [67], the National Institute for Health and Care Excellence (NICE) [68], and the Norwegian Dietary Guidelines [69].

Integration With SSN Ontology and SNOMED CT

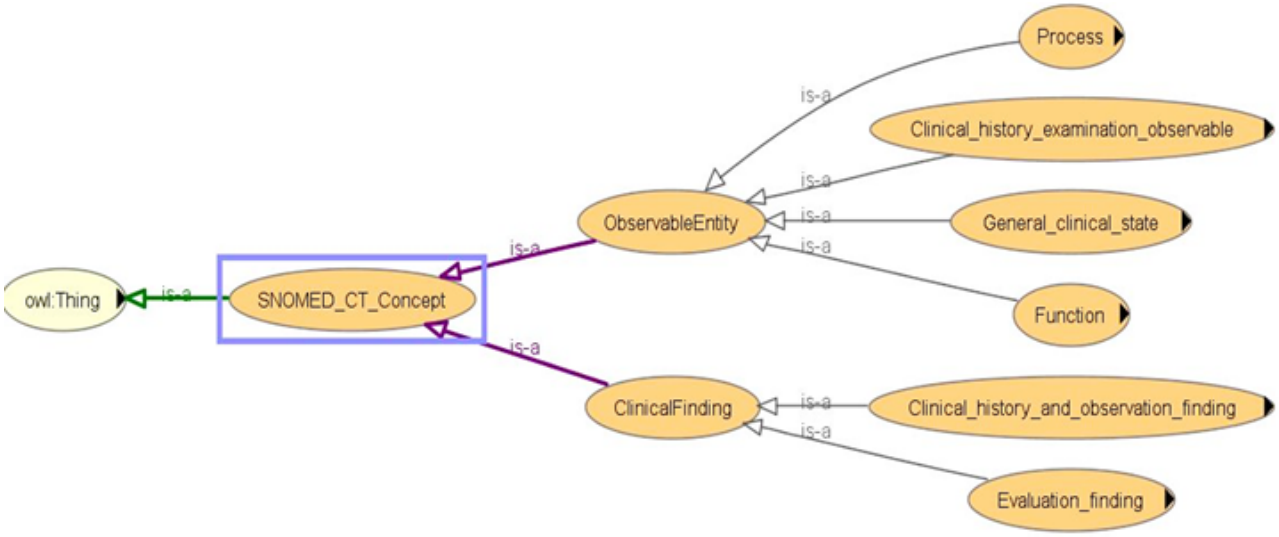
We integrated the SSN ontology [30,36,70-72] into our proposed eHealth ontology to describe sensors (activity sensors and external weather sensors), their observations, and methods adopted for sensing individual activities and context (Figure 5). Observation data related to activity and external weather are annotated with SSN ontology concepts and object properties.

Figure 5. Asserted hierarchy for sensor-based data collection with OWLViz.

Concepts and object properties in the ontology are commented and connected with "rdfs:label," "rdfs:isDefinedBy," "rdfs:seeAlso," "rdfs:comment," "dc:source," "isProxyFor," "has value," "is produced by," "has property," "hasTimeStamp," "isRegionFor," "attached system," "in deployment," "has measurement capability," "detects," "hasOutput," "observes," "implements," "has deployment," "has operating range," "has subsytem," "has survival range," "on platform," "deployment process part," "deployed on platform," "deployed system," "is property of," "feature of interest," "observation result time," "observation sampling time," "observed property," "quality of observation," "sensing method used," "includesEvent," and "observedBy." The SSN ontology is constructed on the foundation of a central ontology design pattern, so-called the stimulus–sensor–observation pattern to describe relationships between sensors, stimulus, and observations [30], and the same concept is reused in our proposed eHealth ontology model. The perspectives of SSN ontology can be classified as follows [30]: a sensor perspective, an observation perspective, a system perspective, and a feature and property perspective. Namespaces for the SSN and DUL ontologies are reused in our ontology prefixing concepts and properties as ssn: and dul:, respectively. "PhysicalDeviceThing" (a class), which behaves as a superclass of classes related to sensor-based observations, is a subclass of "owl:Thing," the universal ontology superclass.

We incorporated selected concepts from SNOMED CT [73] into our proposed ontology model to define how information about the participant's state is to be structured and processed. The SNOMED CT ontology combines hierarchical "is-a" relationships and other related relationships for vital signs, process, body measurements, and observations to describe clinical attributes as depicted in Figure 6. SNOMED CT simplifies the search for respective diseases, process, function, clinical state, measurements, and vital signs, and every concept is identified with an SCTID or SNOMED CT identifier with an object property "hasSCTID" (eg, Obese_finding hasSCTID value "414915002"^^xsd:long) [74].

Figure 6. Asserted hierarchy of SNOMED_CT concept with OWLViz.



Figures 7-9 describe the class hierarchy to process participant's clinical information using the SNOMED CT hierarchy for the vital signs (eg, BP, pulse) and body measurement (eg, obese or overweight) based on the observable entities [75-79]. Observable entities and clinical findings are linked with the objectProperty: isFoundBy. The proposed ontology model can be extended for additional clinical findings [73,74].

Figure 7. SNOMED CT class hierarchy based on selected concepts.

Figure 8. SNOMED CT ontology visualization with OntoGraf based on selected concepts.

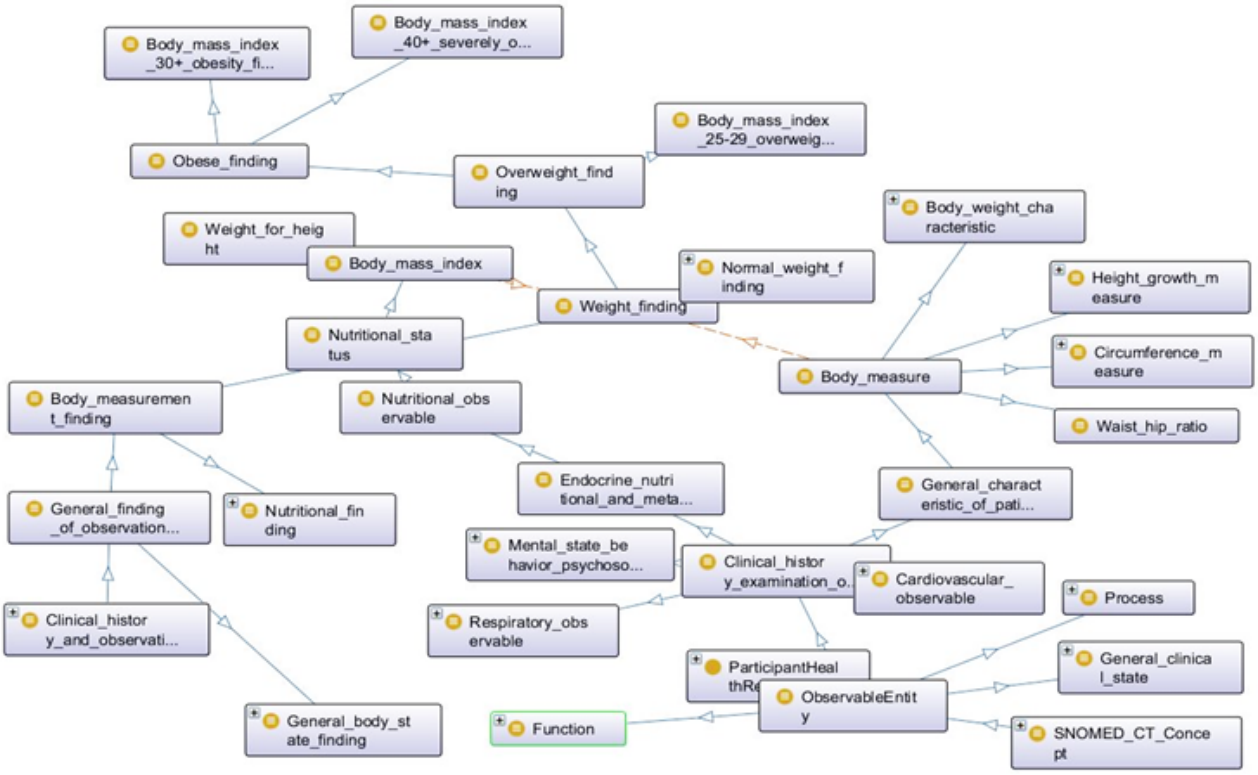
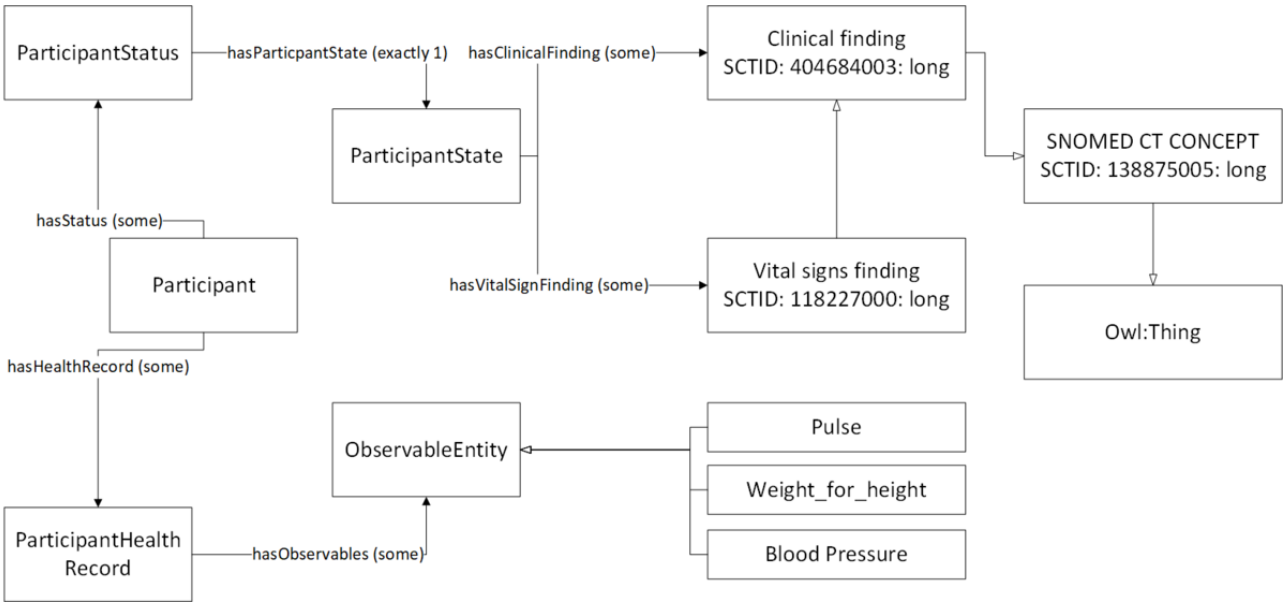


Figure 9. Selected concepts from SNOMED CT Ontology for vital signs, body measurement, and observations.



Ontology Implementation

In Figure 10 we describe how we implemented the proposed eHealth ontology for our future eCoach system with required classes, object properties, and data properties to annotate collected data. The administrator, health professionals, researchers, and participants are subclasses of the "Human" class. They have their designated role, password, and userId to authorize themselves in the system with the following associated objectProperties: hasRole, hasPassword, and hasUserId, respectively. Administrator, health professionals, and researchers have their office address (hasOfficeAddress), and personal data (hasPersonalData) to describe themselves. Their office address consists of a phone number, a postcode, and a room number with the following associated dataProperties: hasOfficePhone, hasOfficePostCode, and hasRoomNo, respectively. Their personal data include age, designation, email, first name, last name, gender, and mobile number with the corresponding dataProperties hasAge, hasDesignation, hasEmail, hasFirstName, hasLastName, hasGender, and hasMobile. The "Participant" is an important class and participants are at the core of the system. Participants have their health record, personal data obtained through interview process by trained health professionals, status (active/inactive), and recommendation with

the associated objectProperties hasHealthRecord, hasInterviewPersonalData, hasStatus, and hasReceivedRecommendation as depicted in Figure 11. “ActivityData,” “BaselineData,” “HabitData,” “NutritionData,” “PhysiologicalData” are subclasses of the “ParticipantHealthRecord” class as depicted in Figure 11. Activity data are an observable entity and are planned to be collected via activity sensors (activity bouts, steps, sleep time, activity duration, sedentary bouts, metabolic rate, nonwear time) and questionnaire (duration of intensive activity and nonwear sensor time) daily. Intensive activities are running, weightlifting, cycling, swimming, and skiing. Based on the activity type, participants can be classified into the following 4 groups: sedentary, light active, moderate active, and active. Baseline data (blood glucose, waist-to-hip ratio, BP, lipid profile, height, weight, BMI, and physical condition) are planned to be collected by trained health professionals at the time of recruitment of participants and on a monthly basis following an interview process. Habit data (smoking, snus, and alcohol consumption) and nutrition data (types of foods and drinks with amount) are planned to be collected daily with a pre-set questionnaire. Physiological data (pulse, weight, and BMI) are planned to be collected daily via activity sensors and pre-set questionnaire, as depicted in Figure 12. Personal data (age, gender, education, mobile, email, income group, social participation status, habit, sleep duration, and postcode) of healthy participants are planned to be collected following an interview process by trained health professionals during recruitment. Gender, education, income range, and social participation are essential for demographic classifications. The data properties related to data collection are depicted in Figure 13.

Figure 10. Proposed eHealth Ontology implementation in Protégé 5.x.

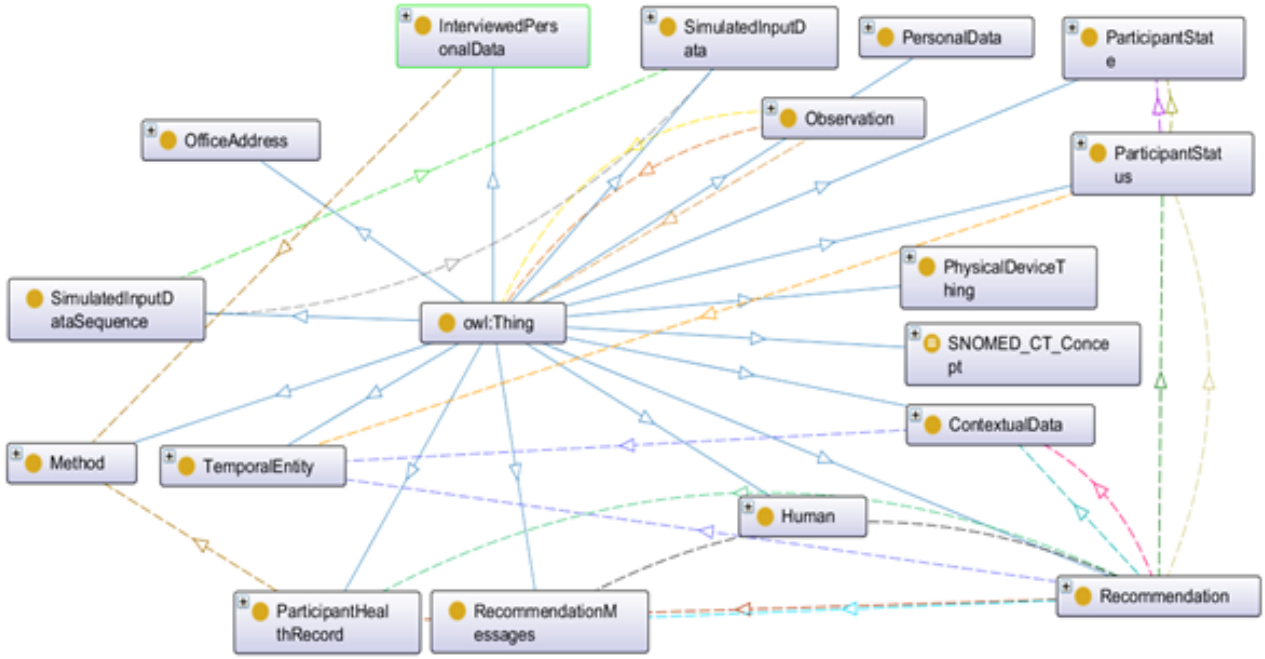


Figure 11. The class hierarchy of the proposed eHealth ontology and the description of participant class.

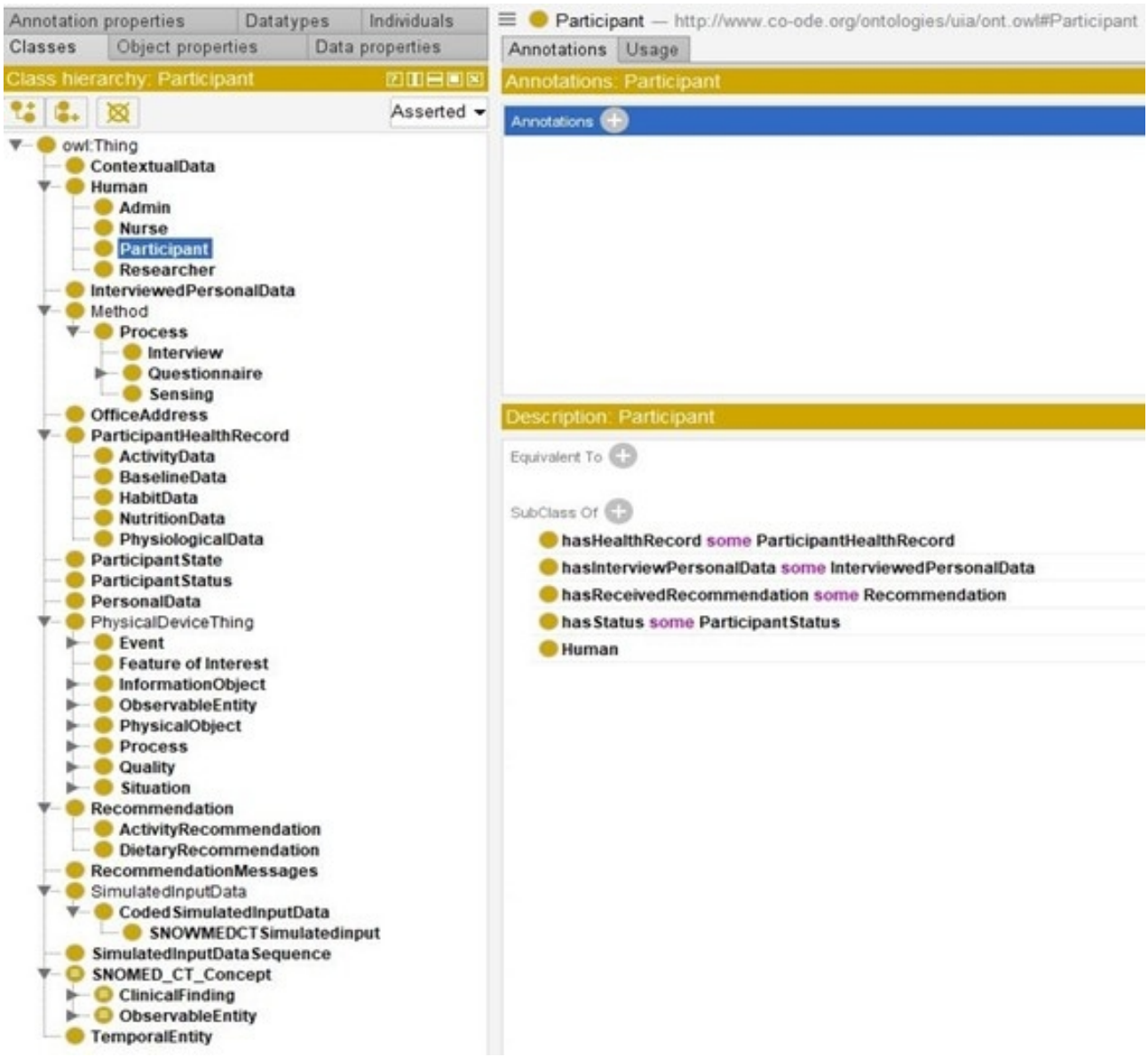


Figure 12. The asserted class hierarchy of participant's health record with OWLViz.

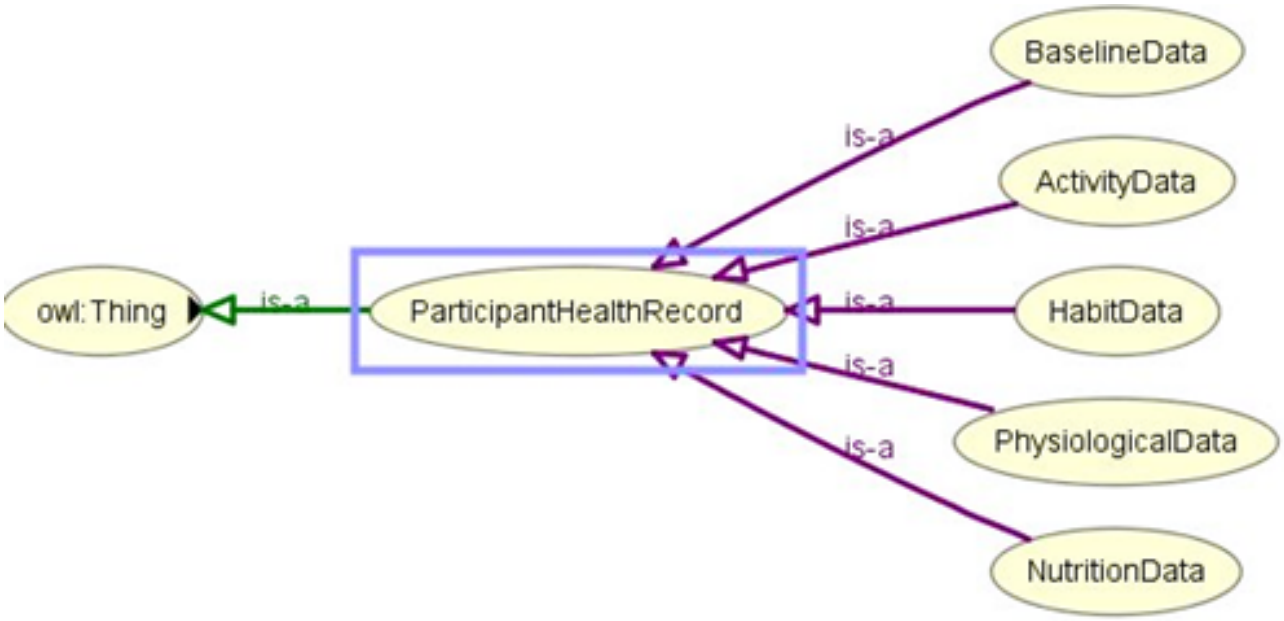
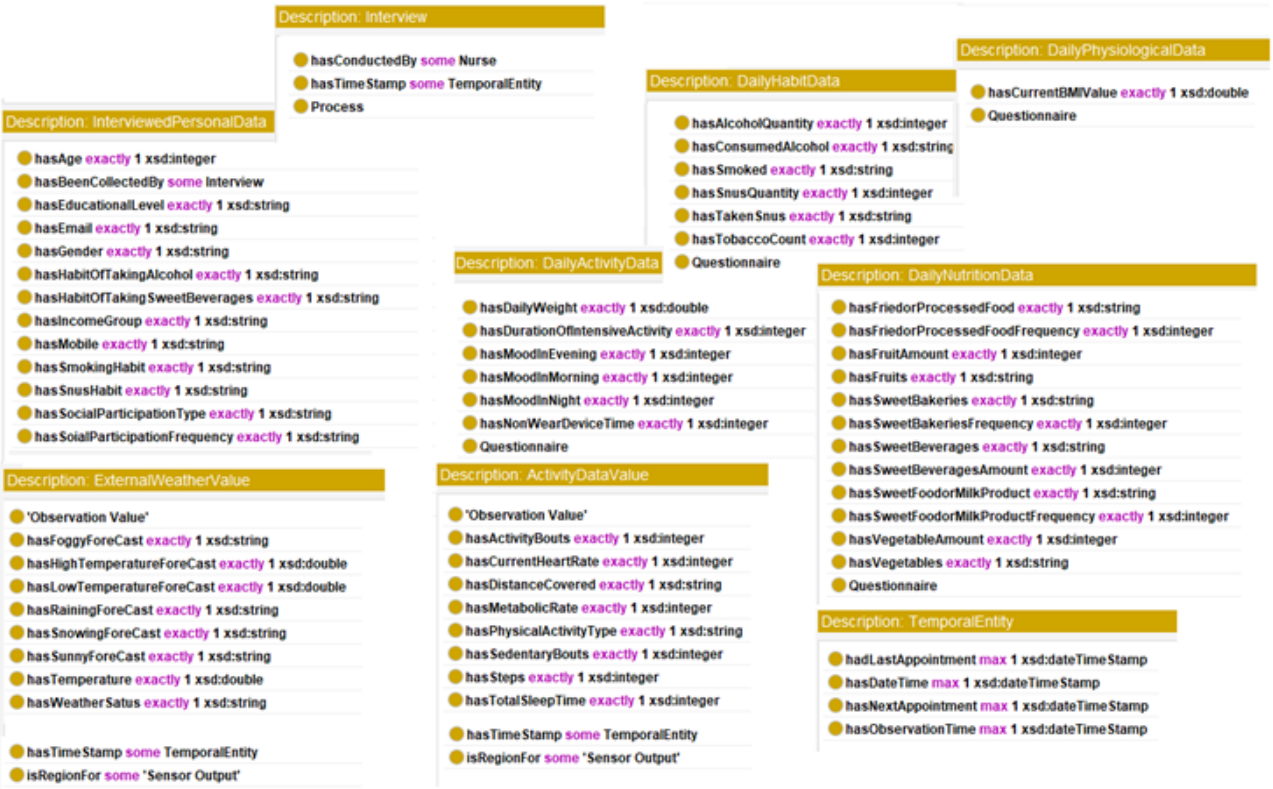


Figure 13. Data properties related to data collection.



The asserted class hierarchy of the methods used for participant's data collection is depicted in Figure 14. Each method ensures a collection of simulated data sequences, maintaining a timestamp as depicted in Figure 15. Contextual data are observable weather-related data (weather status, current temperature, rain forecast, snow forecast, storm forecast, sunny forecast, high and low temperature forecast, fog forecast), which are planned to be collected daily via sensing devices. The relationship between data and data collection methods are linked with the objectProperty: hasBeenCollectedBy and hasConductedBy (for interview).

Figure 14. The asserted class hierarchy of participant's data collection methods with OWLViz.

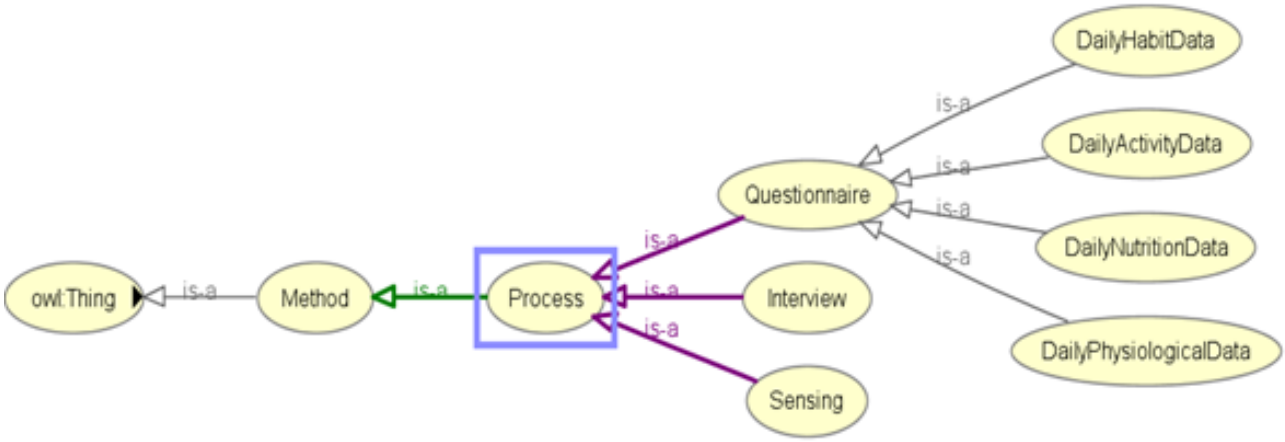
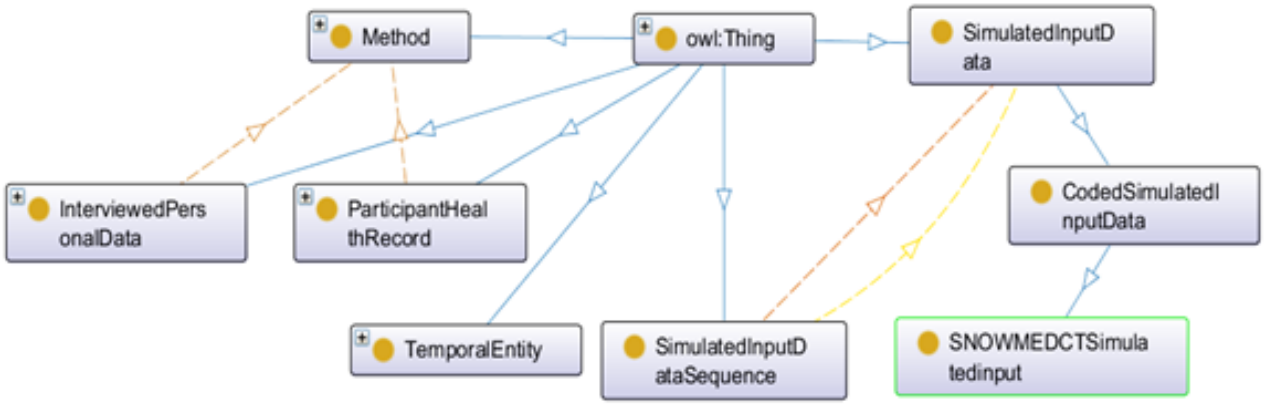
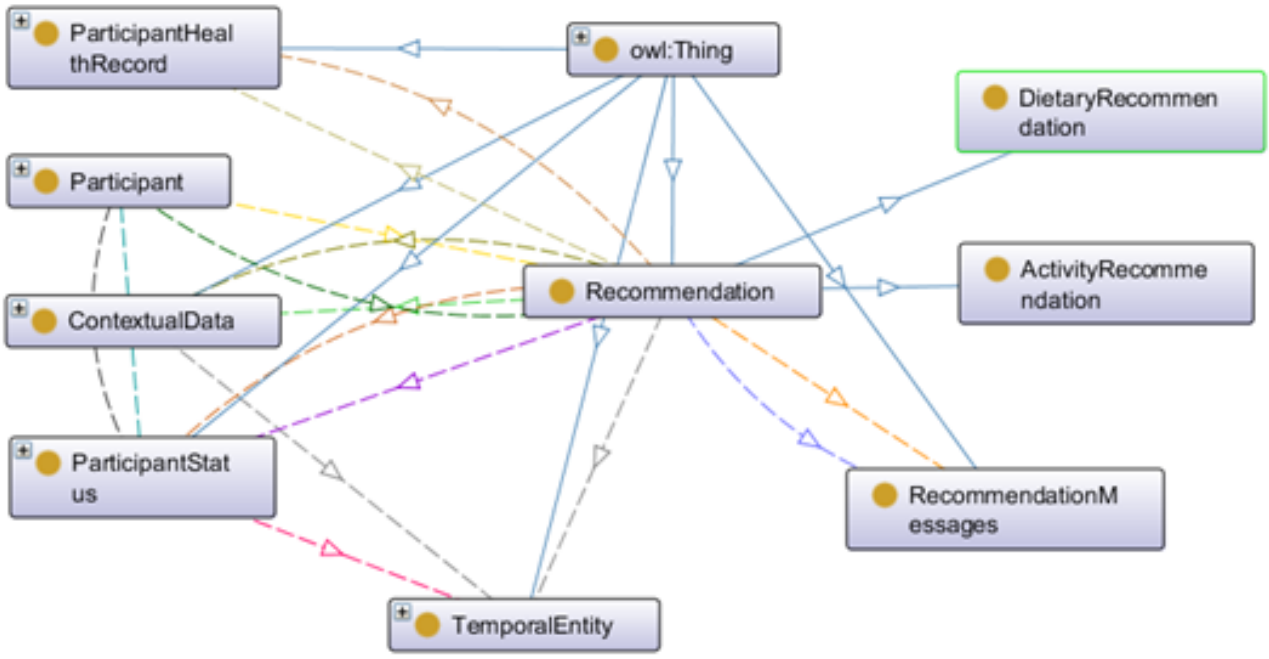


Figure 15. Ontology for data collection from simulated input.



Behavioral recommendations for a healthy lifestyle can be classified in the following 2 categories: activity (A) and dietary (D). Each recommendation is personalized and contextual. Therefore, the recommendation generation depends on evaluating participants' health status (health risk, vital signs, body measurement data) and contextual information. Each generated recommendation consists of a message and the corresponding timestamp (Figure 16). A bad habit (H) has a significant impact on healthy dietary practice. Activities are related to the context (C). Contextual data help recommend participants to plan for indoor/outdoor activities based on the following day's external weather conditions. The data properties of "RecommendationMessages" for activities are "hasActivityMessages" and "hasContextualMessages," whereas those for the diet are "hasDietaryMessages" and "hasHabitRelatedMessages." The identified set of recommendation messages for test setup (ontology verification) is presented in Multimedia Appendix 3, and is prepared based on the positive psychology [79] and the persuasion [80] concept.

Figure 16. Ontology for recommendation generation.

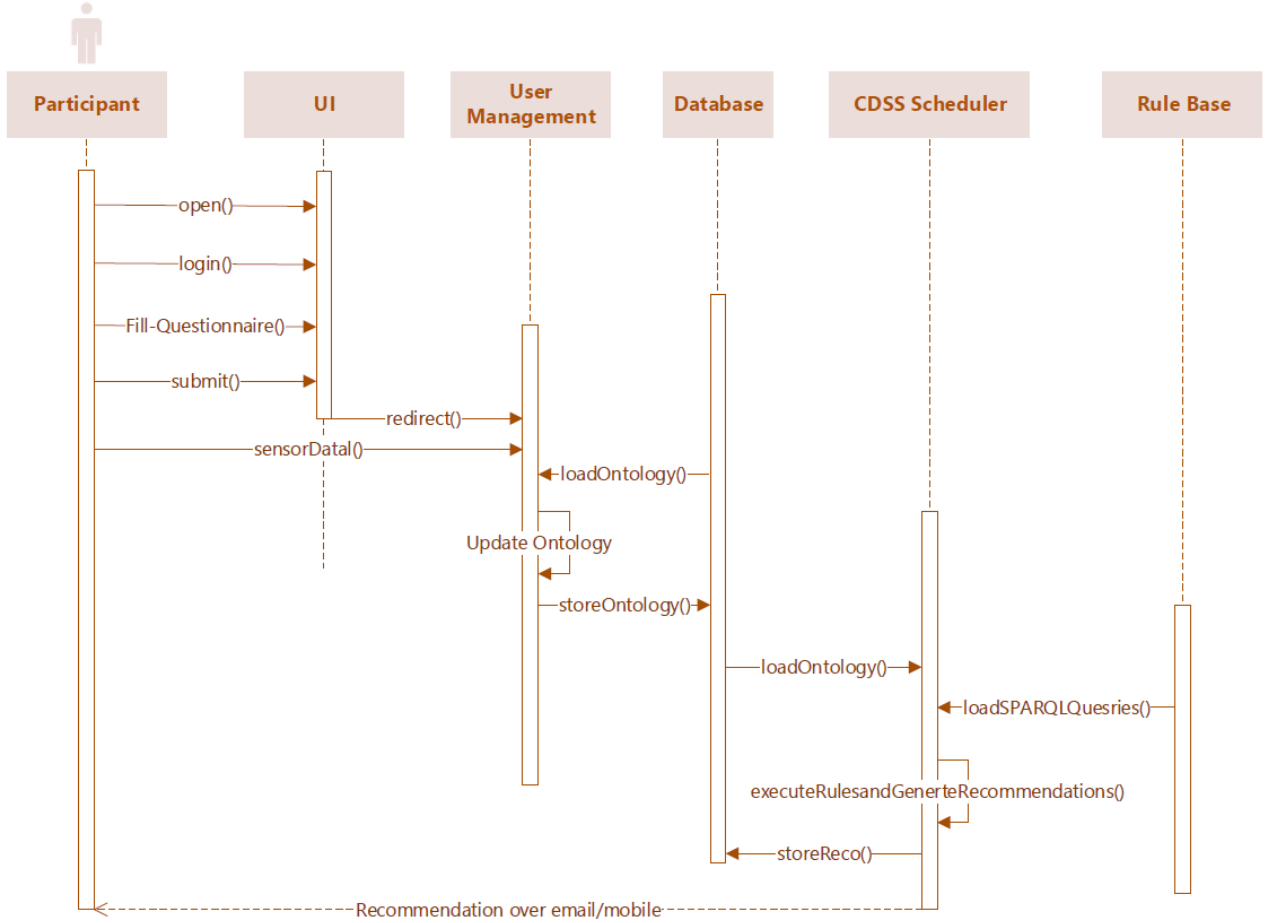


Description logic [35,81] is a formal knowledge representation of the ontology language that offers a good trade-off between expressivity, complexity, and efficiency in knowledge representation and reasoning about structured knowledge. To ensure that the paper is perfectly understood, we have the propositional variables with their linked recommendation messages. Now, we need a set of clauses such that some models will assign these variables to true and thus trigger the sending of a recommendation. The description logic SROIQ [82,83], which is logic providing a formal underpinning of OWL2, has been used as the formal logic to reason in this paper (Multimedia Appendix 4).

Rule Creation for Querying, Recommendation Generation, and Ensuring Satisfiability

A rule consists of a premise (antecedent) and a conclusion (premise). For every condition mentioned in Multimedia Appendix 3, DSS executes SPARQL queries daily to determine what type of recommendation message is to be delivered to each participant as depicted in the unified modeling language sequence diagram (Figure 17). The execution of every predefined semantic rule as specified in Multimedia Appendix 4 relies on the SPARQL query execution, and the rules are created following clinical guidelines, as stated in Multimedia Appendix 5 [62,84-92]. In this study, 20 semantic rules are subdivided into activity-level classification (8), habit-related classification (3), dietary classification (4), weather-level classification (1), obesity-level classification (3), and satisfiability (1) (please also see Multimedia Appendix 4). Moreover, except for the already-existing ontologies used, to ensure some consistency regarding what a participant is, what are the participant health records, etc., the concepts and the rules added are relatively easy to follow, and therefore they will be relatively easy to use.

Figure 17. UML sequence diagram for recommendation generation and delivery.



The observable and measurable parameters associated with activities, habit, nutrition, and context (as described in Multimedia Appendix 4) for individual participants on a timestamp are obtained based on the execution of SPARQL queries by DSS on a daily scheduled interval as specified in Multimedia Appendix 6. The rules 17-19 in Multimedia Appendix 4 assign truth values to variables that ensure consistency with concepts already existing in the SNOMED CT ontology, where the body measurement is defined. We have confirmed with HermiT that for 4 specific cases the correct recommendation messages are triggered. However, one would need to ensure that there is not a combination of variables such that the whole formula is unsatisfiable (ie, no model can satisfy the procedure). One would also need to ensure that only 1 message can be triggered at a time. In this study, we have a formal guarantee that 2 “once-a-day” messages can neither be triggered simultaneously nor for every possible combination of variables, there is, every time, a model output by HermiT. If we put the different variables used in the first 19 rules (Multimedia Appendix 4) into propositional variables, we would have an exponential number of “possible participants.” One formal way to ensure a model’s existence is to negate all our rules and ensure the same. Then, the formula is indeed unsatisfiable. As 2 messages cannot be triggered at the same time, and to satisfy the same, we added a rule (rule 20) on the variables used in the recommendations started “once-a-day.” If rule 20 is false, then the whole set of rules (considered as a large conjunction) will be set to false. It will result in “no execution” of the proposition (see Multimedia Appendix 3) and will help us to debug our defined semantic rules (rules 1-19) as defined in Multimedia Appendix 4. If it is set to true, we have a formal guarantee that no 2 “once-a-day” messages can be triggered at the same time, no matter the truth values we put into our ABox.

Results

The test setup to verify the proposed eHealth ontology’s performance and reliability consisted of a DSS module (health risk prediction and recommendation generation for a healthy lifestyle), SPARQL, and rule base. As an outcome of ontology verification, we generated personalized and contextual recommendations (behavioral) following semantic rules to balance individual weight change with adopting healthy behavior to balance a trade-off between physical activity, healthy habit, and a healthy diet as depicted in Figure 18. We executed all the semantic rules as stated in Multimedia Appendix 4 in the form of SPARQL queries using the Jena ARQ engine on each participant’s simulated data as mentioned in Multimedia Appendix 2. We then determined what type of recommendation messages would be required to be delivered for each participant to manage his/her healthy lifestyle. These findings are detailed in Table 1.

Figure 18. Behavioral recommendation generation (pivot) for the management of healthy lifestyle (a trade-off between physical activity, healthy habit, and healthy diet).

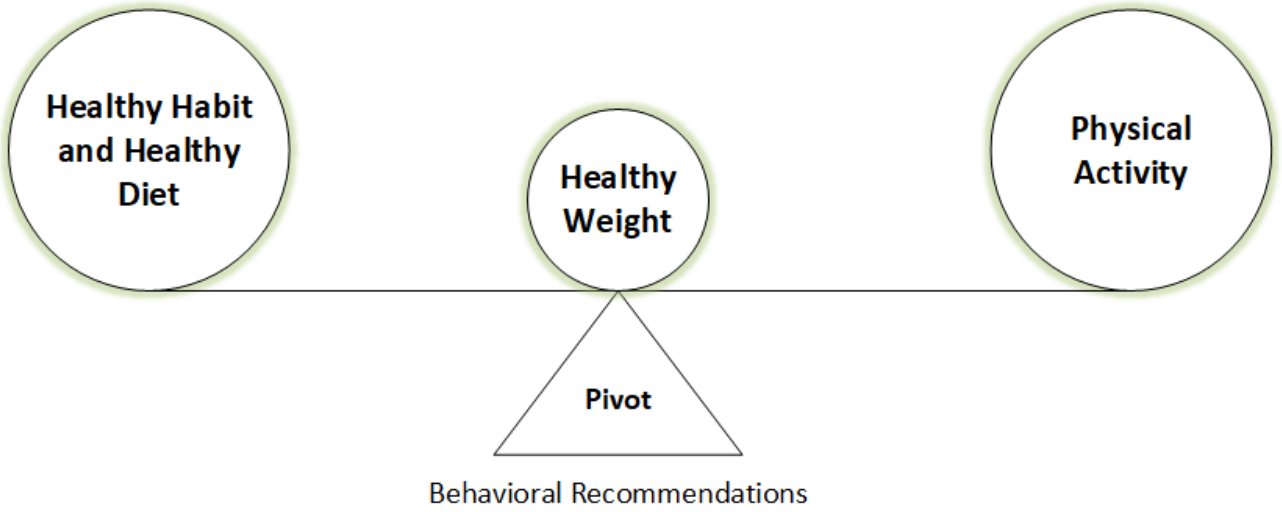


Table 1. Recommendation generation for participants for Day-(n+1) [n>0].

Participant	Profile	SCTID	Healthy habit on Day-n	Healthy diet on Day-n	Physically active on Day-n	Recommendation(s) for Day-(n+1)
Individual_1	Normal weight	43664005	No	No	Yes	H-1, D-2, D-3, A-4, C-1
Individual_2	Normal weight	43664005	No	Yes	No	H-1, H-2, D-4, A-3, C-1
Individual_3	Overweight	162863004	No	No	No	H-2, D-1, D-2, D-3, A-2, A-5, C-1
Individual_4	Overweight	162863004	Yes	No	No	D-1, D-2, A-1, C-1

Discussion

Principal Findings

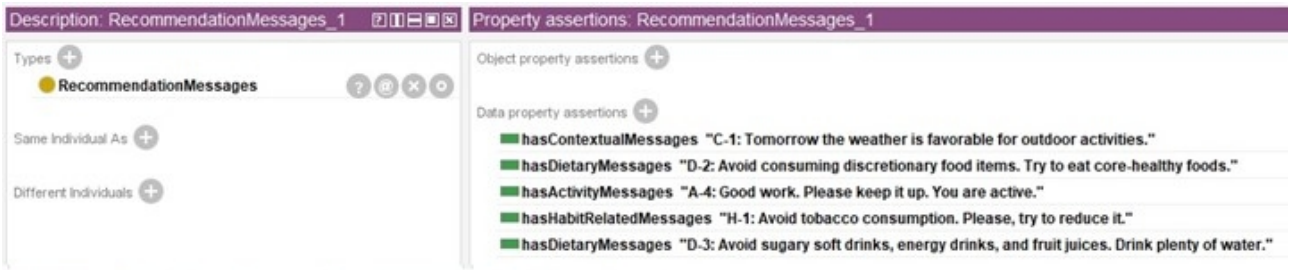
According to Table 1, “Individual_1” and” Individual_2” are healthy weight participants, and “Individual_3” and “Individual_4” are overweight participants as assessed based on their daily (“Day-n”) BMI (weight/height^2) value. According to Figure 1, a healthy weight is a trade-off between healthy habits, healthy diet, and physical activity. On “Day-n” (n>0), “Individual_1” has been physically active, and this is the reason he has been encouraged to keep up the same activity level (A-4). By contrast, he has shown some addiction toward “snus,” sweet beverages, and fried/processed foods, which might grow negative behavior in the participant and increase his weight. Therefore, he has been recommended to reduce tobacco consumptions (H-1) and to refrain from discretionary food items (D-2 and D-3). The simulated data for “Individual_2” has demonstrated that she is inclined to a healthy diet (D-4), but growing some negative behavior with consumption of alcohol and tobacco (H-1, H-2). She is just one step behind to become physically active (A-3). Hence, she has been recommended to take a healthy dietary plan, refrain from tobacco and alcohol, and increase activity level to become active. “Individual_3” is neither physically active nor adhered to healthy habits or healthy dietary plans. He is addicted to alcohol, fried/processed foods, sweet beverages, sweet food/milk products. His consumed number of vegetables and fruits is not adequate for a healthy diet (<400 g). Therefore, he has been recommended to reduce alcohol consumption (H-1), to follow a healthy dietary habit (D-1, D-2, D-3), and to become more physically active (A-2) with adequate sleeping (A-5). The fabricated data for “Individual_4” has shown that she has an unhealthy diet plan, and she is mostly leading a sedentary lifestyle. Therefore, she has been recommended to stay away from discretionary food items (D-2), to incline on “core-foods” (D-1), and to increase activity level by one step (A-1). The analysis of contextual data reveals that the weather on “Day-(n+1)” is suitable for outdoor activities. The purpose of the individualized recommendation generation is to guide and encourage individual participants to keep up a healthy lifestyle by maintaining a balance between healthy habit, healthy diet, and physical activity. It encourages people with a normal weight to maintain their healthy weight, and those with obesity/overweight to reduce their weight.

The rule-based decision support has generated personalized and contextual recommendations (Table 1) using SPARQL queries, as depicted in Figure 19, based on the proposed ontology without any “false-positive” case. The proposed ontology’s reasoning time has been measured as <30.0 seconds in Protégé with HermiT reasoner without reporting any inconsistencies. The reading time of the ontology after loading it in the Jena workspace was about 2.0-3.5 seconds with the “OWL_MEM_MICRO_RULE_INF” ontology specification in the “TTL” format (OWL full), “in-memory” storage, and “optimized rule-based reasoner with OWL rules.” Then, we queried ontology classes, ontologies, “predicate, subject, and object” of every statement using Jena in <1.5 seconds, <0.5 seconds, and <3.5 seconds, respectively. Each ontology model (complete RDF graph) is related to a document manager (default

global document manager: "OntDocumentManager") to assist with the processing and handling of ontology documents. All the classes in the ontology API that represent ontology values have "OntResource" as a common super-class with attributes (versionInfo, comment, label, seeAlso, isDefinedBy, sameAs, and differentFrom) and methods (add, set, list, get, has, and remove). We used the implementation of the RDF interface, provided by Jena, to store the modeled ontology and its instances persistently in the tuple database and load it back to process further. Jena Fuseki is tightly integrated with tuple database to provide a robust, transactional persistent storage layer (Figure 20).

In the future study, the recommendation process can be automated with the amalgamation of a hybrid DSS system (rule based and data driven) and AI algorithms. The scope of the proposed ontology can be enhanced with the integration of (1) real sensor activity devices; (2) mood assessment of participants; (3) collection of nutrition data on a detailed level through multiple questionnaires (daily, on every alternative day, and weekly); (4) semantic annotation of the recommended messages; (5) weekly suggestion generation after evaluating daily generated recommendations, and followed by a ranking of participants based on their weekly performances; (6) help-desk management for technical support; (7) assessment of baseline data; (8) trend analysis of health risks as a function of habit, diet, and activity with machine intelligence; and (9) automated interview management by trained health professionals (nurses).

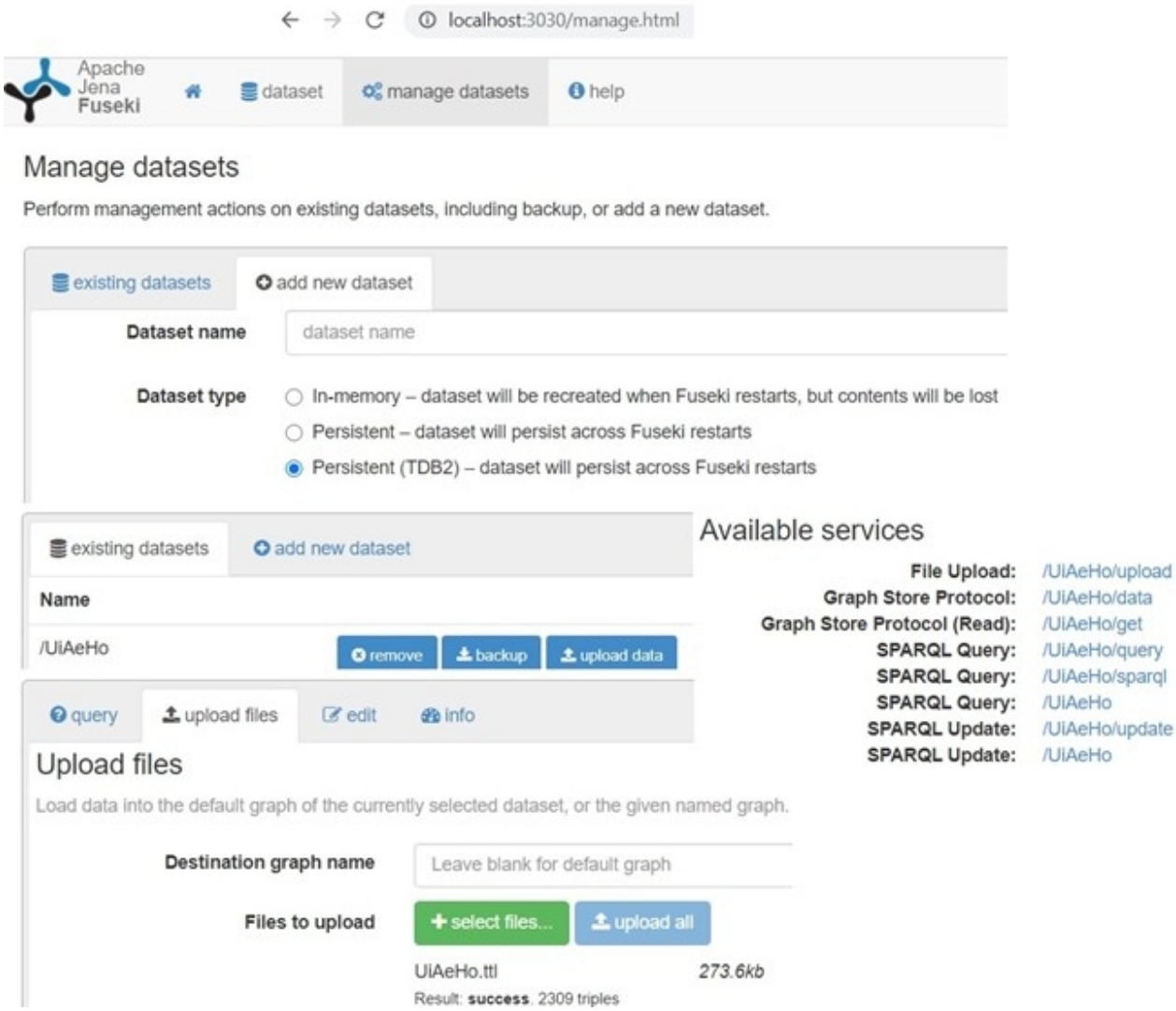
Figure 19. Sample SPARQL query for recommendation finding (e.g., "Individual_1").




```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX ssn: <http://purl.oclc.org/NET/ssnx/ssn#>
PREFIX : <http://www.co-ode.org/ontologies/uia/ont.owl#>

SELECT  ?participant ?activitymessages ?habitmessages ?dietarymessages ?contextualmessages ?timedateinfo
WHERE {
  ?participant :hasReceivedRecommendation ?recommendation .
  ?recommendation :hasMessages ?recommendationmessages .
  ?timestamp :hasDateTime ?timedateinfo .
  ?recommendationmessages ssn:hasActivityMessages ?activitymessages .
  ?recommendationmessages ssn:hasHabitRelatedMessages ?habitmessages .
  ?recommendationmessages ssn:hasDietaryMessages ?dietarymessages .
  ?recommendationmessages ssn:hasContextualMessages ?contextualmessages .
}
LIMIT 25
```

Figure 20. Integration of TDB with Jena Fuseki for ontology store in the "ttl" format and querying.



Conclusions

In health care, with the research advancement on the IoT domain, an increasing number of sensors, actuators, mobile, and web-based health monitoring devices are deployed into our daily life for remote health monitoring. It produces enormous personalized health and wellness observable and measurable data with hidden patterns. Data collected by multichannel sensors or devices demonstrate significant differences in data formats, types, and domains, which might lead to a problem in machine understandability. Therefore, a semantic representation of collected health and wellness data from heterogeneous sources is necessary, and the ontology serves the purpose. In this pilot study, we have proposed an eHealth ontology model in association with SSN and SNOMED CT, to support a semantic representation of collected observable and measurable data to manage a healthy lifestyle focusing on obesity as a case study. The ontology represents collected data with OWL-based web language in RDF triple-store format. The performance of the proposed ontology has been evaluated with the simulated data (eg, sensor, interview, and questionnaire) of 4 dummy participants. The proposed ontology's structural and logical consistency has been evaluated with a Protégé reasoner (HermiT 1.4.3.x). The proposed ontology model has been used by a rule-based DSS to generate personalized and contextual recommendations with the execution of SPARQL queries against a preset rule base (with the help of Apache Jena library) to promote a healthy lifestyle for obesity management. In the future study, we will recruit real participants following inclusion and exclusion criteria and provide them real activity devices to replicate the whole scenario and evaluate the efficacy of the recommendation generation plan. The proposed ontology can be extended to annotate observable and measurable data for other related lifestyle diseases, such as diabetes type II, chronic obstructive pulmonary diseases, cardiovascular diseases, and mental health.

Acknowledgments

The authors acknowledge the funding and infrastructure from the University of Agder, Center for e-Health, Norway, to carry out this research.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Proposed ontology model's OWL file with annotated participant data.
[DOCX File , 80 KB - jmir_v23i4e24656_app1.docx]

Multimedia Appendix 2
Simulated data for 4 participants.
[XLSX File (Microsoft Excel File), 14 KB - jmir_v23i4e24656_app2.xlsx]

Multimedia Appendix 3
Propositional variables with their linked recommendation messages.
[XLSX File (Microsoft Excel File), 10 KB - jmir_v23i4e24656_app3.xlsx]

Multimedia Appendix 4
Scoped recommendation conditions, and corresponding rules (rule-base) for test set-up.
[XLSX File (Microsoft Excel File), 10 KB - jmir_v23i4e24656_app4.xlsx]

Multimedia Appendix 5
Health parameters and corresponding clinical rules [70-78].
[XLSX File (Microsoft Excel File), 10 KB - jmir_v23i4e24656_app5.xlsx]

Multimedia Appendix 6
Prefixes and queries.
[XLSX File (Microsoft Excel File), 13 KB - jmir_v23i4e24656_app6.xlsx]

References

1. Chatterjee A, Gerdes MW, Martinez SG. Identification of Risk Factors Associated with Obesity and Overweight—A Machine Learning Overview. Sensors 2020 May 11;20(9):2734. [doi: 10.3390/s20092734]
2. Chatterjee A, Gerdes M, Martinez S. eHealth Initiatives for The Promotion of Healthy Lifestyle and Allied Implementation Difficulties. In: WiMob. 2019 Presented at: International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob); 2019; Barcelona, Spain p. 1-8. [doi: 10.1109/wimob.2019.8923324]
3. Wagner K, Brath H. A global view on the development of non communicable diseases. Prev Med 2012 May;54 Suppl:S38-S41. [doi: 10.1016/j.ypmed.2011.11.012] [Medline: 22178469]
4. Gerdes M, Martinez S, Tjondronegoro D. Conceptualization of a personalized ecoach for wellness promotion. 2017 Presented at: Proceedings of the 11th EAI International Conference on Pervasive Computing Technologies for Healthcare; 2017; Barcelona, Spain p. 365-374. [doi: 10.1145/3154862.3154930]
5. Rutjes H, Willemsen MC, IJsselsteijn WA. Understanding effective coaching on healthy lifestyle by combining theory- and data-driven approaches. 2016 Presented at: 11th International Conference on Persuasive Technology (PERSUASIVE); 2016; Salzburg, Austria p. 26-29. [doi: 10.1007/978-3-319-31510-2]
6. Dijkhuis TB, Blaauw FJ, van Ittersum MW, Velthuijsen H, Aiello M. Personalized Physical Activity Coaching: A Machine Learning Approach. Sensors (Basel) 2018 Feb 19;18(2):623 [FREE Full text] [doi: 10.3390/s18020623] [Medline: 29463052]
7. Jin W, Kim DH. Design and Implementation of e-Health System Based on Semantic Sensor Network Using IETF YANG. Sensors (Basel) 2018 Feb 20;18(2):629 [FREE Full text] [doi: 10.3390/s18020629] [Medline: 29461493]
8. Alamri A. Ontology Middleware for Integration of IoT Healthcare Information Systems in EHR Systems. Computers 2018 Oct 08;7(4):51. [doi: 10.3390/computers7040051]
9. Rodgers MM, Pai VM, Conroy RS. Recent Advances in Wearable Sensors for Health Monitoring. IEEE Sensors J 2015 Jun;15(6):3119-3126. [doi: 10.1109/jsen.2014.2357257]
10. Mackensen E, Lai M, Wendt T. Bluetooth Low Energy (BLE) based wireless sensors. 2012 IEEE Sensors 2012:1-4. [doi: 10.1109/icsens.2012.6411303]

11. Nykänen P. Decision support systems from a health informatics perspective. Tampere, Finland: Tampere University Press; 2000.
12. Allemang D, Hendler J. Semantic Web for the Working Ontologist: Effective Modeling in RDFS and OWL. Waltham, MA: Elsevier; Jul 05, 2011.
13. Gyrard A, Datta S, Bonnet C, Boudaoud K. Cross-Domain Internet of Things Application Development: M3 Framework and Evaluation. 2015 Presented at: 3rd International Conference on Future Internet of Things and Cloud; 2015; Rome, Italy. [doi: 10.1109/ficloud.2015.10]
14. Kolozali S, Elsaleh T, Barnaghi PM. A Validation Tool for the W3C SSN Ontology based Sensory Semantic Knowledge. In: The Semantic Web – ISWC 2014. 2014 Presented at: 13th International Semantic Web Conference; October 19-23, 2014; Riva del Garda, Italy p. 83-88.
15. Julina J, Thenmozhi D. Ontology based EMR for decision making in health care using SNOMED CT. 2012 Presented at: International Conference on Recent Trends in Information Technology; April 19-21, 2012; Chennai, Tamil Nadu, India p. 514-519. [doi: 10.1109/icrtit.2012.6206787]
16. Kim H, Park H, Min YH, Jeon E. Development of an obesity management ontology based on the nursing process for the mobile-device domain. J Med Internet Res 2013;15(6):e130 [FREE Full text] [doi: 10.2196/jmir.2512] [Medline: 23811542]
17. Sojic A, Terkaj W, Contini G, Sacco M. Modularising ontology and designing inference patterns to personalise health condition assessment: the case of obesity. J Biomed Semantics 2016 May 04;7(1):12 [FREE Full text] [doi: 10.1186/s13326-016-0049-1] [Medline: 29764473]
18. Kim H, Mentzer J, Taira R. Developing a Physical Activity Ontology to Support the Interoperability of Physical Activity Data. J Med Internet Res 2019 Apr 23;21(4):e12776 [FREE Full text] [doi: 10.2196/12776] [Medline: 31012864]
19. Lasierra N, Alesanco A, O'Sullivan D, García J. An autonomic ontology-based approach to manage information in home-based scenarios: From theory to practice. Data & Knowledge Engineering 2013 Sep;87:185-205. [doi: 10.1016/j.datak.2013.06.004]
20. Lasierra N, Alesanco A, Guillén S, García J. A three stage ontology-driven solution to provide personalized care to chronic patients at home. J Biomed Inform 2013 Jun;46(3):516-529 [FREE Full text] [doi: 10.1016/j.jbi.2013.03.006] [Medline: 23567539]
21. Yao W, Kumar A. CONFlexFlow: Integrating Flexible clinical pathways into clinical decision support systems using context and rules. Decision Support Systems 2013 May;55(2):499-515. [doi: 10.1016/j.dss.2012.10.008]
22. Chi Y, Chen T, Tsai W. A chronic disease dietary consultation system using OWL-based ontologies and semantic rules. J Biomed Inform 2015 Feb;53:208-219 [FREE Full text] [doi: 10.1016/j.jbi.2014.11.001] [Medline: 25451101]
23. Rhayem A, Ahmed Mhiri MB, Salah MB, Gargouri F. Ontology-based system for patient monitoring with connected objects. Procedia Computer Science 2017;112:683-692. [doi: 10.1016/j.procs.2017.08.127]
24. Galopin A, Bouaud J, Pereira S, Seroussi B. An Ontology-Based Clinical Decision Support System for the Management of Patients with Multiple Chronic Disorders. Stud Health Technol Inform 2015;216:275-279. [Medline: 26262054]
25. Sherimon PC, Krishnan R. OntoDiabetic: An Ontology-Based Clinical Decision Support System for Diabetic Patients. Arab J Sci Eng 2015 Dec 16;41(3):1145-1160. [doi: 10.1007/s13369-015-1959-4]
26. Hristoskova A, Sakkalis V, Zacharioudakis G, Tsiknakis M, De TF. Ontology-driven monitoring of patient's vital signs enabling personalized medical detection and alert. Sensors (Basel) 2013;14(1):1598-1628 [FREE Full text] [doi: 10.3390/s140101598] [Medline: 24445411]
27. Riaño D, Real F, López-Vallverdú JA, Campana F, Ercolani S, Mecocci P, et al. An ontology-based personalization of health-care knowledge to support clinical decisions for chronically ill patients. J Biomed Inform 2012 Jun;45(3):429-446 [FREE Full text] [doi: 10.1016/j.jbi.2011.12.008] [Medline: 22269224]
28. Ganguly P, Chattopadhyay S, Paramesh N, Ray P. An ontology-based framework for managing semantic interoperability issues in e-health. 2008 Presented at: HealthCom 10th International Conference on e-health Networking, Applications and Services; July 7-9, 2008; Singapore p. 73-78. [doi: 10.1109/health.2008.4600114]
29. Elhefny M, Elmogy M, Elfetouh A. Building OWL ontology for obesity related cancer. 2014 Presented at: 9th International Conference on Computer Engineering & Systems (ICCES); December 22-23, 2014; Cairo, Egypt p. 177-183. [doi: 10.1109/icces.2014.7030953]
30. Compton M, Barnaghi P, Bermudez L, García-Castro R, Corcho O, Cox S, et al. The SSN ontology of the W3C semantic sensor network incubator group. Journal of Web Semantics 2012 Dec;17:25-32. [doi: 10.1016/j.websem.2012.05.003]
31. Henson C, Sheth A, Thirunarayan K. Semantic Perception: Converting Sensory Observations to Abstractions. IEEE Internet Computing 2012 Mar;16(2):26-34. [doi: 10.1109/mic.2012.20]
32. Barnaghi P, Meissner S, Presser M, Moessner K. Sense and sens' ability: Semantic data modelling for sensor networks. 2009 Presented at: Conference Proceedings of ICT Mobile Summit; June 2009; Santander, Spain.
33. Haller A, Janowicz K, Cox SJ, Lefrançois M, Taylor K, Le Phuoc D, et al. The modular SSN ontology: A joint W3C and OGC standard specifying the semantics of sensors, observations, sampling, and actuation. Semantic Web 10.1 2018 Dec 28;10(1):9-32. [doi: 10.3233/sw-180320]
34. Shi F, Li Q, Zhu T, Ning H. A Survey of Data Semantization in Internet of Things. Sensors (Basel) 2018 Jan 22;18(1):313 [FREE Full text] [doi: 10.3390/s18010313] [Medline: 29361772]

35. Baader F, Calvanese D, McGuinness D, Patel-Schneider P, Nardi D. The description logic handbook: Theory, implementation and applications. Cambridge, England: Cambridge University Press; 2003.
36. Raggett D. The Web of Things: Challenges and Opportunities. Computer 2015 May;48(5):26-32. [doi: 10.1109/mc.2015.149]
37. Sezer O, Can S, Dogdu E. Development of a smart home ontology and the implementation of a semantic sensor network simulator: An Internet of Things approach. 2015 Presented at: International Conference on Collaboration Technologies and Systems (CTS); June 1-5, 2015; Atlanta, GA p. 12-18. [doi: 10.1109/cts.2015.7210389]
38. Yu AC. Methods in biomedical ontology. Journal of Biomedical Informatics 2006 Jun;39(3):252-266. [doi: 10.1016/j.jbi.2005.11.006]
39. Herre HB, Burek P, Hoehndorf R, Loebe F, Michalek H. General Formal Ontology: Part 1 Basic principles, Version 1. In: Onto-Med Report. Leipzig, Germany: Institute of Medical Informatics, Statistics and Epidemiology, University of Leipzig; 2006.
40. Jeon B, Ko I. Ontology-Based Semi-automatic Construction of Bayesian Network Models for Diagnosing Diseases in E-health Applications. 2007 Presented at: 2007 Frontiers in the Convergence of Bioscience and Information Technologies; 2007; Jeju, Korea p. 592-602. [doi: 10.1109/fbit.2007.63]
41. Dimitrov DV. Medical Internet of Things and Big Data in Healthcare. Healthc Inform Res 2016 Jul;22(3):156-163 [FREE Full text] [doi: 10.4258/hir.2016.22.3.156] [Medline: 27525156]
42. Ram S, Jinsoo Park. Semantic conflict resolution ontology (SCROL): an ontology for detecting and resolving data and schema-level semantic conflicts. IEEE Trans. Knowl. Data Eng 2004 Feb;16(2):189-202. [doi: 10.1109/tkde.2004.1269597]
43. Zouaq A, Nkambou R. Evaluating the Generation of Domain Ontologies in the Knowledge Puzzle Project. IEEE Trans. Knowl. Data Eng 2009 Nov;21(11):1559-1572. [doi: 10.1109/tkde.2009.25]
44. Alhir S. Introduction to the Unified Modeling Language (UML). In: Guide to Applying the UML. Berlin, Germany: Springer; 2002:1-11.
45. Sir M, Bradac Z, Fiedler P. Ontology versus Database. IFAC-PapersOnLine 2015;48(4):220-225. [doi: 10.1016/j.ifacol.2015.07.036]
46. Holsapple CW, Joshi KD. A collaborative approach to ontology design. Commun. ACM 2002 Feb;45(2):42-47. [doi: 10.1145/503124.503147]
47. Bibault J, Zapletal E, Rance B, Giraud P, Burgun A. Labeling for Big Data in radiation oncology: The Radiation Oncology Structures ontology. PLoS One 2018;13(1):e0191263 [FREE Full text] [doi: 10.1371/journal.pone.0191263] [Medline: 29351341]
48. Vossen G, Lytras M, Koudas N. Editorial: Revisiting the (Machine) Semantic Web: The Missing Layers for the Human Semantic Web. IEEE Trans. Knowl. Data Eng 2007 Feb;19(2):145-148. [doi: 10.1109/tkde.2007.30]
49. Hitzler P, Krotzsch M, Rudolph S. Foundations of Semantic Web Technologies. Boca Raton, FL: CRC Press; Aug 06, 2009.
50. Sirin E, Parsia B, Cuenca Grau B, Kalyanpur A, Katz Y. Pellet: A Practical OWL-DL Reasoner. Journal of Web Semantics 2007;5(2):51-53. [doi: 10.2139/ssrn.3199351]
51. Parsia B, Matentzoglu N, Gonçalves RS, Glimm B, Steigmiller A. The OWL Reasoner Evaluation (ORE) 2015 Competition Report. J Autom Reason 2017;59(4):455-482 [FREE Full text] [doi: 10.1007/s10817-017-9406-8] [Medline: 30069067]
52. Knublauch H, Fergerson R, Noy N, Musen M. The Protégé OWL Plugin: An Open Development Environment for Semantic Web Applications. In: McIlraith SA, Plexousakis D, van Harmelen F, editors. The Semantic Web – ISWC 2004. Berlin, Germany: Springer; 2004:229-243.
53. Editors. URL: http://semanticweb.org/wiki/Editors [accessed 2020-09-28]
54. Reasoners. URL: http://semanticweb.org/wiki/Reasoners [accessed 2020-09-28]
55. Shearer R, Motik B, Horrocks I. HermiT: A Highly-Efficient OWL Reasoner. 2008. URL: https://www.cs.ox.ac.uk/boris.motik/pubs/smh08HermiT.pdf [accessed 2021-03-22]
56. Tsarkov D, Horrocks I. FaCT++ Description Logic Reasonerystem Description. In: Furbach U, Shankar N, editors. Automated Reasoning. Berlin, Germany: Springer; 2006:292-297.
57. Haarslev V, Möller R. RACER System Description. In: Automated Reasoning. Berlin, Germany: Springer; 2001:701-705.
58. Apache Jena Framework. URL: https://jena.apache.org/getting_started/index.html [accessed 2020-09-28]
59. Jena Ontology API. URL: http://jena.apache.org/documentation/ontology/ [accessed 2020-09-28]
60. Poole DL, Mackworth AK. Artificial Intelligence: Foundations of Computational Agents. Cambridge, England: Cambridge University Press; 2010.
61. Marinica C, Guillet F. Knowledge-Based Interactive Postmining of Association Rules Using Ontologies. IEEE Trans. Knowl. Data Eng 2010 Jun;22(6):784-797. [doi: 10.1109/tkde.2010.29]
62. Nutrition. URL: https://www.healthline.com/nutrition [accessed 2020-09-28]
63. Motik B, Shearer R, Horrocks I. Optimized Reasoning in Description Logics Using Hypertableaux. In: Automated Deduction – CADE-21. Berlin, Germany: Springer; 2007:67-83.
64. SPARQL 1.1 Query Language. W3C Recommendation 21 March. 2013. URL: http://www.w3.org/TR/2013/REC-sparql11-query-20130321/ [accessed 2020-09-28]

65. Appreciating SPARQL CONSTRUCT more, Bob DuCharme'sweblog. URL: http://www.snee.com/bobdc.blog/2009/09/appreciating-sparql-construct.html [accessed 2020-09-28]
66. URL: https://nsd.no/ [accessed 2020-09-28]
67. Obesity and overweight. URL: https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight [accessed 2020-09-28]
68. NICE. URL: https://www.nice.org.uk/ [accessed 2020-09-28]
69. Norwegian Dietary Guidelines. URL: https://www.helsedirektoratet.no/brosjyrer/helsedirektoratets-kostrad-brosjyre-og-plakat/Helsedirektoratets%20kostr%C3%A5d%20-%20engelsk.pdf/_/attachment/inline/80f68126-68af-4cec-b2aa-d04069d02471:dcb8efdbe6b6129470ec4969f6639be21a8afd82/Helsedirektoratets%20kostr%C3%A5d%20-%20engelsk.pdf [accessed 2020-09-28]
70. Sheth A, Henson C, Sahoo SS. Semantic Sensor Web. IEEE Internet Comput 2008 Jul;12(4):78-83. [doi: 10.1109/mic.2008.87]
71. Calbimonte JP, Jeung H, Corcho O, Aberer K. Semantic sensor data search in a large-scale federated sensor network. 2011 Presented at: 4th International Workshop on Semantic Sensor Networks; October 23, 2011; Barcelona, Spain p. 23-38.
72. Liu J, Li Y, Tian X, Sangaiah AK, Wang J. Towards Semantic Sensor Data: An Ontology Approach. Sensors (Basel) 2019 Mar 08;19(5):303-306 [FREE Full text] [doi: 10.3390/s19051193] [Medline: 30857211]
73. Julina J, Thenmozhi D. Ontology based EMR for decision making in health care using SNOMED CT. 2012 Presented at: International Conference on Recent Trends in Information Technology; April 19-21, 2012; Chennai, India p. 514-519. [doi: 10.1109/icrtit.2012.6206787]
74. SNOMED-CT Browser. URL: https://browser.ihtsdotools.org/?perspective=full&conceptId1=297976006&edition=enedition&release=v20180731&server=http://browser.ihtsdotools.org/api/v1/snomed&langRefset=900000000000509007? [accessed 2020-09-28]
75. Dong D, Sun Z, Gao F. PCOPM: A Probabilistic CBR Framework for Obesity Prescription Management. In: Advanced Intelligent Computing Theories and Applications. With Aspects of Artificial Intelligence. Berlin, Germany: Springer; 2010:91-99.
76. Shaban-Nejad A, Buckeridge D, Dubé L. COPE: Childhood Obesity Prevention Knowledge Enterprise. In: Artificial Intelligence in Medicine. Berlin, Germany: Springer; 2011:225-229.
77. Taçyıldız Ö, Çelik Ertuğrul D. A decision support system on the obesity management and consultation during childhood and adolescence using ontology and semantic rules. J Biomed Inform 2020 Oct;110:103554. [doi: 10.1016/j.jbi.2020.103554] [Medline: 32911081]
78. Zaragozá I, Guixeres J, Alcañiz M. Ontologies for Intelligent e-Therapy: Application to Obesity. In: Distributed Computing, Artificial Intelligence, Bioinformatics, Soft Computing, and Ambient Assisted Living. Berlin, Germany: Springer; 2009:894-901.
79. Seligman ME, Csikszentmihalyi M. Positive psychology. An introduction. Am Psychol 2000 Jan;55(1):5-14. [doi: 10.1037//0003-066x.55.1.5] [Medline: 11392865]
80. Kim AR, Park H, Song T. Development and Evaluation of an Obesity Ontology for Social Big Data Analysis. Healthc Inform Res 2017 Jul;23(3):159-168 [FREE Full text] [doi: 10.4258/hir.2017.23.3.159] [Medline: 28875050]
81. Motik B, Cuenca Grau B, Horrocks I, Sattler U. Representing ontologies using description logics, description graphs, and rules. Artificial Intelligence 2009 Sep;173(14):1275-1309. [doi: 10.1016/j.artint.2009.06.003]
82. Horrocks I, Kutz O, Sattler U. The Even More Irresistible SROIQ. Palo Alto, CA: AAAI Press; 2006 Presented at: KR'06: Proceedings of the Tenth International Conference on Principles of Knowledge Representation and Reasoning; June 2-5, 2006; Lake District, UK p. 57-67.
83. Motik B, Rosati R. A Faithful Integration of Description Logics with Logic Programming. 2007. URL: http://www.ijcai.org/Proceedings/07/Papers/075.pdf [accessed 2021-03-22]
84. Walking: Your steps to health. URL: https://www.health.harvard.edu/staying-healthy/walking-your-steps-to-health [accessed 2020-09-28]
85. Healthy Diet. URL: https://www.who.int/news-room/fact-sheets/detail/healthy-diet [accessed 2020-09-28]
86. Healthy Lifestyle. URL: https://www.mayoclinic.org/healthy-lifestyle [accessed 2020-09-28]
87. Scala PL, Di Pasquale D, Tresoldi D, Lafortuna CL, Rizzo G, Padula M. Ontology-supported clinical profiling for the evaluation of obesity and related comorbidities. Stud Health Technol Inform 2012;180:1025-1029. [Medline: 22874349]
88. Blood sugar level ranges. URL: https://www.diabetes.co.uk/diabetes_care/blood-sugar-level-ranges.html [accessed 2020-09-28]
89. Cholesterol level. URL: https://www.medicalnewstoday.com/articles/315900 [accessed 2020-09-28]
90. BMI. URL: https://www.nhlbi.nih.gov/health/educational/lose_wt/BMI/bmicalc.htm [accessed 2020-09-28]
91. Starkey IR. Recommendations on blood pressure measurement. Br Med J (Clin Res Ed) 1986 Oct 04;293(6551):886 [FREE Full text] [doi: 10.1136/bmj.293.6551.886] [Medline: 3094704]
92. Friestad M, Wright P. The Persuasion Knowledge Model: How People Cope with Persuasion Attempts. J CONSUM RES 1994 Jun;21(1):1. [doi: 10.1086/209380]

Abbreviations

AI: artificial intelligence
API: application programming interface
BLE: Bluetooth low energy
BP: blood pressure
CDSS: clinical decision support system
DSS: decision support system
ICD-11: International Classification of Diseases (11th edition)
ICT: information and communications technology
KB: knowledge base
LOINC: Logical Observation Identifiers Names and Codes
NICE: National Institute for Health and Care Excellence
RDF: resource description framework
RDF: resource description framework
RDFS: RDF schema
SNOMED CT: Systematized Nomenclature of Medicine—Clinical Terms
SPARQL: Simple Protocol and RDF Query Language
SSN: semantic sensor network
SWRL: semantic web rule language
UMLS: Unified Medical Lexicon System
URI: unified resource identifier
WHO: World Health Organization

Edited by G Eysenbach; submitted 29.09.20; peer-reviewed by D Parry, V Montmirail; comments to author 01.12.20; revised version received 26.01.21; accepted 08.02.21; published 09.04.21.

Please cite as:
Chatterjee A, Prinz A, Gerdes M, Martinez S
An Automatic Ontology-Based Approach to Support Logical Representation of Observable and Measurable Data for Healthy Lifestyle Management: Proof-of-Concept Study
J Med Internet Res 2021;23(4):e24656
URL: https://www.jmir.org/2021/4/e24656
doi:10.2196/24656
PMID:33835031

Original Paper

Fast Healthcare Interoperability Resources (FHIR)–Based Quality Information Exchange for Clinical Next-Generation Sequencing Genomic Testing: Implementation Study

Donghyeong Seong[1], BSc; Sungwon Jung[1], MSc; Sungchul Bae[2], PhD; Jongsuk Chung[3], PhD; Dae-Soon Son[4*], PhD; Byoung-Kee Yi[2,5*], PhD

[1]Department of Health Sciences and Technology, Samsung Advanced Institute for Health Sciences & Technology, Sungkyunkwan University, Seoul, Republic of Korea

[2]Smart Healthcare Research Institute, Samsung Medical Center, Seoul, Republic of Korea

[3]Samsung Genome Institute, Samsung Medical Center, Seoul, Republic of Korea

[4]School of Big Data Science, Data Science Convergence Research Center, Hallym University, Chuncheon, Republic of Korea

[5]Department of Digital Health, Samsung Advanced Institute for Health Sciences & Technology, Sungkyunkwan University, Seoul, Republic of Korea

[*]these authors contributed equally

Corresponding Author:
Byoung-Kee Yi, PhD
Smart Healthcare Research Institute
Samsung Medical Center
81 Irwon-ro
Gangnam-gu
Seoul, 06351
Republic of Korea
Phone: 82 2 3410 1944
Email: byoungkeeyi@gmail.com

Abstract

Background: Next-generation sequencing (NGS) technology has been rapidly adopted in clinical practice, with the scope extended to early diagnosis, disease classification, and treatment planning. As the number of requests for NGS genomic testing increases, substantial efforts have been made to deliver the testing results clearly and unambiguously. For the legitimacy of clinical NGS genomic testing, quality information from the process of producing genomic data should be included within the results. However, most reports provide insufficient quality information to confirm the reliability of genomic testing owing to the complexity of the NGS process.

Objective: The goal of this study was to develop a Fast Healthcare Interoperability Resources (FHIR)–based web app, NGS Quality Reporting (NGS-QR), to report and manage the quality of the information obtained from clinical NGS genomic tests.

Methods: We defined data elements for the exchange of quality information from clinical NGS genomic tests, and profiled a FHIR genomic resource to enable information exchange in a standardized format. We then developed the FHIR-based web app and FHIR server to exchange quality information, along with statistical analysis tools implemented with the R Shiny server.

Results: Approximately 1000 experimental data entries collected from the targeted sequencing pipeline CancerSCAN designed by Samsung Medical Center were used to validate implementation of the NGS-QR app using real-world data. The user can share the quality information of NGS genomic testing and verify the quality status of individual samples in the overall distribution.

Conclusions: This study successfully demonstrated how quality information of clinical NGS genomic testing can be exchanged in a standardized format. As the demand for NGS genomic testing in clinical settings increases and genomic data accumulate, quality information can be used as reference material to improve the quality of testing. This app could also motivate laboratories to perform diagnostic tests to provide high-quality genomic data.

(J Med Internet Res 2021;23(4):e26261) doi:10.2196/26261

KEYWORDS

FHIR; clinical NGS genomic testing; clinical massive parallel sequencing; quality control; genomic reporting

Introduction

Next-generation sequencing (NGS) technology has been rapidly adopted in clinical practice, with the scope extended to early diagnosis, disease classification, and treatment planning [1-3]. To implement clinical NGS applications, health care providers operate their own NGS laboratories or request genomic tests to external laboratories. The NGS genomic testing reports provide information regarding genomic variations and related data. However, different templates and data elements are used depending on the laboratory, and most reports are provided in text or PDF format [4,5]. As the number of requests for NGS genomic testing increases, considerable efforts have been made to deliver the testing results clearly and unambiguously [6-10]. Moreover, international standards development organizations have developed reporting standards such as the Fast Healthcare Interoperability Resources (FHIR) Genomics Reporting Implementation Guide and International Organization for Standardization (ISO) standards to exchange complex clinical genomic data and interpretations [11-14].

For the legitimacy of clinical NGS genomic testing, the quality information from the process of producing the genomic data should be included within the test results [15-17]. The demand for genomic testing of small tissue samples and needle biopsies is increasing [18], and it remains challenging to determine the reliability of the genomic test results. When testing small quantities or low-quality samples, including low-purity specimens and formalin-fixed paraffin-embedded (FFPE) specimens, quality information can be considered for the interpretation of diagnostic results as evidence. Confirming that there is no specific variation is a particular challenge, especially without quality information. For the validity and utility of clinical NGS genomic testing, the US Centers for Medicare and Medicaid Services regulates all laboratory tests performed on humans through the Clinical Laboratory Improvement Amendments, and the US Food and Drug Administration (FDA) requires information regarding the clinical validity for genomic tests [19,20]. In Korea, the Ministry of Food and Drug Safety (MFDS) has initiated the clinical laboratory accreditation program since 2017 [21]; however, this provides accreditation of the entire NGS process rather than the reliability of individual samples. When a clinician performs NGS genomic testing for diagnostic purposes, quality information can be used to interpret the results for individual samples. In addition, when constructing a reference database using these samples, it is essential to include this quality information.

The successful practice of precision medicine depends on clinical genomic data sharing and knowledge-based interpretations of genomic variant data at the point of care [11]. To improve interoperability as part of precision medicine, health care stakeholders encourage the use of application programming interfaces (APIs) and app-based ecosystems such as SMART on FHIR, CDS Hooks, and SMART Markers [22-26]. These platforms enable easy implementation for health care use cases and facilitate functional extensibility. There are several apps based on clinical genomics use cases on the SMART on FHIR platforms, such as the SMART Precision Cancer Medicine and SMART Cancer Navigator apps [27-29].

In this study, we developed the NGS Quality Reporting (NGS-QR) app to exchange the quality information of clinical NGS genomic testing. To exchange the information in a standardized format, we profiled a FHIR genomic resource based on ISO/TS 22692:2020 Genomics Informatics-Quality Control Metrics for DNA sequencing, which defines the quality-related data for the entire NGS process, including sample preparation, library preparation, sequencing, and data processing [15]. This app enables the performance comparison of clinical NGS genomic testing and monitoring of the quality status of the current sample as the data accumulate.

Methods

Overview

This study describes the development of the NGS-QR app, a FHIR-based web app, to report, manage, and monitor the quality information from the process of producing genomic data. As shown in Table 1, the development was composed of the following phases: (1) requirement analysis, (2) design, (3) implementation, and (4) testing. In the requirement analysis phase, we defined use cases and selected data elements such as DNA purity and integrity, library input amount and size, and sequencing running quality, in accordance with standards and guidelines. In the design phase, we profiled a FHIR resource on the existing genomic resource to exchange quality information in a standardized manner. We also designed user interfaces and functions for the NGS-QR app. In the development phase, we developed the three following components: web app (NGS-QR app), FHIR server, and R Shiny server. The web app includes a FHIR resource handler to generate and parse FHIR resources. Finally, in the testing phase, the NGS-QR app was validated using real-world data collected from the NGS pipeline, a targeted sequencing platform.

Table 1. Overview of development of the next-generation sequencing-quality reporting (NGS-QR) app.

Development phase	Description
Requirement analysis	
Use cases	Define use cases for quality information exchange of clinical NGS[a] genomic testing
Data elements	Select data elements in accordance with standards and guidelines (eg, ISO[b], US FDA[c], ACMG[d])
Design	
FHIR[e] genomic profile	Profile a FHIR resource on the existing genomic resource, MolecularSequence
System components	Design user interfaces and functions of the NGS-QR app
Implementation	
Web app (NGS-QR app)	Develop user interfaces and functions of the web app; develop a FHIR resource handler to generate and parse FHIR resources
FHIR server	Develop a FHIR server and repository; apply the FHIR QcMetrics profile to the FHIR server
R Shiny server	Develop R code for statistical analysis of NGS experimental data
Testing	
Data collection	Collect real-world data generated from the NGS pipeline
Data exchange	Exchange real-world data between the FHIR server and NGS-QR app using FHIR APIs[f]
Quality management	Check the quality information of NGS genomic testing using the dashboard and statistical analysis

[a]NGS: next-generation sequencing.

[b]ISO: International Organization for Standardization.

[c]US FDA: US Food and Drug Administration.

[d]ACMG: American College of Medical Genetics.

[e]FHIR: Fast Healthcare Interoperability Resources.

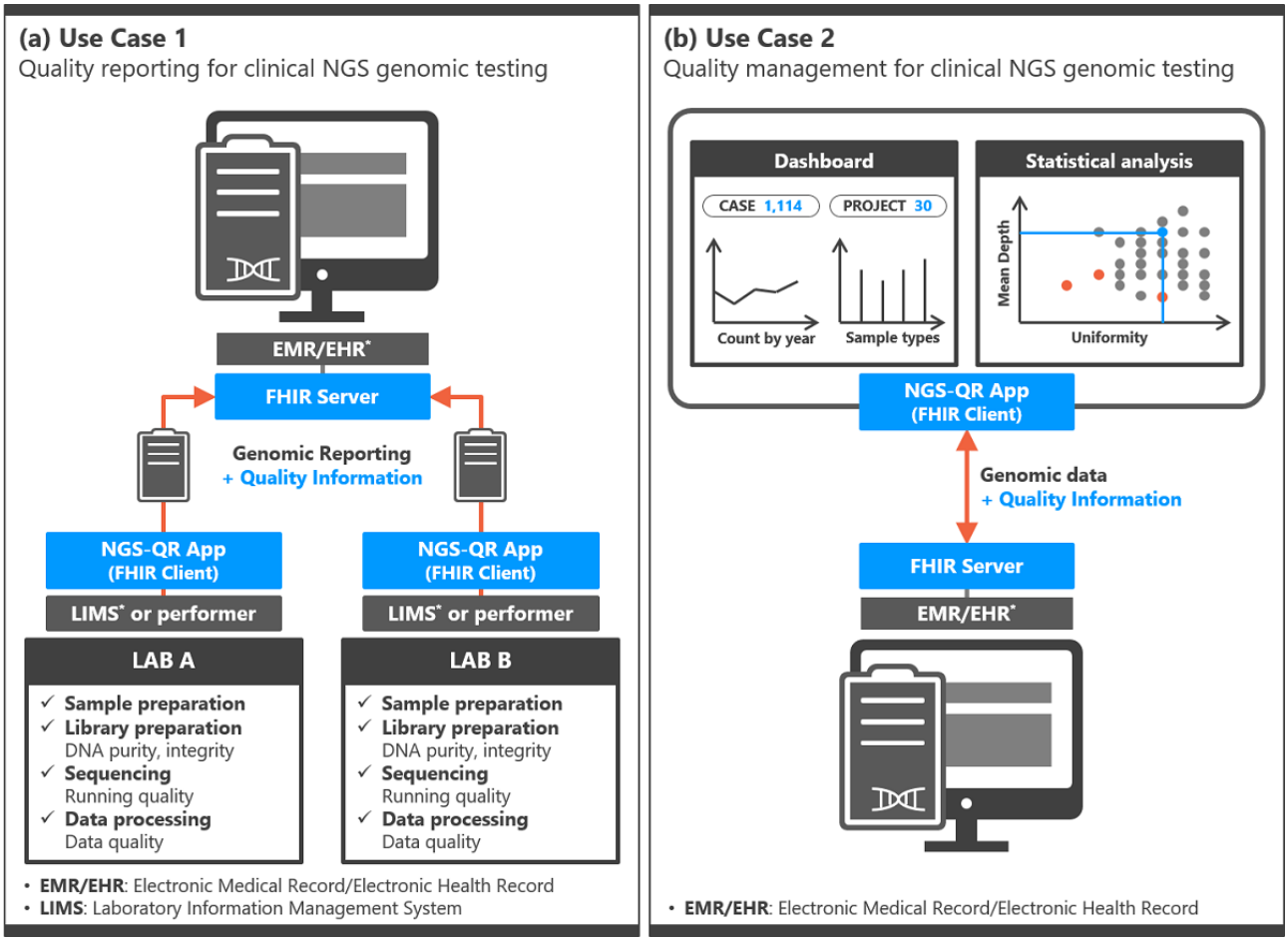
[f]APIs: application programming interfaces.

Use Cases

We defined two use cases: quality information reporting and quality management for clinical NGS genomic testing. The first use case is related to the quality information reporting for clinical NGS genomic testing, as shown in Figure 1a. The laboratory reports the results of NGS genomic testing to the hospital along with quality information. This use case assumes that the hospital information system provides the FHIR APIs. The laboratory requested for the NGS genomic testing sends the results using the NGS-QR app, which acts as an FHIR client. When the user fills out the reporting form in the NGS-QR app, a FHIR resource is created in JSON format and then the resource is sent to the server via the FHIR API. Hospitals that request NGS genomic testing to multiple laboratories can receive information in the same format through this standardized method.

The second use case is related to the quality management for clinical NGS genomic testing, as shown in Figure 1b. Users such as health care providers and performers of NGS testing manage and monitor the quality status of the genomic data received from the laboratories. Through the NGS-QR app, users can retrieve experimental data from the FHIR server and view the summary such as the total number of genomic tests and the number of tests based on the year or specimen type for all genomic test results. Moreover, quality information such as DNA purity, integrity, and data quality for each sample can be compared through statistical analysis. This step can be used to determine the reliability of clinical NGS genomic testing.

Figure 1. Use cases for exchanging the quality information of clinical next-generation sequencing (NGS) genomic testing. (a) Quality reporting for clinical NGS genomic testing. (b) Quality management for clinical NGS genomic testing. FHIR: Fast Healthcare Interoperability Resources.



Data Elements Selection

The NGS workflow involves complex procedures consisting of several steps, which are broadly divided into sample preparation, library preparation, sequencing, and data processing. In this process, various types of experimental conditions and results are generated and captured. As shown in Table 2, we selected data elements based on the ISO 22692 [15] and guidelines such as those of the FDA [20], American College of Medical Genetics [30], and College of American Pathologists [31] for quality management and comparative analysis. For this study, approximately 30 data elements were selected.

Table 2. Next-generation sequencing (NGS) workflow and data elements.

NGS workflow	Data elements
Sample preparation	
Sample sequencing type	sequencing type, target gene
Sample information	specimen type, sampling date
Library preparation	
DNA extraction	DNA extraction kit, DNA purity (eg, OD[a] 260/280, OD 260/230), DNA integrity (eg, DNA median size)
Library construction	library input amount, library input size, library construction kit
Sequencing	
Sequencing information	sequencing instrument, read length, sequencing direction, running mode
Running quality	error rate, percent data quality (>Q30)
Data processing	
Data quality	total reads, mean coverage, uniformity, on-target rate, Q30, PR[b] score
Sequencing alignment	mapping algorithm, sequencing alignment software
Variant calling	variant calling software, quality score, allelic read percentage
Variant filtering and annotation	germline filter criteria, reference database

[a]OD: optical density.

[b]PR: pass rate.

FHIR Genomics Resource Profile

We profiled a FHIR genomic resource by defining constraints and extensions for exchanging quality information with a standardized method. Since the results of NGS genomic testing include genomic variant information and its quality-related data, the FHIR R4 MolecularSequence resource [32] was used for tailoring the NGS data to the use cases in this study. We created a QcMetrics element (a FHIR extension) in the resource and added data elements for each process under QcMetrics. Textbox 1 shows an example of a FHIR genomics resource that includes the QcMetrics profile. We uploaded the profile to the public FHIR registry SIMPLIFHIR.NET [33].

Textbox 1. Example of the Fast Healthcare Interoperability Resources genomic resource that includes the QcMetrics profile.


```
{
“resourceType”: “MolecularSequence”,
“extension”: [
{
“url”: “http://example.org/fhir/StructureDefinition/QcMetrics”,
“extension”: [
{
“url”: “dnaExtraction”,
“extension”: [
{
“url”: “dnaExtractionKit”,
“valueString”: “qiagen allprep DNA/RNA mini kit”
},
{
“url”: “dnaPurity”,
“extension”: [
{
“url”: “od260280”,
“valueDecimal”: 2.1
},
{
“url”: “od260230”,
“valueDecimal”: 2.3
}]
},
{
“url”: “dnaIntegrity”,
“units”: “bp”,
“valueDecimal”: 60000
}]
},
…
}
```

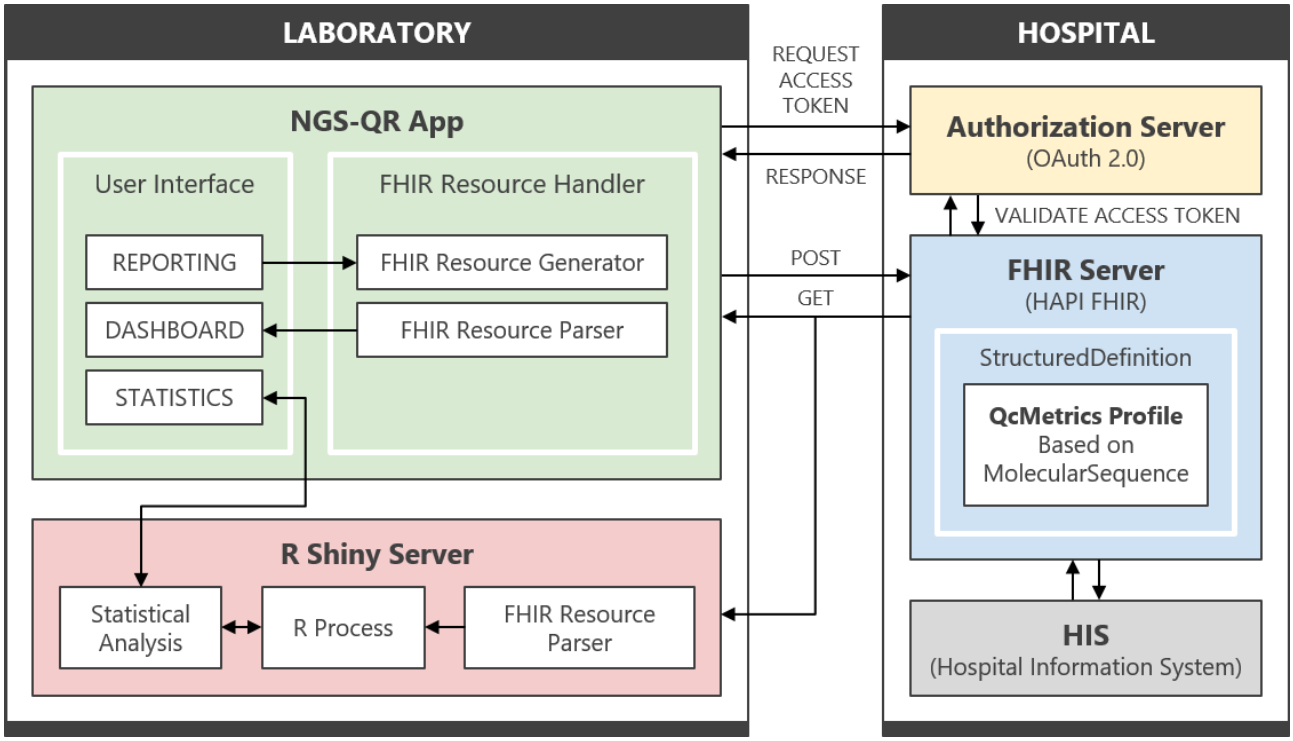

Implementation

We developed the following three major components: the NGS-QR app (FHIR client), FHIR server, and R Shiny server, as shown in Figure 2. The NGS-QR app was developed using Node.js (v12.13.1), which is composed of user interfaces (UIs), a FHIR resource handler, and a REST API module. The app has been deployed to the App Gallery of SMART on FHIR, which is an open platform for substitutable third-party health apps to connect to electronic medical record (EMR)/electronic health record (EHR) systems with appropriate security guarantees [34,35]. The source code of the app is available at GitHub [36]. The FHIR server was locally installed using the HAPI FHIR server (v4.0.0), which is an open-source Java implementation of the FHIR specification. The QcMetrics profile was added to the FHIR server. The source code for operating the R Shiny server was written in RStudio Cloud (R version 3.6.3) and the server was deployed to the R cloud platform, Shinyapp.io [37]. The R Shiny server fetched the data used for statistical analysis from the FHIR server. The R source code is available at GitHub [38].

FHIR has a set of security recommendations that identifies communications security, authentication, authorization, access control, and auditing [39]. We applied Transport Layer Security

(TLS) and OAuth 2.0, which are industry-standard protocols for communications security and authentication [40,41]. The TLS was used for the encryption of FHIR resources transmitted between the NGS-QR app and the FHIR server. The OAuth 2.0 protocol was used to grant the NGS-QR app access to the FHIR server. The NGS-QR app requests an access token by authenticating with the Authorization Server. The Authorization Server authenticates the NGS-QR app and issues an access token. The access token was used for security credentials for the NGS-QR app to make API requests on behalf of a user.

Figure 2. Concept model of the NGS-QR app, FHIR server, and R Shiny server. FHIR: Fast Healthcare Interoperability Resources.



Results

Data Collection

Samsung Medical Center (SMC) has developed and utilized a cancer panel sequencing pipeline, namely CancerSCAN, to determine effective treatment methods for patients based on data from more than 15,000 panel sequencing studies since 2014 [17,42,43]. The SMC received clinical laboratory accreditation by the MFDS in 2017, and CancerSCAN is clinically used for the diagnosis and prognosis of cancer patients [17]. For this study, approximately 1000 data entries were collected from CancerSCAN. They contained the experimental conditions and results for FFPE specimens, fresh cells, and cell lines.

User Interfaces and Functions

In the reporting UI, input data are converted to a FHIR genomic resource in JSON format using the FHIR resource generator, and the resource is sent to the FHIR server using the POST method, as shown in Figure 3. In the dashboard UI, genomic resources in the FHIR server are compiled using the GET method to display the summary of data such as the total number of genomic tests and the number of tests based on the year and specimen type for the complete genomic test results. In the statistics UI, the user can select the group of samples, and view the distribution and threshold of each experimental parameter. The user can also check the current quality status of each sample in the overall distribution, as shown in Figure 4.

Figure 3. Screenshots of the user interface: reporting quality information of clinical next-generation sequencing (NGS) genomic testing.

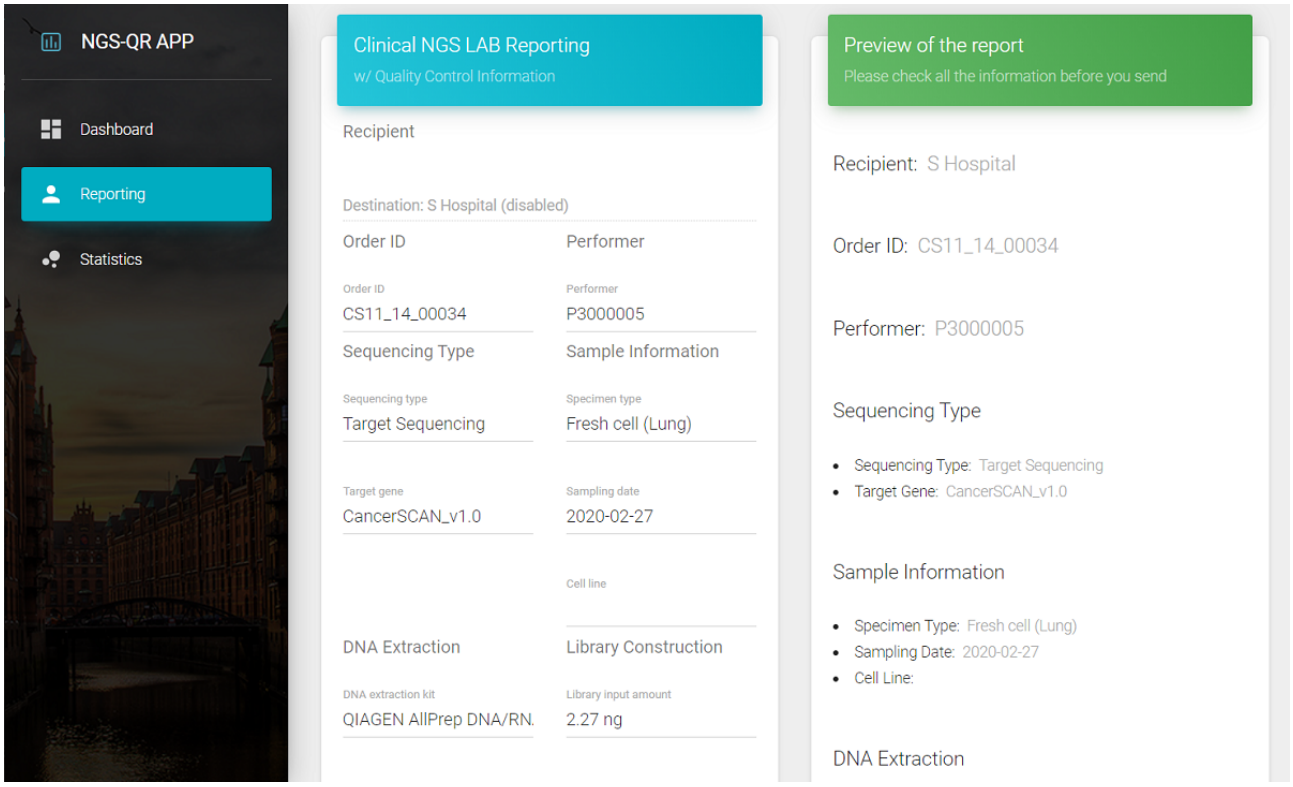
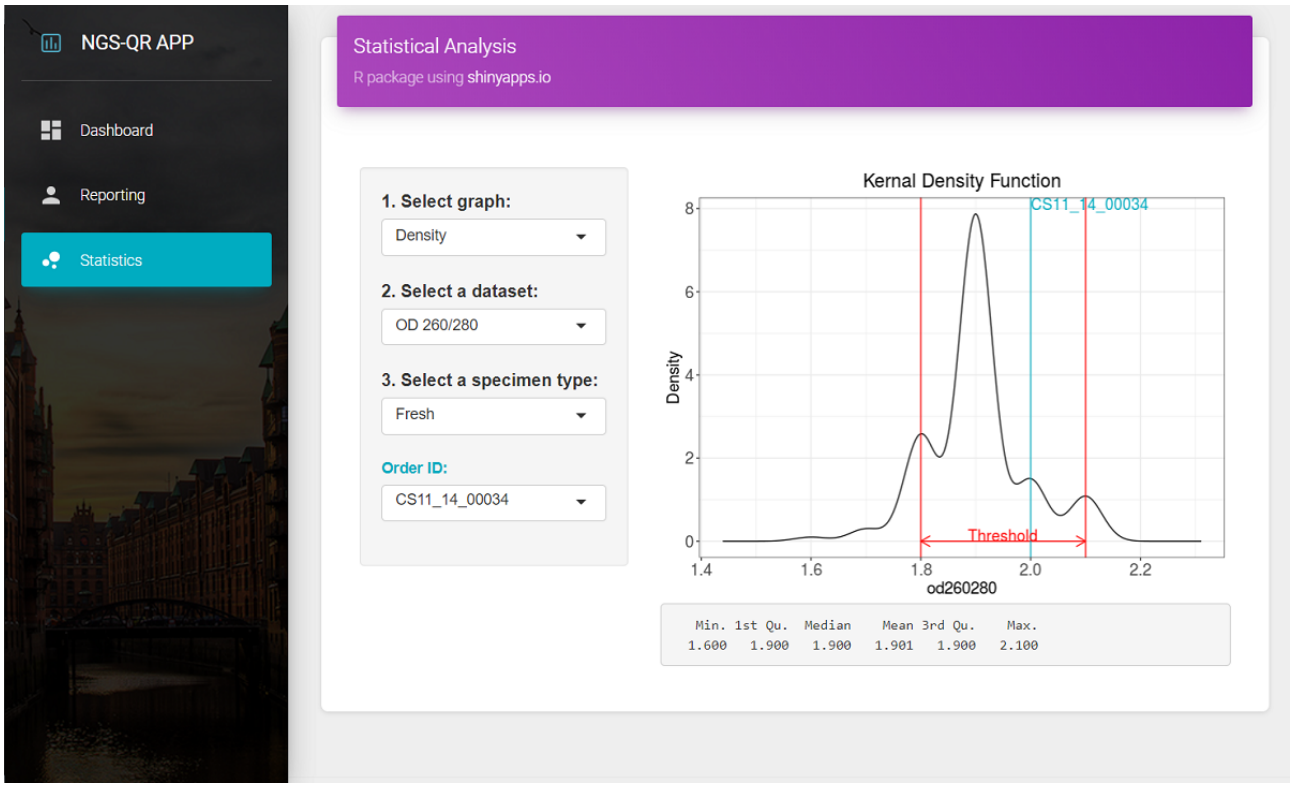


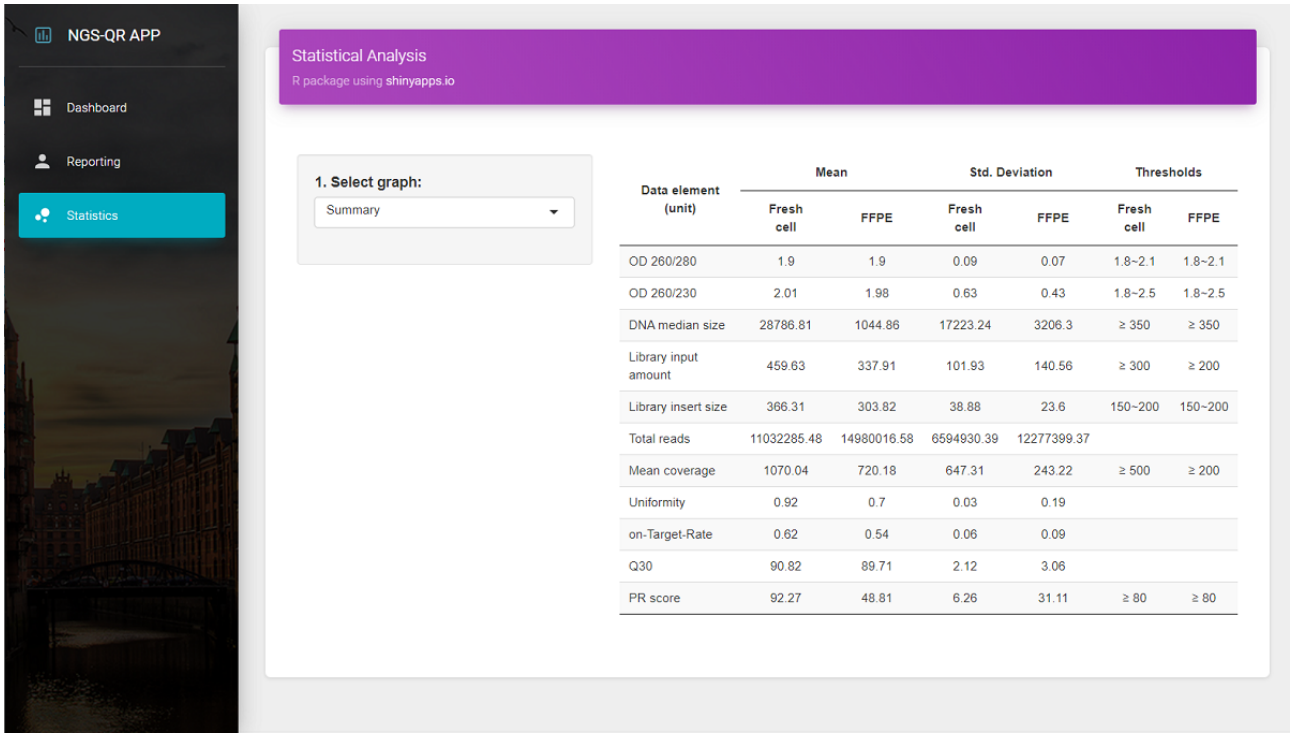
Figure 4. Screenshots of the user interface: quality management of clinical next-generation sequencing genomic testing.



Data Statistics

We validated the utility of the NGS-QR app using real-world data from the NGS pipeline CancerSCAN in SMC. All data were converted to FHIR genomic resources and sent to the FHIR server using the NGS-QR app. Figure 5 shows the statistics of the experimental data used in this study. As expected, fresh cell samples had better quality than FFPE samples in almost every category.

Figure 5. Statistics of the experimental data used in this study. FFPE: formalin-fixed paraffin-embedded; OD: optical density; PR: pass rate.



Data element (unit)	Mean		Std. Deviation		Thresholds	
	Fresh cell	FFPE	Fresh cell	FFPE	Fresh cell	FFPE
OD 260/280	1.9	1.9	0.09	0.07	1.8~2.1	1.8~2.1
OD 260/230	2.01	1.98	0.63	0.43	1.8~2.5	1.8~2.5
DNA median size	28786.81	1044.86	17223.24	3206.3	≥ 350	≥ 350
Library input amount	459.63	337.91	101.93	140.56	≥ 300	≥ 200
Library insert size	366.31	303.82	38.88	23.6	150~200	150~200
Total reads	11032285.48	14980016.58	6594930.39	12277399.37		
Mean coverage	1070.04	720.18	647.31	243.22	≥ 500	≥ 200
Uniformity	0.92	0.7	0.03	0.19		
on-Target-Rate	0.62	0.54	0.06	0.09		
Q30	90.82	89.71	2.12	3.06		
PR score	92.27	48.81	6.26	31.11	≥ 80	≥ 80

Discussion

Principal Results

NGS technology has been widely adopted in clinical practice, which will play a prominent role in precision medicine. Although many countries and international institutions have developed NGS guidelines to implement clinical NGS applications, the quality criteria for determining the reliability of NGS genomic testing have not yet been standardized due to the complexity of the NGS technology. Since the scope and purpose of NGS applications are very diverse, it remains challenging to evaluate the validity and utility of clinical NGS genomic testing. At present, most individual laboratories have their own processes and criteria; thus, it is important to share quality information from the process of generating genomic data.

In this study, we propose the NGS-QR app to exchange quality information for clinical NGS genomic testing. This study provides the following main contributions to the field. First, we demonstrated that the quality information managed only by individual laboratories could be shared using a standardized format. In the NGS workflow, the experimental conditions change depending on the purpose of the tests and the status of the samples. Thus far, it has been difficult to determine the conditions under which the genome data were generated. This study defined quality-related elements and profiled the FHIR genomic resource to report the experimental conditions and results produced in NGS genomic testing. Since our proposed method is based on ISO/TS 22692 and HL7 FHIR standards to interoperably share the quality-related data of clinical genomic testing, the app can communicate with any EMR/EHR systems that conform to these standards.

Second, this study facilitated the verification of the quality status of each sample as experimental data accumulate. The quality of genome data is determined through comparative analysis based on the characteristics of NGS technology. As shown in Figure 6, the results from different types of specimens cannot be directly compared because this may lead to incorrect conclusions. Therefore, it is important to select the target groups and determine the performance of individual samples within them. The NGS-QR app allows users to select various experimental conditions and compare the produced data with the correct target group to check the exact data quality.

Figure 6. Example of comparative analysis based on specimen type using the NGS-QR app. FFPE: formalin-fixed paraffin-embedded; PR: pass rate.

Conclusions

This study successfully demonstrated how the quality information of clinical NGS genomic testing can be exchanged using a standardized method. As the demand for NGS genomic testing increases and genomic data accumulate, quality information can be used as reference material for improving the quality of testing. This approach can also motivate laboratories to perform diagnostic tests to provide high-quality genomic data.

Acknowledgments

This work was supported by the Industrial Strategic Technology Development Program (grant number 10078282) funded by the Ministry of Trade, Industry & Energy (MOTIE) and by the Korea Health Technology R&D Project (grant number HI19C1026) through the Korea Health Industry Development Institute (KHIDI) funded by the Ministry of Health & Welfare, Republic of Korea.

Conflicts of Interest

None declared.

References

1. Kamps R, Brandão RD, Bosch BJVD, Paulussen ADC, Xanthoulea S, Blok MJ, et al. Next-generation sequencing in oncology: genetic diagnosis, risk prediction and cancer classification. Int J Mol Sci 2017 Jan 31;18(2):308 [FREE Full text] [doi: 10.3390/ijms18020308] [Medline: 28146134]
2. Biesecker LG, Green RC. Diagnostic clinical genome and exome sequencing. N Engl J Med 2014 Jun 19;370(25):2418-2425. [doi: 10.1056/NEJMra1312543] [Medline: 24941179]
3. Katsanis SH, Katsanis N. Molecular genetic testing and the future of clinical genomics. Nat Rev Genet 2013 Jun;14(6):415-426 [FREE Full text] [doi: 10.1038/nrg3493] [Medline: 23681062]
4. Tarczy-Hornoch P, Amendola L, Aronson SJ, Garraway L, Gray S, Grundmeier RW, et al. A survey of informatics approaches to whole-exome and whole-genome clinical reporting in the electronic health record. Genet Med 2013 Oct;15(10):824-832 [FREE Full text] [doi: 10.1038/gim.2013.120] [Medline: 24071794]
5. Davis KW, Hamby Erby L, Fiallos K, Martin M, Wassman ER. A comparison of genomic laboratory reports and observations that may enhance their clinical utility for providers and patients. Mol Genet Genomic Med 2019 Jul;7(7):e00551. [doi: 10.1002/mgg3.551] [Medline: 31115190]

6. Knoppers BM, Zawati MH, Sénécal K. Return of genetic testing results in the era of whole-genome sequencing. Nat Rev Genet 2015 Sep;16(9):553-559. [doi: 10.1038/nrg3960] [Medline: 26239711]
7. Kearney HM, Thorland EC, Brown KK, Quintero-Rivera F, South ST, Working Group of the American College of Medical Genetics Laboratory Quality Assurance Committee. American College of Medical Genetics standards and guidelines for interpretation and reporting of postnatal constitutional copy number variants. Genet Med 2011 Jul;13(7):680-685. [doi: 10.1097/GIM.0b013e3182217a3a] [Medline: 21681106]
8. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, ACMG Laboratory Quality Assurance Committee. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med 2015 May;17(5):405-424 [FREE Full text] [doi: 10.1038/gim.2015.30] [Medline: 25741868]
9. Li MM, Datto M, Duncavage EJ, Kulkarni S, Lindeman NI, Roy S, et al. Standards and guidelines for the interpretation and reporting of sequence variants in cancer: A Joint Consensus Recommendation of the Association for Molecular Pathology, American Society of Clinical Oncology, and College of American Pathologists. J Mol Diagn 2017 Jan;19(1):4-23 [FREE Full text] [doi: 10.1016/j.jmoldx.2016.10.002] [Medline: 27993330]
10. Haga SB, Mills R, Pollak KI, Rehder C, Buchanan AH, Lipkus IM, et al. Developing patient-friendly genetic and genomic test reports: formats to promote patient engagement and understanding. Genome Med 2014;6(7):58 [FREE Full text] [doi: 10.1186/s13073-014-0058-6] [Medline: 25473429]
11. Alterovitz G, Heale B, Jones J, Kreda D, Lin F, Liu L, et al. FHIR Genomics: enabling standardization for precision medicine use cases. NPJ Genom Med 2020;5:13. [doi: 10.1038/s41525-020-0115-6] [Medline: 32194985]
12. ISO/TS 20428:2017 Health informatics — Data elements and their metadata for describing structured clinical genomic sequence information in electronic health records. ISO. URL: https://www.iso.org/standard/67981.html [accessed 2021-04-20]
13. ISO/PRF TS 22693 Genomics informatics — Structured clinical gene fusion report in electronic health records. ISO. URL: https://www.iso.org/standard/73694.html [accessed 2021-04-20]
14. Ryu B, Shin S, Baek R, Kim J, Heo E, Kang I, et al. Clinical genomic sequencing reports in electronic health record systems based on international standards: implementation study. J Med Internet Res 2020 Aug 10;22(8):e15040 [FREE Full text] [doi: 10.2196/15040] [Medline: 32773368]
15. ISO/TS 22692:2020 Genomics informatics— Quality control metrics for DNA sequencing. ISO. URL: https://www.iso.org/standard/73693.html [accessed 2021-04-20]
16. ISO/DTS 22690 Genomics informatics — Reliability assessment criteria for high-throughput gene-expression data. ISO. URL: https://www.iso.org/standard/73691.html [accessed 2021-04-20]
17. Seong D, Chung J, Lee K, Kim S, Kim B, Song J, et al. Benchmark database for process optimization and quality control of clinical cancer panel sequencing. Biotechnol Bioproc E 2019 Oct 25;24(5):793-798. [doi: 10.1007/s12257-019-0202-7]
18. Patel NM, Jo H, Eberhard DA, Yin X, Hayward MC, Stein MK, et al. Improved tumor purity metrics in next-generation sequencing for clinical practice: the Integrated Interpretation of Neoplastic Cellularity and Sequencing Results (IINCaSe) approach. Appl Immunohistochem Mol Morphol 2019;27(10):764-772 [FREE Full text] [doi: 10.1097/PAI.0000000000000684] [Medline: 30102605]
19. Clinical Laboratory Improvement Amendments (CLIA). US Centers for Medicare and Medicaid Services (CMS). URL: https://www.cms.gov/Regulations-and-Guidance/Legislation/CLIA [accessed 2021-04-20]
20. Considerations for Design, Development, and Analytical Validation of Next Generation Sequencing-Based In Vitro Diagnostics Intended to Aim in the Diagnosis of Suspected Germline Diseases. US Food and Drug Administration (FDA). 2018. URL: https://www.fda.gov/media/99208/download [accessed 2021-04-20]
21. Kim J, Park W, Kim NKD, Jang SJ, Chun S, Sung C, Molecular Pathology Study Group of Korean Society of Pathologists. Good Laboratory Standards for clinical next-generation sequencing cancer panel tests. J Pathol Transl Med 2017 May;51(3):191-204. [doi: 10.4132/jptm.2017.03.14] [Medline: 28535585]
22. Lehne M, Sass J, Essenwanger A, Schepers J, Thun S. Why digital medicine depends on interoperability. NPJ Digit Med 2019;2:79. [doi: 10.1038/s41746-019-0158-1] [Medline: 31453374]
23. Mandl KD, Mandel JC, Kohane IS. Driving innovation in health systems through an apps-based information economy. Cell Syst 2015 Jul;1(1):8-13 [FREE Full text] [doi: 10.1016/j.cels.2015.05.001] [Medline: 26339683]
24. Mandel JC, Kreda DA, Mandl KD, Kohane IS, Ramoni RB. SMART on FHIR: a standards-based, interoperable apps platform for electronic health records. J Am Med Inform Assoc 2016 Sep;23(5):899-908 [FREE Full text] [doi: 10.1093/jamia/ocv189] [Medline: 26911829]
25. Dolin RH, Boxwala A, Shalaby J. A pharmacogenomics clinical decision support service based on FHIR and CDS hooks. Methods Inf Med 2018 Dec;57(S02):e115-e123 [FREE Full text] [doi: 10.1055/s-0038-1676466] [Medline: 30605914]
26. Sayeed R, Gottlieb D, Mandl KD. SMART Markers: collecting patient-generated health data as a standardized property of health information technology. NPJ Digit Med 2020;3:9. [doi: 10.1038/s41746-020-0218-6] [Medline: 31993507]
27. Alterovitz G, Warner J, Zhang P, Chen Y, Ullman-Cullere M, Kreda D, et al. SMART on FHIR Genomics: facilitating standardized clinico-genomic apps. J Am Med Inform Assoc 2015 Nov;22(6):1173-1178. [doi: 10.1093/jamia/ocv045] [Medline: 26198304]

28. Warner JL, Rioth MJ, Mandl KD, Mandel JC, Kreda DA, Kohane IS, et al. SMART precision cancer medicine: a FHIR-based app to provide genomic information at the point of care. J Am Med Inform Assoc 2016 Jul;23(4):701-710 [FREE Full text] [doi: 10.1093/jamia/ocw015] [Medline: 27018265]
29. Warner JL, Prasad I, Bennett M, Arniella M, Beeghly-Fadiel A, Mandl KD, et al. SMART Cancer Navigator: A framework for implementing ASCO Workshop recommendations to enable precision cancer medicine. JCO Precis Oncol 2018;2018:PO.17.00292 [FREE Full text] [doi: 10.1200/PO.17.00292] [Medline: 30238071]
30. Rehm HL, Bale SJ, Bayrak-Toydemir P, Berg JS, Brown KK, Deignan JL, Working Group of the American College of Medical GeneticsGenomics Laboratory Quality Assurance Commitee. ACMG clinical laboratory standards for next-generation sequencing. Genet Med 2013 Sep;15(9):733-747 [FREE Full text] [doi: 10.1038/gim.2013.92] [Medline: 23887774]
31. Aziz N, Zhao Q, Bry L, Driscoll DK, Funke B, Gibson JS, et al. College of American Pathologists' laboratory standards for next-generation sequencing clinical tests. Arch Pathol Lab Med 2015 Apr;139(4):481-493 [FREE Full text] [doi: 10.5858/arpa.2014-0250-CP] [Medline: 25152313]
32. FHIR Resource MolecularSequence. HL7.org. URL: https://www.hl7.org/fhir/molecularsequence.html [accessed 2021-04-20]
33. Quality control metrics for DNA sequencing. SIMPLIFIER.NET. URL: https://simplifier.net/qualtiycontorlmetricsfordnasequencing [accessed 2021-04-20]
34. Deployment of NGS-QR app. Heroku.com. URL: https://ngs-qr.herokuapp.com [accessed 2021-04-20]
35. NGS Quality Reporting (NGS-QR) App. SMART App Gallery. URL: https://apps.smarthealthit.org/app/ngs-quality-reporting-ngs-qr-app [accessed 2021-04-20]
36. NGS Quality Reporting App. GitHub.com. URL: https://github.com/dhseong/ngs-qr-app-client [accessed 2021-04-20]
37. Deployment of Shiny app for statistical analysis in NGS-QR app. Shinyapp.io. URL: https://dhseong.shinyapps.io/NGS-QR-App [accessed 2021-04-20]
38. R code for statistical analysis in NGS-QR app. GitHub.com. URL: https://github.com/dhseong/Statistics-for-NGS-QR-app [accessed 2021-04-20]
39. FHIR Security. HL7.org. URL: https://www.hl7.org/fhir/security.html [accessed 2021-04-20]
40. Transport Layer Security (TLS) Protocol. URL: https://tools.ietf.org/html/rfc5246 [accessed 2021-04-20]
41. OAuth 2.0 Authorization Framework. URL: https://tools.ietf.org/html/rfc6749 [accessed 2021-04-20]
42. Kim ST, Kim K, Kim NKD, Park JO, Ahn S, Yun J, et al. Clinical application of targeted deep sequencing in solid-cancer patients and utility for biomarker-selected clinical trials. Oncologist 2017 Oct;22(10):1169-1177. [doi: 10.1634/theoncologist.2017-0020] [Medline: 28701572]
43. Shin H, Choi Y, Yun JW, Kim NKD, Kim S, Jeon HJ, et al. Prevalence and detection of low-allele-fraction variants in clinical cancer samples. Nat Commun 2017 Nov 09;8(1):1377. [doi: 10.1038/s41467-017-01470-y] [Medline: 29123093]

Abbreviations

API: application programming interface
EHR: electronic health record
EMR: electronic medical record
FDA: Food and Drug Administration
FFPE: formalin-fixed paraffin-embedded
FHIR: Fast Healthcare Interoperability Resource
ISO: International Organization for Standardization
MFDS: Ministry of Food and Drug Safety
NGS: next-generation sequencing
NGS-QR: Next-Generation Sequencing Quality Reporting
SMC: Samsung Medical Center
TLS: Transport Layer Security
UI: user interface

Edited by G Eysenbach; submitted 04.12.20; peer-reviewed by T Sagi, J Lee, L Guo; comments to author 29.12.20; revised version received 10.01.21; accepted 03.04.21; published 28.04.21.

Please cite as:
Seong D, Jung S, Bae S, Chung J, Son DS, Yi BK
Fast Healthcare Interoperability Resources (FHIR)–Based Quality Information Exchange for Clinical Next-Generation Sequencing Genomic Testing: Implementation Study
J Med Internet Res 2021;23(4):e26261
URL: https://www.jmir.org/2021/4/e26261
doi:10.2196/26261
PMID:33908889

Corrigenda and Addenda

Correction: Spelling Errors and Shouting Capitalization Lead to Additive Penalties to Trustworthiness of Online Health Information: Randomized Experiment With Laypersons

Harry J Witchel[1], PhD; Georgina A Thompson[1], BMBS; Christopher I Jones[2], CSci, CStat, PhD; Carina E I Westling[3], PhD; Juan Romero[4]; Alessia Nicotra[4], MBBS, PhD, FRCP; Bruno Maag[4], TGmeF, HFG; Hugo D Critchley[1], MBChB, DPhil, FRCPsych

[1]Department of Neuroscience, Brighton and Sussex Medical School, Brighton, United Kingdom

[2]Department of Primary Care and Public Health, Brighton and Sussex Medical School, Brighton, United Kingdom

[3]Faculty of Media and Communication, Bournemouth University, Bournemouth, United Kingdom

[4]Dalton Maag Ltd, London, United Kingdom

Corresponding Author:
Harry J Witchel, PhD
Department of Neuroscience
Brighton and Sussex Medical School
Trafford Centre for Medical Research
Brighton,
United Kingdom
Phone: 44 1273 873 549
Email: h.witchel@bsms.ac.uk

Related Article:

Correction of: https://www.jmir.org/2020/6/e15171

(J Med Internet Res 2021;23(4):e29452) doi:10.2196/29452

In "Spelling Errors and Shouting Capitalization Lead to Additive Penalties to Trustworthiness of Online Health Information: Randomized Experiment With Laypersons" (J Med Internet Res 2020;22(6):e15171) one error was noted after publication.

One reviewer of this paper, L Sbaffi, was listed twice in the originally published paper due to a system error. The duplicate instance of the reviewer's name has been removed from the corrected version.

The correction will appear in the online version of the paper on the JMIR Publications website on April 13, 2021, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

Submitted 07.04.21; this is a non–peer-reviewed article;accepted 07.04.21; published 13.04.21.

Please cite as:
Witchel HJ, Thompson GA, Jones CI, Westling CEI, Romero J, Nicotra A, Maag B, Critchley HD
Correction: Spelling Errors and Shouting Capitalization Lead to Additive Penalties to Trustworthiness of Online Health Information: Randomized Experiment With Laypersons
J Med Internet Res 2021;23(4):e29452
URL: https://www.jmir.org/2021/4/e29452
doi:10.2196/29452
PMID:33848255

Corrigenda and Agenda

Correction: Theory Integration for Lifestyle Behavior Change in the Digital Age: An Adaptive Decision-Making Framework

Chao Zhang[1], PhD; Daniël Lakens[1], PhD; Wijnand A IJsselsteijn[1], PhD

Human-Technology Interaction Group, Department of Industrial Engineering & Innovation Sciences, Eindhoven University of Technology, Eindhoven, Netherlands

Corresponding Author:
Chao Zhang, PhD
Human-Technology Interaction Group, Department of Industrial Engineering & Innovation Sciences, Eindhoven University of Technology
PO Box 513
Eindhoven, 5600 MB
Netherlands
Phone: 31 624749479
Email: chao.zhang87@gmail.com

Related Article:

Correction of: https://www.jmir.org/2021/4/e17127

(J Med Internet Res 2021;23(4):e29629) doi:10.2196/29629

In "Theory Integration for Lifestyle Behavior Change in the Digital Age: An Adaptive Decision-Making Framework" (J Med Internet Res 2021;23(4):e17127) the authors noted one error.

In the originally published manuscript, the heading "Mapping Digital Intervention Techniques to the Framework" was incorrectly set as a level 3 subheading rather than a level 2 heading. This has been changed to a level 2 heading in the corrected version of the manuscript.

The correction will appear in the online version of the paper on the JMIR Publications website on April 15, 2021, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

Submitted 14.04.21; this is a non–peer-reviewed article;accepted 14.04.21; published 15.04.21.

Please cite as:
Zhang C, Lakens D, IJsselsteijn WA
Correction: Theory Integration for Lifestyle Behavior Change in the Digital Age: An Adaptive Decision-Making Framework
J Med Internet Res 2021;23(4):e29629
URL: https://www.jmir.org/2021/4/e29629
doi:10.2196/29629
PMID:33857012

Corrigenda and Addenda

Correction: Assessing Public Interest Based on Wikipedia's Most Visited Medical Articles During the SARS-CoV-2 Outbreak: Search Trends Analysis

Jędrzej Chrzanowski[1*]; Julia Sołek[1,2*], MD; Wojciech Fendler[1], MD, PhD; Dariusz Jemielniak[3], PhD

[1]Department of Biostatistics and Translational Medicine, Medical University of Łódź, Łódź, Poland

[2]Department of Pathology, Medical University of Łódź, Łódź, Poland

[3]Management in Networked and Digital Societies, Kozminski University, Warszawa, Poland

[*]these authors contributed equally

Corresponding Author:
Wojciech Fendler, MD, PhD
Department of Biostatistics and Translational Medicine
Medical University of Łódź
Mazowiecka 15
Łódź, 92-215
Poland
Phone: 48 422722585
Email: wojciech.fendler@umed.lodz.pl

Related Article:

Correction of: https://jmir.org/2021/4/e26331

(J Med Internet Res 2021;23(4):e29598) doi:10.2196/29598

In "Assessing Public Interest Based on Wikipedia's Most Visited Medical Articles During the SARS-CoV-2 Outbreak: Search Trends Analysis" (J Med Internet Res 2021;23(4):e26331) the authors noted two errors.

Due to a system error, the name of one author, *Wojciech Fendler*, was replaced with the name of another author on the paper, *Dariusz Jemielniak*. In the originally published paper, the order of authors was listed as follows:

> *Jędrzej Chrzanowski; Julia Sołek; Dariusz Jemielniak; Dariusz Jemielniak*

This has been corrected to:

> *Jędrzej Chrzanowski; Julia Sołek; Wojciech Fendler; Dariusz Jemielniak*

In the originally published paper, the ORCID of author 'Wojciech Fendler' was incorrectly published as follows:

> *Wojciech Fendler: 0000-0002-3745-7931*

This has been corrected to:

> *Wojciech Fendler: 0000-0002-5083-9168*

The correction will appear in the online version of the paper on the JMIR Publications website on April 15, 2021, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

Submitted 13.04.21; this is a non–peer-reviewed article;accepted 13.04.21; published 15.04.21.

Please cite as:
Chrzanowski J, Sołek J, Fendler W, Jemielniak D
Correction: Assessing Public Interest Based on Wikipedia's Most Visited Medical Articles During the SARS-CoV-2 Outbreak: Search Trends Analysis
J Med Internet Res 2021;23(4):e29598
URL: https://www.jmir.org/2021/4/e29598
doi:10.2196/29598
PMID:33857009

Corrigenda and Addenda

Correction: Potential Correlates of Internet Gaming Disorder Among Indonesian Medical Students: Cross-sectional Study

Kristiana Siste[1], MD, PhD; Enjeline Hanafi[1], BMedSci, MD; Lee Thung Sen[1], MRES, MD; Petra Octavian Perdana Wahjoepramono[2], MD, BMedSci; Andree Kurniawan[3], MD; Ryan Yudistiro[2], MD

[1]Department of Psychiatry, Faculty of Medicine, Universitas Indonesia - Dr Cipto Mangunkusumo General Hospital, Jakarta, Indonesia

[2]Faculty of Medicine, Universitas Pelita Harapan, Siloam Hospitals, Tangerang, Indonesia

[3]Department of Internal Medicine, Faculty of Medicine, Universitas Pelita Harapan Siloam Hospital, Tangerang, Indonesia

Corresponding Author:
Petra Octavian Perdana Wahjoepramono, MD, BMedSci
Faculty of Medicine
Universitas Pelita Harapan
Siloam Hospitals
Jl Siloam No 6
Tangerang, 15811
Indonesia
Phone: 62 811971169
Email: petra.wahjoepramono@uph.edu

Related Article:

Correction of: https://jmir.org/2021/4/e25468

(J Med Internet Res 2021;23(4):e29790) doi:10.2196/29790

In “Potential Correlates of Internet Gaming Disorder Among Indonesian Medical Students: Cross-sectional Study” (J Med Internet Res 2021;23(4):e25468) the authors noted two errors.

Due to a system error, the name of one author, *Andree Kurniawan*, was replaced with the name of another author on the paper, *Ryan Yudistiro*. In the originally published paper, the order of authors was listed as follows:

> *Kristiana Siste; Enjeline Hanafi; Lee Thung Sen; Petra Octavian Perdana Wahjoepramono; Ryan Yudistiro; Ryan Yudistiro*

This has been corrected to:

> *Kristiana Siste; Enjeline Hanafi; Lee Thung Sen; Petra Octavian Perdana Wahjoepramono; Andree Kurniawan; Ryan Yudistiro*

In the originally published paper, the ORCID number of author *Ryan Yudistiro* was incorrectly published as follows:

> *0000-0002-5219-9029*

This has been corrected to:

> *0000-0003-1418-2661*

The correction will appear in the online version of the paper on the JMIR Publications website on April 21, 2021, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

Submitted 20.04.21; this is a non–peer-reviewed article;accepted 20.04.21; published 21.04.21.

Please cite as:
Siste K, Hanafi E, Sen LT, Wahjoepramono POP, Kurniawan A, Yudistiro R
Correction: Potential Correlates of Internet Gaming Disorder Among Indonesian Medical Students: Cross-sectional Study
J Med Internet Res 2021;23(4):e29790
URL: https://www.jmir.org/2021/4/e29790
doi:10.2196/29790
PMID:33882024

Corrigenda and Addenda

Correction: Health Care Cybersecurity Challenges and Solutions Under the Climate of COVID-19: Scoping Review

Ying He[1], PhD; Aliyu Aliyu[2], MSc; Mark Evans[2], PhD; Cunjin Luo[3,4], PhD

[1]School of Computer Science, University of Nottingham, Nottingham, United Kingdom

[2]School of Computer Science and Informatics, De Montfort University, Leicester, United Kingdom

[3]School of Computer Science and Electronic Engineering, University of Essex, Colchester, United Kingdom

[4]Key Lab of Medical Electrophysiology, Ministry of Education, Institute of Cardiovascular Research, Southwest Medical University, Luzhou, China

Corresponding Author:
Cunjin Luo, PhD
School of Computer Science and Electronic Engineering
University of Essex
Wivenhoe Park
Colchester, CO4 3SQ
United Kingdom
Phone: 44 7493622995
Email: cunjin.luo@essex.ac.uk

Related Article:

Correction of: https://www.jmir.org/2021/4/e21747

(J Med Internet Res 2021;23(4):e29877) doi:10.2196/29877

In "Health Care Cybersecurity Challenges and Solutions Under the Climate of COVID-19: Scoping Review" (J Med Internet Res 2021;23(4):e21747) the authors noted one error.

In the originally published manuscript, a footnote was erroneously displayed in the authorship list noting equal contribution only for author Cunjin Luo.

This footnote has been removed from the corrected manuscript, as no equal contribution should be noted in the authorship list.

The correction will appear in the online version of the paper on the JMIR Publications website on April 28, 2021, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

Submitted 23.04.21; this is a non–peer-reviewed article;accepted 23.04.21; published 28.04.21.

Please cite as:
He Y, Aliyu A, Evans M, Luo C
Correction: Health Care Cybersecurity Challenges and Solutions Under the Climate of COVID-19: Scoping Review
J Med Internet Res 2021;23(4):e29877
URL: https://www.jmir.org/2021/4/e29877
doi:10.2196/29877
PMID:33909589

Corrigenda and Addenda

Correction: Work-Related and Personal Factors Associated With Mental Well-Being During the COVID-19 Response: Survey of Health Care and Other Workers

Bradley A Evanoff[1,2], MD, MPH; Jaime R Strickland[1,2], MA; Ann Marie Dale[1,2], PhD, OTR/L; Lisa Hayibor[1], MPH; Emily Page[3], BFA; Jennifer G Duncan[1], MD; Thomas Kannampallil[1], PhD; Diana L Gray[1], MD

[1]Washington University School of Medicine, St. Louis, MO, United States

[2]Healthier Workforce Center of the Midwest, University of Iowa, Iowa City, IA, United States

[3]Human Resources, Washington University in St. Louis, St. Louis, MO, United States

Corresponding Author:
Bradley A Evanoff, MD, MPH
Washington University School of Medicine
4523 Clayton Avenue
Box 8005
St. Louis, MO, 63110
United States
Phone: 1 3144548340
Email: bevanoff@wustl.edu

Related Article:

Correction of: https://www.jmir.org/2020/8/e21366/

(J Med Internet Res 2021;23(4):e29069) doi:10.2196/29069

In "Work-Related and Personal Factors Associated With Mental Well-Being During the COVID-19 Response: Survey of Health Care and Other Workers" (J Med Internet Res 2020;22(8):e21366) the authors noted two errors in the reference list.

In the originally published paper, the following reference was omitted from the reference list:

> *Wu AW, Connors C, Everly GS. COVID-19: Peer Support and Crisis Communication Strategies to Promote Institutional Resilience. Annals of Internal Medicine 2020 Jun 16;172(12):822-823. [doi: 10.7326/m20-1236]*

This citation has been inserted as Reference 3 in the corrected version of the paper.

In addition, References 20 and 21 were duplicate references in the originally published paper. In the corrected version, this reference (McGinty et al) remains as Reference 21. The remaining references have been renumbered accordingly to adjust for these changes. The corrected and renumbered list of references appears below and in the corrected paper.

The correction will appear in the online version of the paper on the JMIR Publications website on April 9, 2021, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

References

1. Lai J, Ma S, Wang Y, Cai Z, Hu J, Wei N, et al. Factors Associated With Mental Health Outcomes Among Health Care Workers Exposed to Coronavirus Disease 2019. JAMA Netw Open 2020 Mar 02;3(3):e203976 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.3976] [Medline: 32202646]
2. Bettinsoli M, Di RD, Napier J, Moretti L, Bettinsoli P, Delmedico M, et al. Psychological Impact and Contextual Factors Associated With Physical and Mental Health Conditions of Italian Healthcare Professionals During the Covid-19 Disease Outbreak. PsyArXiv Preprints 2020:1-22. [doi: 10.31234/osf.io/w89fz]
3. Wu AW, Connors C, Everly GS. COVID-19: Peer Support and Crisis Communication Strategies to Promote Institutional Resilience. Annals of Internal Medicine 2020 Jun 16;172(12):822-823. [doi: 10.7326/m20-1236]

4. Ripp J, Peccoralo L, Charney D. Attending to the Emotional Well-Being of the Health Care Workforce in a New York City Health System During the COVID-19 Pandemic. Acad Med 2020 Aug;95(8):1136-1139. [doi: 10.1097/acm.0000000000003414]
5. Shanafelt T, Ripp J, Trockel M. Understanding and Addressing Sources of Anxiety Among Health Care Professionals During the COVID-19 Pandemic. JAMA 2020 Apr 07;323(21):2133. [doi: 10.1001/jama.2020.5893] [Medline: 32259193]
6. Galea S, Merchant RM, Lurie N. The Mental Health Consequences of COVID-19 and Physical Distancing: The Need for Prevention and Early Intervention. JAMA Intern Med 2020 Apr 10:2020. [doi: 10.1001/jamainternmed.2020.1562] [Medline: 32275292]
7. Petterson S, Westfall J, Miller B. Projected Deaths of Despair from COVID-19. Well Being Trust. 2020 Apr. URL: https://wellbeingtrust.org/wp-content/uploads/2020/05/WBT_Deaths-of-Despair_COVID-19-FINAL-FINAL.pdf [accessed 2020-08-20]
8. Kannampallil TG, Goss CW, Evanoff BA, Strickland JR, McAlister RP, Duncan J. Exposure to COVID-19 patients increases physician trainee stress and burnout. PLoS One 2020 Aug 6;15(8):e0237301. [doi: 10.1371/journal.pone.0237301] [Medline: 32760131]
9. Hammer LB, Ernst Kossek E, Bodner T, Crain T. Measurement development and validation of the Family Supportive Supervisor Behavior Short-Form (FSSB-SF). J Occup Health Psychol 2013 Jul;18(3):285-296 [FREE Full text] [doi: 10.1037/a0032612] [Medline: 23730803]
10. Lovibond SH, Lovibond PF. Manual for the depression anxiety stress scales. Sydney: Psychology Foundation of Australia; 1996.
11. Trockel M, Bohman B, Lesure E, Hamidi MS, Welle D, Roberts L, et al. A Brief Instrument to Assess Both Burnout and Professional Fulfillment in Physicians: Reliability and Validity, Including Correlation with Self-Reported Medical Errors, in a Sample of Resident and Practicing Physicians. Acad Psychiatry 2018 Feb 1;42(1):11-24 [FREE Full text] [doi: 10.1007/s40596-017-0849-3] [Medline: 29196982]
12. Serxner S, Kichlu R, Ratelis E. Consumer sentiment during a time of global crisis. Optum Inc. 2020. URL: https://www.optum.com/content/dam/optum3/optum/en/resources/PDFs/2020-comparative-insights-during-global-crisis.pdf [accessed 2020-08-20]
13. R Foundation for Statistical Computing. R: A language environment for statistical computing. URL: https://www.r-project.org/ [accessed 2020-08-20]
14. RStudio Inc. RStudio: Integrated Development for R. Boston, MA URL: https://rstudio.com/ [accessed 2020-08-20]
15. Kossek EE, Pichler S, Bodner T, Hammer LB. Workplace social support and work-family conflict: a meta-analysis clarifying the influence of general and work-family-specific supervisor and organizational support. Pers Psychol 2011;64(2):289-313 [FREE Full text] [doi: 10.1111/j.1744-6570.2011.01211.x] [Medline: 21691415]
16. King R, Karuntzos G, Casper L. Work-family balance issues and work-leave policies. In: Handbook of occupational health and wellness. Boston, MA: Springer; 2012:323-340.
17. Hammer LB, Kossek EE, Yragui NL, Bodner TE, Hanson GC. Development and Validation of a Multidimensional Measure of Family Supportive Supervisor Behaviors (FSSB). J Manage 2009 Aug;35(4):837-856 [FREE Full text] [doi: 10.1177/0149206308328510] [Medline: 21660254]
18. Chew NW, Lee GK, Tan BY, Jing M, Goh Y, Ngiam NJ, et al. A multinational, multicentre study on the psychological outcomes and associated physical symptoms amongst healthcare workers during COVID-19 outbreak. Brain Behav Immun 2020 Aug;88:559-565 [FREE Full text] [doi: 10.1016/j.bbi.2020.04.049] [Medline: 32330593]
19. Kisely S, Warren N, McMahon L, Dalais C, Henry I, Siskind D. Occurrence, prevention, and management of the psychological effects of emerging virus outbreaks on healthcare workers: rapid review and meta-analysis. BMJ 2020 May 05;369:m1642 [FREE Full text] [doi: 10.1136/bmj.m1642] [Medline: 32371466]
20. Kirzinger A, Kearney A, Hamel L, Brodie M. KFF Health Tracking Poll - Early April 2020: The Impact of Coronavirus on Life in America. Kaiser Family Foundation. 2020 Apr 2. URL: https://www.kff.org/coronavirus-covid-19/report/kff-health-tracking-poll-early-april-2020/ [accessed 2020-08-20]
21. McGinty EE, Presskreischer R, Han H, Barry CL. Psychological Distress and Loneliness Reported by US Adults in 2018 and April 2020. JAMA 2020 Jun 03;324(1):94 [FREE Full text] [doi: 10.1001/jama.2020.9740] [Medline: 32492088]
22. De Brier N, Stroobants S, Vandekerckhove P, De Buck E. Factors affecting mental health of health care workers during coronavirus disease outbreaks: a rapid systematic review. PsyArXiv Preprints 2020:1-17. [doi: 10.31234/osf.io/w9uxs]
23. Dzau VJ, Kirch D, Nasca T. Preventing a Parallel Pandemic — A National Strategy to Protect Clinicians' Well-Being. N Engl J Med 2020 Aug 06;383(6):513-515. [doi: 10.1056/nejmp2011027]
24. Panagioti M, Panagopoulou E, Bower P, Lewith G, Kontopantelis E, Chew-Graham C, et al. Controlled Interventions to Reduce Burnout in Physicians: A Systematic Review and Meta-analysis. JAMA Intern Med 2017 Feb 01;177(2):195-205. [doi: 10.1001/jamainternmed.2016.7674] [Medline: 27918798]
25. Odle-Dusseau HN, Hammer LB, Crain TL, Bodner TE. The influence of family-supportive supervisor training on employee job performance and attitudes: An organizational work-family intervention. J Occup Health Psychol 2016 Jul;21(3):296-308. [doi: 10.1037/a0039961] [Medline: 26652264]

26. Hammer LB, Kossek EE, Anger WK, Bodner T, Zimmerman KL. Clarifying work-family intervention processes: the roles of work-family conflict and family-supportive supervisor behaviors. J Appl Psychol 2011 Jan;96(1):134-150 [FREE Full text] [doi: 10.1037/a0020927] [Medline: 20853943]

Submitted 24.03.21; this is a non–peer-reviewed article;accepted 25.03.21; published 09.04.21.

Please cite as:
Evanoff BA, Strickland JR, Dale AM, Hayibor L, Page E, Duncan JG, Kannampallil T, Gray DL
Correction: Work-Related and Personal Factors Associated With Mental Well-Being During the COVID-19 Response: Survey of Health Care and Other Workers
J Med Internet Res 2021;23(4):e29069
URL: https://www.jmir.org/2021/4/e29069
doi:10.2196/29069
PMID:33835934

Original Paper

Comparison of Public Responses to Containment Measures During the Initial Outbreak and Resurgence of COVID-19 in China: Infodemiology Study

Xinyu Zhou[1*], BSc; Yi Song[1*], BSc; Hao Jiang[1], MSc; Qian Wang[1], MSc; Zhiqiang Qu[1], BSc; Xiaoyu Zhou[1], BSc; Mark Jit[2], PhD; Zhiyuan Hou[1,3], PhD; Leesa Lin[2], PhD

[1]School of Public Health, Fudan University, Shanghai, China

[2]Department of Infectious Disease Epidemiology, London School of Hygiene & Tropical Medicine, London, United Kingdom

[3]NHC Key Laboratory of Health Technology Assessment, Fudan University, Shanghai, China

[*]these authors contributed equally

Corresponding Author:
Zhiyuan Hou, PhD
School of Public Health
Fudan University
Mailbox 250, 138# Yixueyuan Road, Xuhui District
Shanghai, 200032
China
Phone: 86 21 33563935
Email: zyhou@fudan.edu.cn

Abstract

Background: COVID-19 cases resurged worldwide in the second half of 2020. Not much is known about the changes in public responses to containment measures from the initial outbreak to resurgence. Monitoring public responses is crucial to inform policy measures to prepare for COVID-19 resurgence.

Objective: This study aimed to assess and compare public responses to containment measures during the initial outbreak and resurgence of COVID-19 in China.

Methods: We curated all COVID-19–related posts from Sina Weibo (China's version of Twitter) during the initial outbreak and resurgence of COVID-19 in Beijing, China. With a Python script, we constructed subsets of Weibo posts focusing on 3 containment measures: lockdown, the test-trace-isolate strategy, and suspension of gatherings. The Baidu open-source sentiment analysis model and latent Dirichlet allocation topic modeling, a widely used machine learning algorithm, were used to assess public engagement, sentiments, and frequently discussed topics on each containment measure.

Results: A total of 8,985,221 Weibo posts were curated. In China, the containment measures evolved from a complete lockdown for the general population during the initial outbreak to a more targeted response strategy for high-risk populations during COVID-19 resurgence. Between the initial outbreak and resurgence, the average daily proportion of Weibo posts with negative sentiments decreased from 57% to 47% for the lockdown, 56% to 51% for the test-trace-isolate strategy, and 55% to 48% for the suspension of gatherings. Among the top 3 frequently discussed topics on lockdown measures, discussions on containment measures accounted for approximately 32% in both periods, but those on the second-most frequently discussed topic shifted from the expression of negative emotions (11%) to its impacts on daily life or work (26%). The public expressed a high level of panic (21%) during the initial outbreak but almost no panic (1%) during resurgence. The more targeted test-trace-isolate measure received the most support (60%) among all 3 containment measures in the initial outbreak, and its support rate approached 90% during resurgence.

Conclusions: Compared to the initial outbreak, the public expressed less engagement and less negative sentiments on containment measures and were more supportive toward containment measures during resurgence. Targeted test-trace-isolate strategies were more acceptable to the public. Our results indicate that when COVID-19 resurges, more targeted test-trace-isolate strategies for high-risk populations should be promoted to balance pandemic control and its impact on daily life and the economy.

(J Med Internet Res 2021;23(4):e26518) doi:10.2196/26518

KEYWORDS

COVID-19; engagement; latent Dirichlet allocation; public response; sentiment; social media; topic modeling

Introduction

In December 2019, COVID-19 emerged in Wuhan, and propagated rapidly across China and worldwide [1]. In response, many countries implemented stringent large-scale containment measures such as lockdowns, quarantines, and suspension of mass gatherings. Following these containment measures, the number of new COVID-19 cases decreased significantly from its peak in the first half of 2020 [2-6]. In China, the government adopted a rapid nationwide lockdown in late January 2020 during the Chinese New Year holiday, which included stay-at-home orders, transportation block, closure of shops and schools, suspension of gatherings, and suspension of work after the national Spring Festival holiday (the Chinese New Year), when most individuals in China would travel across cities or provinces for family gatherings. The lockdown measure aimed to prevent movement and mass gatherings in order to limit community transmission of COVID-19. As the then epidemic was under control, many provinces lifted containment measures from late February 2020 [7].

Although very few COVID-19 cases were thereafter reported in China [7], smaller-scale resurgences occurred. On June 11, 2020, a confirmed case, not linked to international travel, was reported in Beijing, ending the city's almost 2-month span of zero incidence of local infections [8-10]. During the resurgence, Beijing adopted a more targeted response strategy instead of the city-level lockdown that was implemented during the initial outbreak. This targeted response strategy, namely "test-trace-isolate," principally consisted of nucleic acid testing of individuals who had contact with known cases, tracing of close contacts of confirmed or suspected cases, and isolation of vendors and customers of Xinfadi Market (the epicenter of COVID-19 resurgence in Beijing). By rapidly identifying Xinfadi Market as the disease epicenter, most businesses and schools in Beijing were allowed to remain operational. Furthermore, instead of suspending all modes of transport, restrictions were promptly applied to passengers or goods entering or leaving Beijing, which prevented the virus from spreading outside Beijing. Moreover, the health code system was widely used to identify the exposure risk within the population [11]. One month later, no new cases were reported, with a total of 335 confirmed cases, and on July 19, 2020, Beijing lifted its response strategy. Worldwide, there has also been a resurgence of COVID-19 since August 2020, and evidence-based response strategies are thus needed to fully prepare for yet another resurgence [6,12-15].

The public response may change from that during the initial outbreak to resurgence and hence needs to be monitored. Previous studies have reported that the public frequently discussed the containment measures on social media, including lockdown, test-trace-isolate strategies, suspension of gatherings, personal protection, social distancing, travel restrictions, and workplace closures [16-18]. At the early stage of the COVID-19 pandemic, the public complied with containment measures in many countries, but their compliance gradually weakened as the pandemic progressed [19]. Studies also reported that both Weibo and Twitter users expressed more negative sentiments in the early stage of the COVID-19 pandemic [20-22]. In late 2020, many countries faced a COVID-19 resurgence and restarted containment measures. In India, Twitter users held a positive viewpoint to the second lockdown, but the majority held a negative opinion regarding the third lockdown [23]. In the Philippines, the negative sentiments increased owing to food shortage and helplessness during the lockdown [24]. Prolonged containment measures may lead to decreased risk perception, increased negative sentiments, and fatigue with sustaining containment measures [25-27]. Therefore, it is necessary to continuously monitor the public's response to containment measures and design effective public communication strategies. Previous studies have reported that public risk perception and negative sentiments (such as depression, anxiety, and frustration) were highly correlated with the implementation of containment measures [16,26,28,29]. However, to our knowledge, none of these studies has assessed public responses to containment measures during COVID-19 resurgence and compared them across different stages of the pandemic.

With the protraction of pandemics, containment measures need to be adjusted accordingly, and it is imperative to gain timely feedback on containment measures from the public. Social media has been increasingly recognized as a platform for social surveillance [7,30]. Compared to surveys, social media not only allows for the monitoring of fluctuations in public sentiment and responses over a longer period but also is limited by a low recall bias. Using a machine learning approach based on social media data, this study aimed to assess and compare public responses to containment measures during the initial outbreak and resurgence of COVID-19 in China, including the level of public engagement, sentiments expressed, and frequently discussed topics. This study presents primary data on the evolution of public responses toward the progression of the COVID-19 pandemic, which would help inform adjustments in containment measures. Understanding the shifts of public responses could be essential for policymakers to prepare for future resurgence of COVID-19 globally.

Methods

Study Design

This is a comparative study based on social media data. With over 500,000,000 users, Sina Weibo (China's version of Twitter) is the most influential social media platform in China [31]. Weibo allows users to share information and opinions in real time through posts and has been widely used to identify public concerns during the COVID-19 pandemic. We compared the public response to containment measures between the initial outbreak and resurgence on the basis of the following indicators: (1) the number of relevant Weibo posts, (2) the prevalence of negative sentiments in Weibo posts, and (3) the proportion of frequently discussed topics on Weibo. All data in this study are

publicly available, and this study is exempt from ethical approval.

Data Collection

Weibo posts that contain specific words can be retrieved through Sina Weibo's keyword search function. We programmed a crawler in Python to curate publicly available Weibo posts through a keyword search. Since COVID-19 spread across China during the initial outbreak and only propagated in Beijing during resurgence, we retrieved all COVID-19–related Weibo posts in China for the initial outbreak and posts in Beijing for resurgence. For both surges of the pandemic, we curated Weibo posts from 1 week before the outbreak to the time when affected areas began to lift their responses (January 13 to February 28, 2020, for the initial outbreak and June 4 to July 20, 2020, for resurgence). In total, we curated 8,985,221 Weibo posts, and the data set is available on GitHub [32].

Data Preprocessing

The initial pool of Weibo posts was preprocessed using a Python script to exclude duplicates, remove hashtags, links, uniform resource locators, and user handles from each post to clean the text [33,34]. We extracted the last user's comment if it was a repost, and we excluded Weibo posts that were irrelevant to the public's response by matching patterns with Python [35]. Finally, keyword matching was carried out from all COVID-19–related Weibo posts to extract Weibo posts targeting 3 primary containment measures: lockdown, the test-trace-isolate strategy, and suspension of gatherings (Table 1). A flowchart of data collection and preprocessing is shown in Multimedia Appendix 1.

Table 1. Keywords related to containment measures associated with the COVID-19 pandemic.

Containment measure	Keywords
Lockdown	"封城"(city lockdown), "封村"(village lockdown), "封路"(block road), "封闭"(close), "响应"(public health response), "停运"(stop traffic), "暂时关闭"(temporarily closed)
Test-trace-isolate strategy	"体温监测"/"测温"(body temperature monitoring), "检测"(test), "排查"(trace), "隔离"(isolate), "强制"(enforce/enforcement), "控制"(control)
Suspension of gatherings	"暂停"(suspension of gathering), "停止"(call off), "取消"(cancel), "推迟"(put off school opening or returning to work), "延期"(postpone), "延长"(prolong vacation)

Data Analysis

We analyzed public engagement, sentiments, and frequently discussed topics related to each of the 3 containment measures and compared them between COVID-19 resurgence and the initial outbreak. Public engagement toward containment measures was assessed on the basis of the daily number of related Weibo posts, which was compared with the daily number of new COVID-19 cases locally reported by the National Health Commission of China and Beijing Municipal Health Commission. Public sentiment toward containment measures was analyzed using the Baidu open-source sentiment analysis application programming interface and measured from the proportion of Weibo posts with negative sentiments. Frequently discussed topics regarding containment measures on social media were identified through latent Dirichlet allocation (LDA) topic modeling combined with manual annotation.

Topic modeling is an unsupervised machine learning technique that can automatically identify underlying topics or clusters by identifying groups of words that often co-occur in a textual data set (ie, Weibo posts) [9,17,36]. LDA is a widely used topic modeling algorithm to identify the most common topics across social media platforms [36,37]. In LDA, each document (ie, a Weibo post) is assumed to contain different topics, and each topic can be captured from a set of words. This helps map the given documents to the set of topics, such that the words in each document can be mostly captured by those topics. We applied LDA topic modeling by separating all documents into 30 machine-generated topics, and every Weibo post was assigned to a topic that it most likely belonged to according to the LDA model. LDA outputs provide keywords of the 30 LDA-generated topics for each containment measure during the initial outbreak and resurgence (Multimedia Appendix 2).

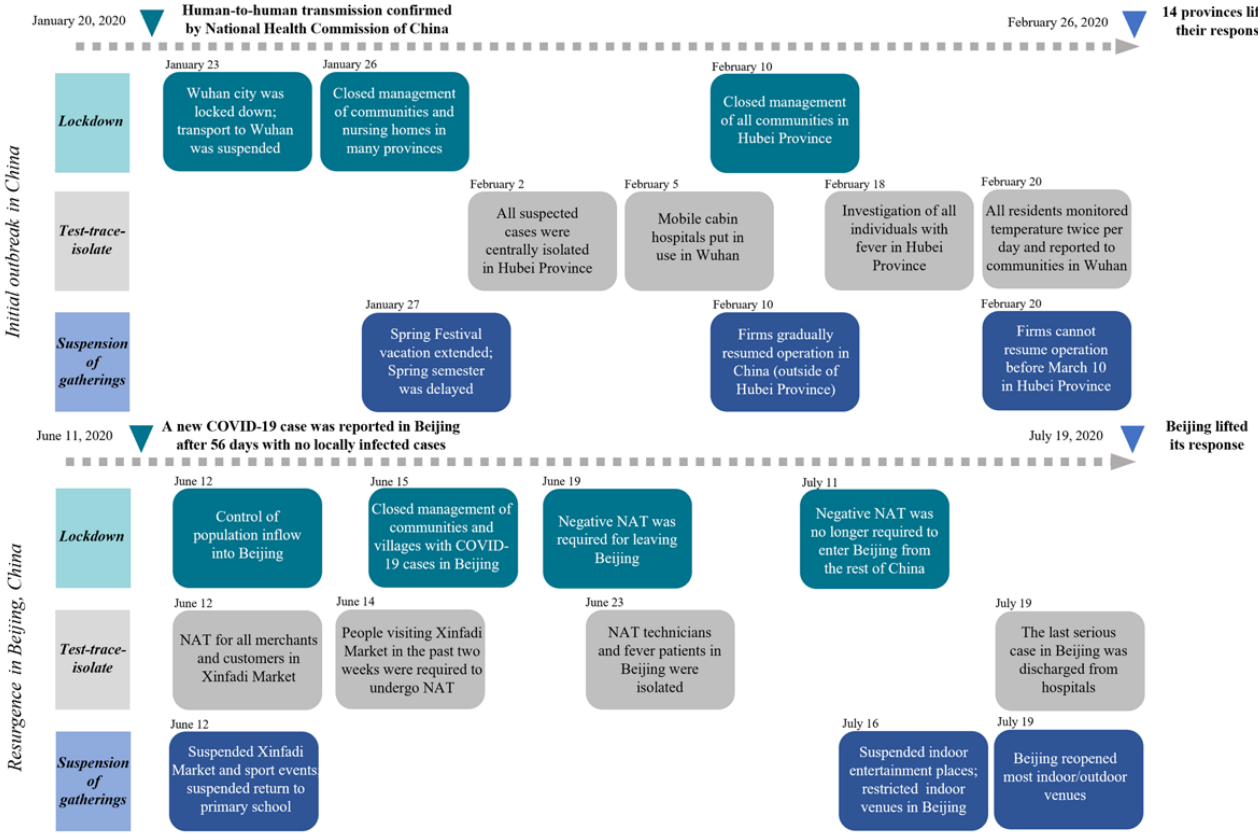
Since LDA is an unsupervised text classification algorithm based on the "bag-of-words model" [38], it may sometimes misclassify documents or misidentify topics [39-41]; therefore, it is important to manually assess representative documents (in our case, Weibo posts). In this study, LDA outputs was verified and improved by 2 independent researchers (YS and QW) to analyze key words generated by the LDA model and manually review sample posts for each topic. If the Weibo posts in one of the machine-generated 30 topics were found to contain several exclusive subtopics, that topic was manually reviewed. Finally, a random sample of Weibo posts (>10%) and their assigned topics were reviewed by 2 independent researchers (YS and QW) for quality control.

Results

Findings Overview

We compared the public sentiment to 3 containment measures during the initial COVID-19 outbreak in China and resurgence in Beijing (Figure 1): lockdown, the test-trace-isolate strategy, and suspension of gatherings. Generally, these measures evolved from a complete lockdown for the general population during the initial outbreak to a more targeted response strategy for high-risk populations during resurgence.

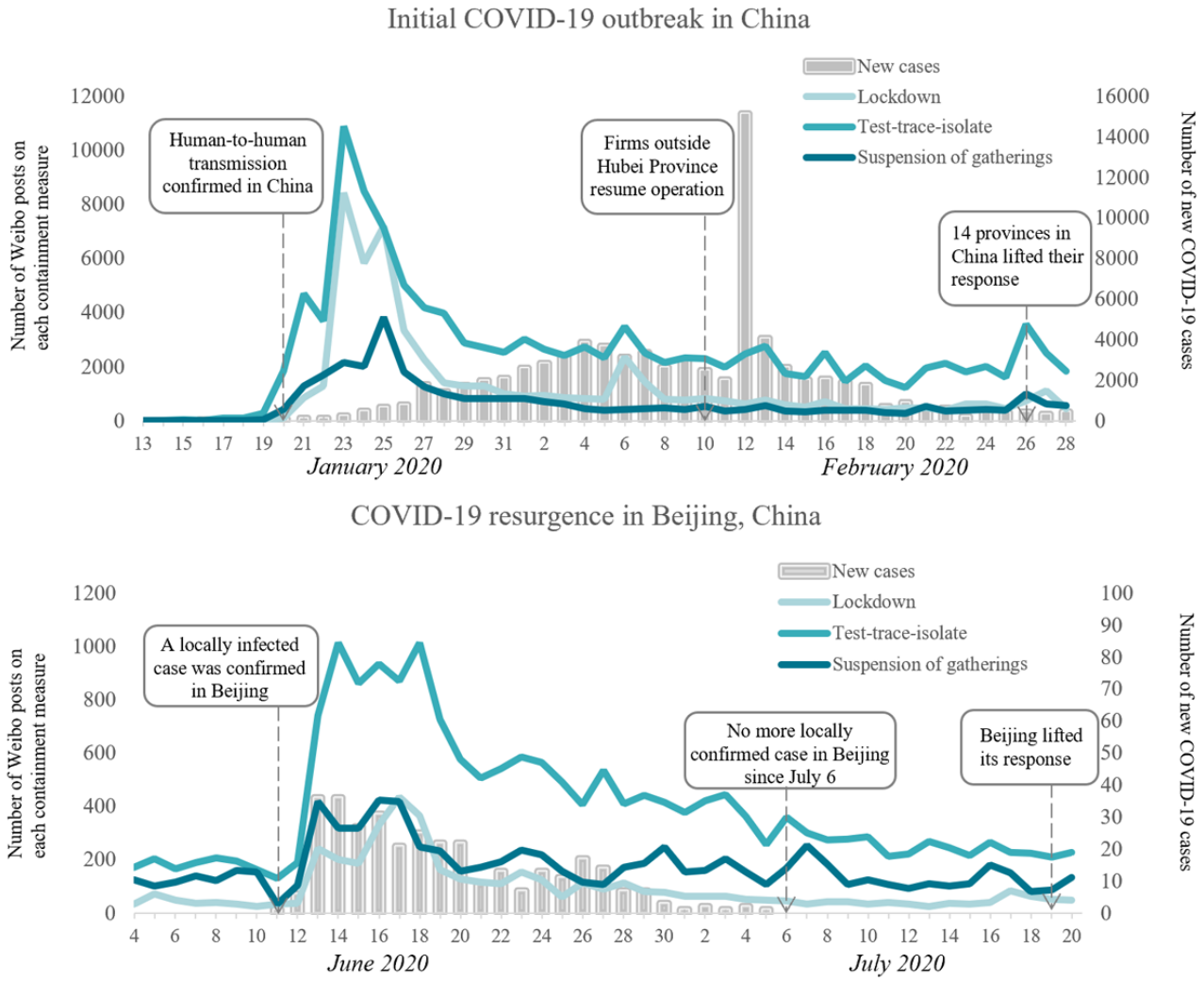
Figure 1. Timeline of containment measures during the initial outbreak and the resurgence of COVID-19 in China. Containment measures relevant to the lockdown, the test-trace-isolate measure, and suspension of gatherings are shown in chronological order. NAT: nucleic acid testing.



Public Engagement With Containment Measures

The COVID-19 pandemic received marked engagement from the public nationwide. During both the initial outbreak and resurgence in China, test-trace-isolate measures had the most engagement from the public, and suspension of gatherings had relatively low engagement; one difference between these 2 periods is that lockdown measures had high engagement during the initial outbreak but the lowest engagement during resurgence (Figure 2). In both periods, the number of Weibo posts related to containment measures increased drastically after outbreak announcement and the implementation of containment measures, remained high for approximately 1-2 weeks, and then plummeted. From 1 week before to 1 week after the confirmation of human-to-human transmission (on January 20, 2020) in China, the average daily number of Weibo posts increased from 14 to 4168 for the lockdown measure, 69 to 2003 for the suspension of gatherings, and 340 to 6298 for the test-trace-isolate measure. Similarly, the average daily number of Weibo posts during the week before resurgence (on June 11, 2020) were 41 for the lockdown measure, 118 for the suspension of gatherings, and 179 for the test-trace-isolate measure. In comparison, this number approached 256 for the lockdown measure, 321 for the suspension of gatherings, and 803 for the test-trace-isolate measure in the week after resurgence. Furthermore, Weibo activity peaked approximately 1-2 weeks before the peak in COVID-19 cases, which indicated that Weibo seemed to track policy changes (eg, lockdown) rather than the actual number of cases.

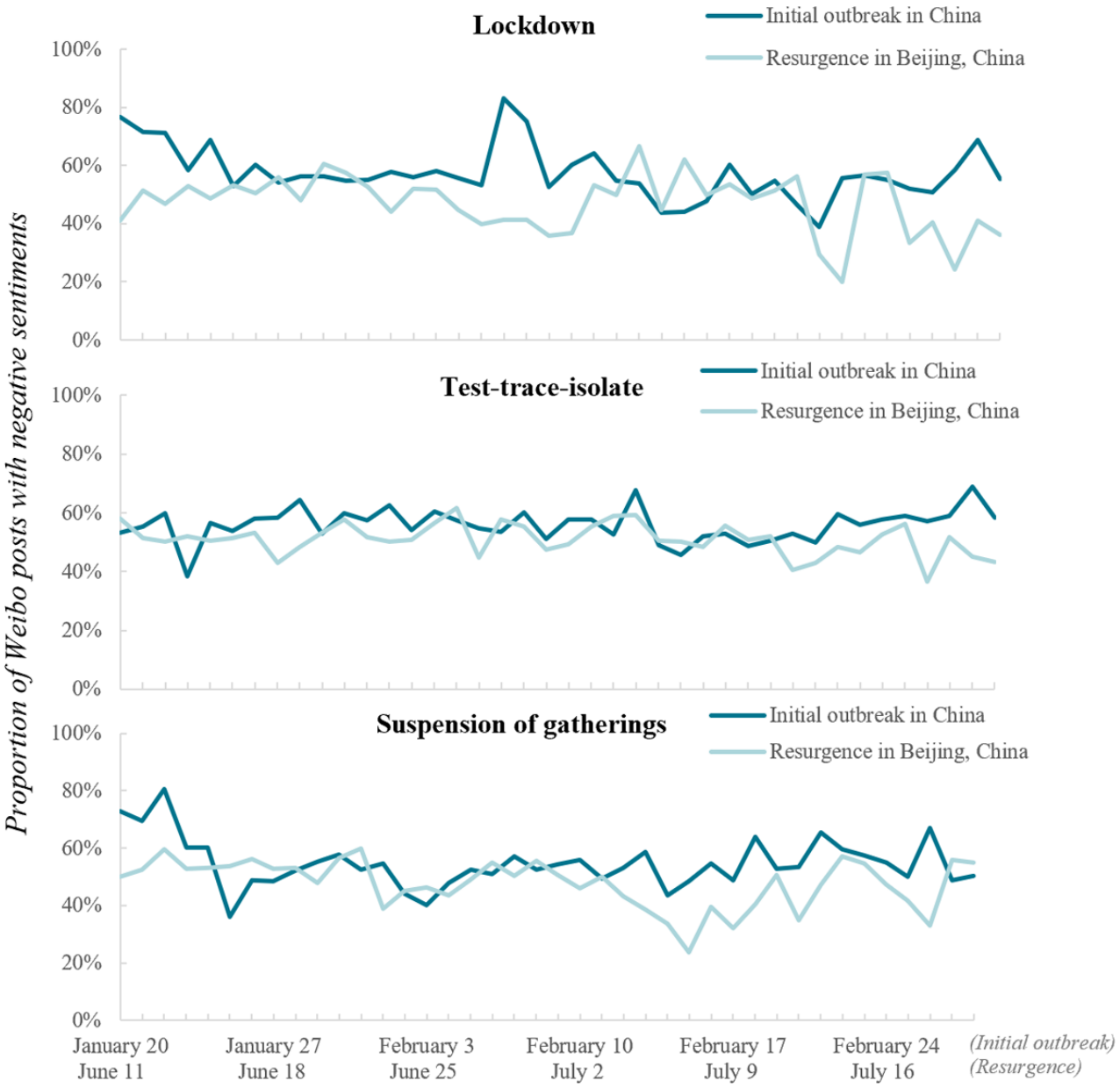
Figure 2. Number of Weibo posts on containment measures and new COVID-19 cases during the initial outbreak and the resurgence of COVID-19 in China. The lines show the daily number of Weibo posts on the lockdown, test-trace-isolate measure, and the suspension of gatherings. The bars show the daily number of new confirmed COVID-19 cases in China and Beijing.



Public Sentiment Toward Containment Measures

The overall negative sentiment toward all 3 containment measures during resurgence (average daily proportion of Weibo posts with a negative sentiment: 47% for the lockdown, 51% for the test-trace-isolate measure, and 48% for the suspension of gatherings) was lower than that during the initial outbreak (average daily proportion: 57% for the lockdown, 56% for the test-trace-isolate measure, and 55% for the suspension of gatherings) (Figure 3). During the initial outbreak, approximately 80% of the public immediately expressed negative sentiments toward the lockdown and suspension of gatherings after a lockdown was imposed in Wuhan, and approximately 60% of people expressed negative sentiments toward the test-trace-isolate measure. Thereafter, negative sentiments rapidly started to decrease for approximately 1 week and fluctuated surrounding a low level. However, for the test-trace-isolate measure, the proportion of negative sentiments varied through a smaller range than that for the other 2 containment measures. One exception is the spike in negative sentiments toward the lockdown on February 6, 2020, which might be related to concerns regarding the lack of medications or treatments for patients with other diseases; this accounted for a large proportion of posts expressing a negative sentiment on that day. During COVID-19 resurgence in Beijing, the negative sentiment toward containment measures was relatively lower and displayed lesser variation than that during the initial outbreak.

Figure 3. Proportion of Weibo posts with negative sentiments on containment measures during the initial outbreak and the resurgence of COVID-19 in China.



Frequently Discussed Topics Related to Containment Measures

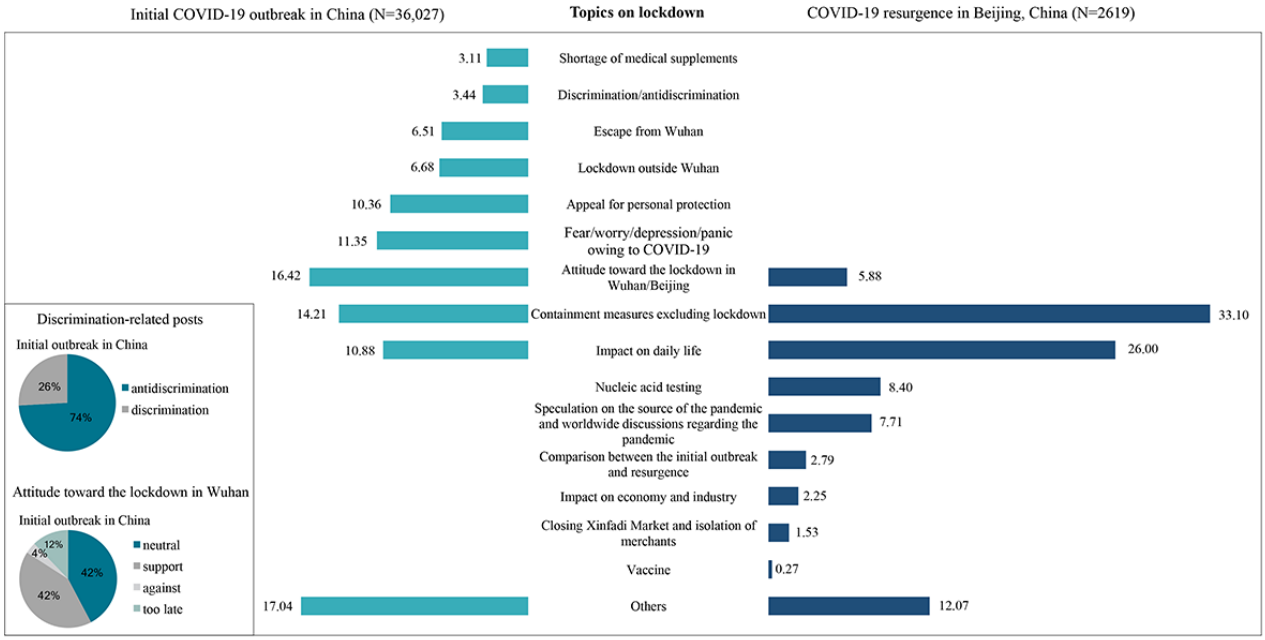
Weibo posts during the first two weeks of the initial outbreak and COVID-19 resurgence were used to identify frequently discussed topics on each containment measure.

Topics Related to the Lockdown Measure

For the initial outbreak, the top 3 topics related to the lockdown measure were attitudes towards the lockdown in Wuhan (n=5915/36,027, 16.42%), discussions on containment measures other than the lockdown (n=5120/36,027, 14.21%), and expressions of negative sentiments owing to the pandemic, such as fear, worry, depression, and panic (n=4090/36,027, 11.35%) (Figure 4). Among posts related to attitudes toward the lockdown in Wuhan, 2473 of 5915 (42%) posts expressed support for the policy, while 220 (4%) expressed opposition toward the policy, along with 720 (12%) posts that indicated that the lockdown should have been implemented earlier. Regarding resurgence in Beijing, the leading 3 topics were public containment measures (n=867/2619, 33.10%), impacts on daily life (n=681/2619, n=26.00%), and nucleic acid tests for COVID-19 (n=220/2619, 8.40%). Since more targeted responses had been used during the resurgence, containment measures excluding lockdown became the main focus of Weibo posts, whereas posts related to the lockdown decreased from 16.42% to 5.88%. After the resumption of work, people paid more attention to how containment measures would affect their daily lives or work, and thus the proportion of relevant posts increased from 10.88% to 26.00%. Consequently, there were discussions about the impact of containment measures on the economy and industry. Noticeably, people expressed more biases by ascribing the emergence and spread of the pandemic

to Wuhan residents during the initial outbreak; in contrast, this was barely mentioned during resurgence in Beijing. Other topics during the initial outbreak included an appeal for personal protection (10.36%), lockdowns outside Wuhan (6.68%), people leaving Wuhan (6.51%); shortages in the supply of medical equipment, including masks, were not widely discussed (3.11%). Topics such as vaccines and nucleic acid tests only emerged during resurgence.

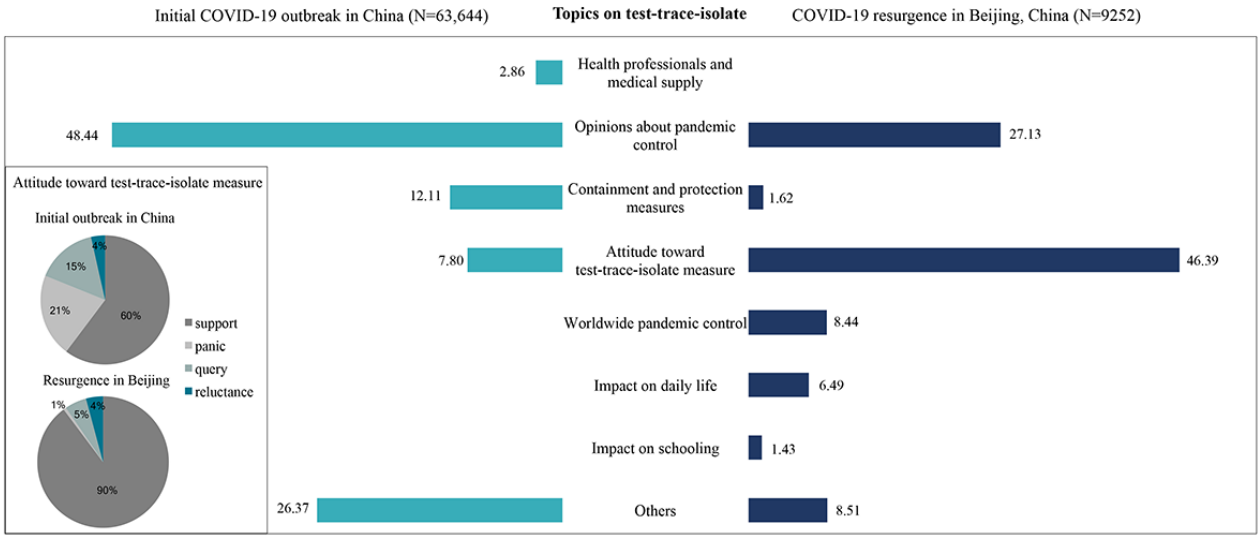
Figure 4. Proportion of posts related to frequently discussed topics on the lockdown measure during the initial outbreak and the resurgence of COVID-19 in China. The columns show the proportion of posts on each topic among all posts related to the lockdown measure. Pie charts show the proportion of various attitudes toward some topics during the initial outbreak.



Topics Related to the Test-Trace-Isolate Measure

During the initial outbreak, 30,826 of 63,644 (48.44%) posts related to the test-trace-isolate measure were opinions on pandemic control, which expressed hope or trust in rapid control of the pandemic and contained suggestions about specific containment measures and descriptions of the status of its spread (Figure 5). During resurgence, the percentage of posts expressing opinions on pandemic control decreased to 27.13% owing to public confidence in effective response measures. The second-most popular topic during the initial outbreak was quarantine and protective measures, which were only indicated in 1.62% of posts during resurgence. In the initial outbreak, 7.8% of posts expressed public attitudes toward the test-trace-isolate measure, while during resurgence, 46.39% of users expressed their attitudes. This further illustrates how the discussion's focus shifted from the status of pandemic control to targeted response measures. Interestingly, public attitudes toward the test-trace-isolate measure differed between these 2 periods of the pandemic in China. In the initial nationwide outbreak, only 3921 of 6506 (60%) posts supported the test-trace-isolate strategy. However, during resurgence in Beijing, the proportion of posts supporting the test-trace-isolate policies increased to 90% (n=3853/4292). Posts indicating panic decreased from 21% to 1%, and those related to queries decreased from 15% to 5%. Users expressing reluctance to be tested or quarantined accounted for 4% of posts during these 2 periods. Furthermore, different topics emerged during these 2 periods. During the initial outbreak, some posts (2.86%) discussed medical equipment and health professionals, while during resurgence, people discussed the impact of the test-trace-isolate measure on schooling, work, daily lives, and the status of the pandemic worldwide.

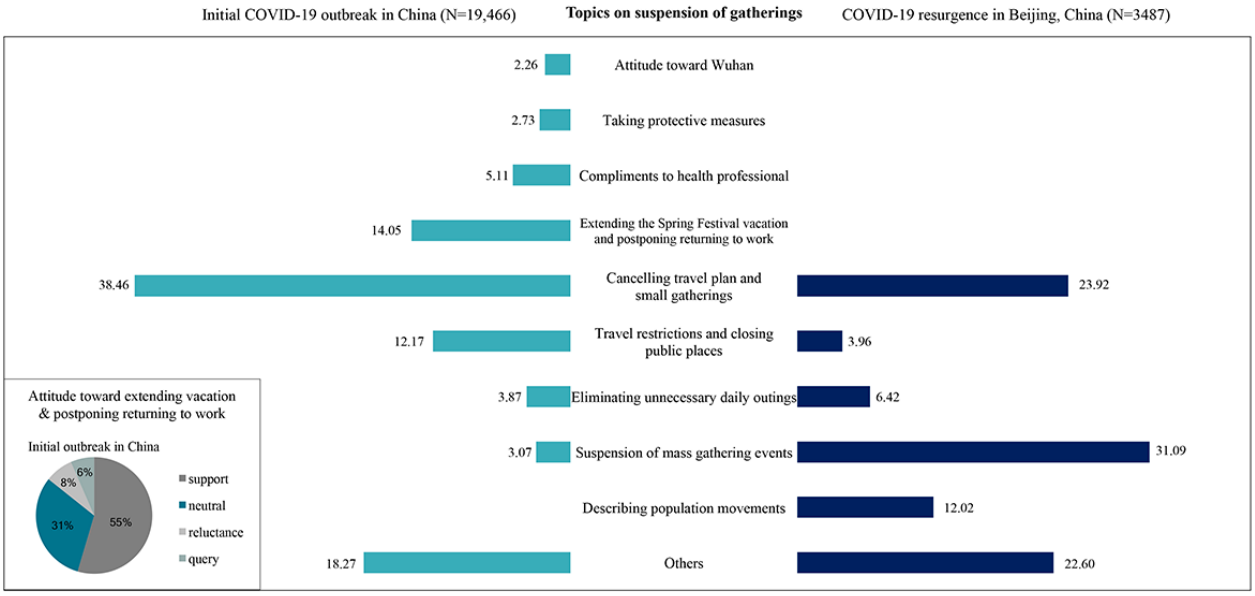
Figure 5. Proportion of posts related to frequently discussed topics on the test-trace-isolate measure during the initial outbreak and the resurgence of COVID-19 in China. The columns show the proportion of posts on each topic among all posts related to the test-trace-isolate measure. Pie charts show the proportion of various attitudes toward the test-trace-isolate measure between the initial outbreak and resurgence.



Topics Related to the Suspension of Gatherings

Among posts on the suspension of gatherings, 7487 of 19,466 (38.46%) and 834 of 3487 (23.92%) posts indicated the cancellation of travel plans and small-scale gatherings during the initial outbreak and resurgence, respectively (Figure 6). The proportion of posts describing the suspension of mass gathering events increased drastically from 3.07% during the initial outbreak to 31.09% during resurgence, which indicates that the public response was much more sensitive to the suspension of mass gathering events such as sports and concerts, during resurgence. Other common topics during these 2 periods included travel restrictions, closure of public places, and elimination of unnecessary daily outings. In addition, 2735 of 19,466 (14.05%) posts discussed the extension of the Spring Festival holiday and postponing of the return to work during the initial outbreak, among which, 55% of users supported the extension of the holiday. During resurgence, discussions on issues regarding population movement emerged, such as visas and immigration.

Figure 6. Proportion of posts related to frequently discussed topics on the suspension of gatherings during the initial outbreak and the resurgence of COVID-19 in China. The columns show the proportion of posts on each topic among all posts related to the suspension of gatherings. Pie charts show the proportion of various attitudes toward some topics during the initial outbreak.



Discussion

Principal Findings

Through social media surveillance, we analyzed and compared the level of public engagement, sentiment expressed, and frequently discussed topics related to 3 primary containment measures during the initial outbreak and the resurgence of COVID-19 in China. As the pandemic control strategy switched from lockdown to a targeted response during resurgence, public responses became more rational. During the initial outbreak,

the number of Weibo posts escalated drastically after the lockdown in Wuhan, a high proportion of posts indicated a negative public sentiment toward all 3 primary containment measures, and topics related to discrimination emerged. During resurgence, however, the public showed less engagement and less negative sentiments, were more supportive towards containment measures, and shifted their focus to the impacts of containment measures on their daily lives or their work.

Monitoring of public responses for pandemic control strategies is crucial for obtaining rapid feedback and informing strategy adjustments during a pandemic. Previous studies have reported that social mobilization and community engagement were central to the Ebola response in West Africa [42-44]. During the COVID-19 pandemic, especially during the initial outbreak, Weibo users in China raised questions regarding the pandemic's source, shortage of medical equipment, and discrimination against Wuhan residents. A considerable number of Weibo posts expressed panic or queries regarding containment measures. This is worth noting, since previous studies have reported that panic and queries regarding containment measures and discrimination may inhibit community engagement and dampen pandemic control [19]. During resurgence, public concerns regarding containment measures changed to focus on their impact on daily life. Policymakers should pay close attention to these changes in the public response to track the most vital needs of the population during the different stages of the pandemic and to address public concerns through timely and effective communication [45,46]. This can improve public compliance and engagement with containment measures and facilitate their implementation [47,48].

During COVID-19 resurgence in Beijing, a more targeted response strategy was applied for the high-risk population, and the general population was less impacted. Frequently discussed topics on Weibo during resurgence suggested that the lives of the public were returning to normal; there were no longer topics regarding discrimination and the shortage of supplies, and the overall negative sentiment toward all containment measures was lower during resurgence than during the initial outbreak. The shifts in discussed topics and the public sentiment might also be related to the reduced stringency of these containment measures. Concurrent with previous reports, stricter measures were followed by a more marked negative sentiment [16]. The public may become fatigued to containment measures when the pandemic resurges. In our study, public responses were more sensitive to the suspension of mass gathering events during resurgence than during the initial outbreak. Therefore, governments should balance the benefits of containment strategies to their impacts on daily life and the economy to formulate tailored response strategies during COVID-19 resurgence.

As more information on COVID-19 and its control has become available, governments have found alternatives to tailor their response strategies to its resurgence. Using the test-trace-isolate measure for high-risk populations, Beijing flattened the contagion curve and gained control over the pandemic within 1 month. Our study indicates that this targeted response strategy better matches public concerns. With the normalization of the pandemic, people cared more about their life and the economy. during the pandemic, such as the return to school or work, going to concerts, traveling, immigration, and sports events. The targeted response strategy focuses on high-risk populations and has less restrictions on most people's lives, which might help reduce the negative public sentiment during resurgence. A modeling study predicted that, compared with the community-wide lockdown strategy, the targeted response strategies would be less costly [49] and could help revive industries and the economy. China's economy data revealed that its gross domestic product decreased by 6.8% in the first quarter of 2020 during the complete lockdown but increased by 3.2% and 4.9% in the second and third quarters of 2020, respectively, with the test-trace-isolate measure targeting high-risk populations [50]. Therefore, more targeted response strategies should be promoted during subsequent COVID-19 resurgence.

Overall, the public reaction to the initial outbreak indicated a drastic escalation in public engagement, as evident from the increased number of negative sentiments after implementation of the lockdown, and, in comparison, targeted containment measures during resurgence gained more rational responses and greater support. The public expressed a high level of panic (21%) during the initial outbreak but virtually no panic (1%) during resurgence. Only 4% of the public expressed reluctance to be tested or quarantined during both the initial outbreak and resurgence, which is an important consideration for public health efforts to achieve universal compliance. The targeted test-trace-isolate measure received the greatest support among all 3 containment measures during the initial outbreak, and its support rate approached approximately 90% during resurgence. The evolution of public responses toward containment measures indicated that targeted test-trace-isolate strategies were more acceptable to the public. Governments should take public responses on social media into consideration, and develop more targeted test-trace-isolate strategies to prepare for the future resurgence of COVID-19.

Limitations

This study has several limitations. First, Weibo is more widely used by younger people, and some users simply read Weibo posts but do not post their own content. Therefore, caution should be exercised when interpreting our findings in the context of the general public. We used Python to curate all relevant Weibo posts for analysis; therefore, our data are valid enough to reflect the opinions and responses of the Weibo users. Second, as an unsupervised text classification algorithm based on the "bag-of-words model" [38], LDA may lead to misclassification of Weibo posts. However, such misclassification is random; that is, it does not specify a certain direction (positive or negative) [51]. Therefore, it does not cause an interpretation bias in a comparative study. To control for this potential bias, we randomly sampled and manually classified Weibo posts to rectify for misclassification and improve accuracy. Future studies should perform real-time social media surveillance with more advanced machine learning techniques (eg, bidirectional encoder representations from transformers), and conduct long-term, multilingual, and multiplatform public response surveillance for the COVID-19 pandemic.

Conclusions

Compared to the initial outbreak, the public became more rational during COVID-19 resurgence. The public expressed less engagement and less negative sentiment on containment measures, were more supportive towards containment measures, and shifted their focus to the impact of containment measures on their daily life or work during resurgence. Targeted test-trace-isolate strategies were more acceptable to the public. This study indicates that pandemic control strategies should be more targeted during subsequent COVID-19 resurgence, such as test-trace-isolate strategies targeting high-risk populations, to balance pandemic control and its impact on daily life and the economy.

Acknowledgments

We are thankful to Linyao Lu, Sihong Zhao, Jialu Pu, Yixin Tong, Longfei Feng, Xiaoli Wang, Qian Lv, Fanxing Du, Simeng Hu, Tianye Wang, Yixuan Zhang, and Shuxian He from Fudan University for their assistance with data analysis. ZH acknowledges support from the National Natural Science Foundation of China (no 71874034), the National Key R&D Program of China (no 2018YFC1312600 and 2018YFC1312604), and the National Institute for Health Research (EPIDZL9012). LL received support from the UK Foreign, Commonwealth and Development Office and the Wellcome Trust (grant# 215373/Z/19/Z). The study funders had no role in the study design; data collection, analysis, and interpretation; or drafting of the manuscript.

Authors' Contributions

ZH and LL conceptualized and designed the study. Xinyu Zhou collected the data. YS, Xinyu Zhou, HJ, QW, ZQ, and Xiaoyu Zhou analyzed the data. Xinyu Zhou and YS drafted the manuscript. ZH, LL, and MJ revised the manuscript for important intellectual content. ZH had full access to all the study data and the final responsibility for the decision to submit the manuscript for publication. All authors approved the final manuscript for submission and were accountable for all aspects of the study.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Flowchart of data collection and preprocessing.
[PNG File , 187 KB - jmir_v23i4e26518_app1.png]

Multimedia Appendix 2
LDA outputs for each containment measure.
[DOCX File , 90 KB - jmir_v23i4e26518_app2.docx]

References

1. Chen N, Zhou M, Dong X, Qu J, Gong F, Han Y, et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. Lancet 2020 Feb 15;395(10223):507-513 [FREE Full text] [doi: 10.1016/S0140-6736(20)30211-7] [Medline: 32007143]
2. Lee V, Chiew C, Khong W. Interrupting transmission of COVID-19: lessons from containment efforts in Singapore. J Travel Med 2020 May 18;27(3):taaa039 [FREE Full text] [doi: 10.1093/jtm/taaa039] [Medline: 32167146]
3. Iacobucci G. Covid-19: UK lockdown is "crucial" to saving lives, say doctors and scientists. BMJ 2020 Mar 24;368:m1204. [doi: 10.1136/bmj.m1204] [Medline: 32209548]
4. Chinazzi M, Davis JT, Ajelli M, Gioannini C, Litvinova M, Merler S, et al. The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. Science 2020 Apr 24;368(6489):395-400 [FREE Full text] [doi: 10.1126/science.aba9757] [Medline: 32144116]
5. Sjödin H, Wilder-Smith A, Osman S, Farooq Z, Rocklöv J. Only strict quarantine measures can curb the coronavirus disease (COVID-19) outbreak in Italy, 2020. Euro Surveill 2020 Apr;25(13):2000280 [FREE Full text] [doi: 10.2807/1560-7917.ES.2020.25.13.2000280] [Medline: 32265005]
6. Wise J. Covid-19: Risk of second wave is very real, say researchers. BMJ 2020 Jun 09;369:m2294. [doi: 10.1136/bmj.m2294] [Medline: 32518177]
7. Hou Z, Du F, Zhou X, Jiang H, Martin S, Larson H, et al. Cross-Country Comparison of Public Awareness, Rumors, and Behavioral Responses to the COVID-19 Epidemic: Infodemiology Study. J Med Internet Res 2020 Aug 03;22(8):e21143 [FREE Full text] [doi: 10.2196/21143] [Medline: 32701460]
8. Wu Z, Wang Q, Zhao J, Yang P, McGoogan JM, Feng Z, et al. Time Course of a Second Outbreak of COVID-19 in Beijing, China, June-July 2020. JAMA 2020 Aug 24:1458-1459 [FREE Full text] [doi: 10.1001/jama.2020.15894] [Medline: 32852518]

9. Han X, Wang J, Zhang M, Wang X. Using Social Media to Mine and Analyze Public Opinion Related to COVID-19 in China. Int J Environ Res Public Health 2020 Apr 17;17(8):2788 [FREE Full text] [doi: 10.3390/ijerph17082788] [Medline: 32316647]
10. Owen J. Covid-19: WHO raises concerns about new cases in Beijing. BMJ 2020 Jun 16;369:m2415. [doi: 10.1136/bmj.m2415] [Medline: 32546470]
11. Increased transmission of COVID-19 in the EU/EEA and the UK – twelfth update. European Centre for Disease Prevention and Control. 2020 Sep 24. URL: https://www.ecdc.europa.eu/sites/default/files/documents/covid-19-risk-assessment-increased-transmission-12th-update-september-2020.pdf [accessed 2021-03-25]
12. Cacciapaglia G, Cot C, Sannino F. Second wave COVID-19 pandemics in Europe: a temporal playbook. Sci Rep 2020 Sep 23;10(1):15514 [FREE Full text] [doi: 10.1038/s41598-020-72611-5] [Medline: 32968181]
13. Mahase E. Covid-19: Medical leaders call for rapid review to prepare for second wave. BMJ 2020 Jun 24;369:m2529. [doi: 10.1136/bmj.m2529] [Medline: 32580991]
14. Murillo-Llorente MT, Perez-Bermejo M. COVID-19: Social Irresponsibility of Teenagers Towards the Second Wave in Spain. J Epidemiol 2020 Oct 05;30(10):483 [FREE Full text] [doi: 10.2188/jea.JE20200360] [Medline: 32779631]
15. Hsu L, Chia P, Vasoo S. A midpoint perspective on the COVID-19 pandemic. Singapore Med J 2020 Jul;61(7):381-383 [FREE Full text] [doi: 10.11622/smedj.2020036] [Medline: 32211911]
16. Doogan C, Buntine W, Linger H, Brunt S. Public Perceptions and Attitudes Toward COVID-19 Nonpharmaceutical Interventions Across Six Countries: A Topic Modeling Analysis of Twitter Data. J Med Internet Res 2020 Sep 03;22(9):e21419 [FREE Full text] [doi: 10.2196/21419] [Medline: 32784190]
17. Abd-Alrazaq A, Alhuwail D, Househ M, Hamdi M, Shah Z. Top Concerns of Tweeters During the COVID-19 Pandemic: Infoveillance Study. J Med Internet Res 2020 Apr 21;22(4):e19016 [FREE Full text] [doi: 10.2196/19016] [Medline: 32287039]
18. Xue J, Chen J, Hu R, Chen C, Zheng C, Su Y, et al. Twitter Discussions and Emotions About the COVID-19 Pandemic: Machine Learning Approach. J Med Internet Res 2020 Nov 25;22(11):e20550 [FREE Full text] [doi: 10.2196/20550] [Medline: 33119535]
19. Seale H, Dyer CEF, Abdi I, Rahman KM, Sun Y, Qureshi MO, et al. Improving the impact of non-pharmaceutical interventions during COVID-19: examining the factors that influence engagement and the impact on individuals. BMC Infect Dis 2020 Aug 17;20(1):607 [FREE Full text] [doi: 10.1186/s12879-020-05340-9] [Medline: 32807087]
20. Li S, Wang Y, Xue J, Zhao N, Zhu T. The Impact of COVID-19 Epidemic Declaration on Psychological Consequences: A Study on Active Weibo Users. Int J Environ Res Public Health 2020 Mar 19;17(6):2032 [FREE Full text] [doi: 10.3390/ijerph17062032] [Medline: 32204411]
21. Boon-Itt S, Skunkan Y. Public Perception of the COVID-19 Pandemic on Twitter: Sentiment Analysis and Topic Modeling Study. JMIR Public Health Surveill 2020 Nov 11;6(4):e21978 [FREE Full text] [doi: 10.2196/21978] [Medline: 33108310]
22. Zhao Y, Cheng S, Yu X, Xu H. Chinese Public's Attention to the COVID-19 Epidemic on Social Media: Observational Descriptive Study. J Med Internet Res 2020 May 04;22(5):e18825 [FREE Full text] [doi: 10.2196/18825] [Medline: 32314976]
23. Chehal D, Gupta P, Gulati P. COVID-19 pandemic lockdown: An emotional health perspective of Indians on Twitter. Int J Soc Psychiatry 2020 Jul 07:20764020940741 [FREE Full text] [doi: 10.1177/0020764020940741] [Medline: 32633185]
24. Pastor CK. Sentiment Analysis of Filipinos and Effects of Extreme Community Quarantine Due to Coronavirus (COVID-19) Pandemic. SSRN Journal 2020. [doi: 10.2139/ssrn.3574385]
25. Pandemic fatigue: Reinvigorating the public to prevent COVID-19. World Health Organization: Regional Office for Europe. 2020. URL: https://apps.who.int/iris/bitstream/handle/10665/335820/WHO-EURO-2020-1160-40906-55390-eng.pdf [accessed 2021-10-08]
26. Hou Z, Du F, Jiang H, Zhou X, Lin L. Assessment of public attention, risk perception, emotional and behavioural responses to the COVID-19 outbreak: social media surveillance in China. medRxiv. Preprint posted online March 17, 2020. [doi: 10.1101/2020.03.14.20035956]
27. Tsao S, Chen H, Tisseverasinghe T, Yang Y, Li L, Butt ZA. What social media told us in the time of COVID-19: a scoping review. Lancet Digit Health 2021 Mar;3(3):e175-e194 [FREE Full text] [doi: 10.1016/S2589-7500(20)30315-0] [Medline: 33518503]
28. Barkur G, Vibha, Kamath GB. Sentiment analysis of nationwide lockdown due to COVID 19 outbreak: Evidence from India. Asian J Psychiatr 2020 Jun;51:102089 [FREE Full text] [doi: 10.1016/j.ajp.2020.102089] [Medline: 32305035]
29. Đogaš Z, Lušić Kalcina L, Pavlinac Dodig I, Demirović S, Madirazza K, Valić M, et al. The effect of COVID-19 lockdown on lifestyle and mood in Croatian general population: a cross-sectional study. Croat Med J 2020 Aug 31;61(4):309-318 [FREE Full text] [Medline: 32881428]
30. Graham MW, Avery EJ, Park S. The role of social media in local government crisis communications. Public Relations Review 2015 Sep;41(3):386-394. [doi: 10.1016/j.pubrev.2015.02.001] [Medline: 2015]
31. Weibo Reports Third Quarter 2020 Unaudited Financial Results. Weibo Corp. 2020 Dec 28. URL: https://weibocorporation.gcs-web.com/news-releases/news-release-details/weibo-reports-third-quarter-2020-unaudited-financial-results [accessed 2021-03-06]

32. Weibo COVID-19 Dataset. GitHub. URL: https://github.com/xinyuuzhou/weibo_covid-19_dataset [accessed 2021-03-26]
33. Effrosynidis D, Symeonidis S, Arampatzis A. A Comparison of Pre-processing Techniques for Twitter Sentiment Analysis. In: Research and Advanced Technology for Digital Libraries. 2017 Sep 02 Presented at: International Conference on Theory and Practice of Digital Libraries; 17-21 September 2017; Thessaloniki, Greece p. 394-406 URL: https://link.springer.com/chapter/10.1007/978-3-319-67008-9_31#citeas
34. Gupta I, Joshi N. Tweet normalization: A knowledge based approach. 2017 Presented at: 2017 International Conference on Infocom Technologies and Unmanned Systems; 18-20 December 2017; Dubai, United Arab Emirates. [doi: 10.1109/ictus.2017.8285996]
35. To H, Agrawal S, Kim S, Shahabi C. On Identifying Disaster-Related Tweets: Matching-Based or Learning-Based? 2017 Presented at: 2017 IEEE Third International Conference on Multimedia Big Data; 19-21 April 2017; Laguna Hills, CA. [doi: 10.1109/bigmm.2017.82]
36. Blei D, Ng A, Jordan M. Latent Dirichlet Allocation. J Mach Learn Res 2003;3:993-1022 [FREE Full text]
37. Luo R, Xu J, Zhang Y, Ren X, Sun X. PKUSEG: A Toolkit for multi-domain Chinese word segmentation. arXiv. Preprint posted online June 26, 2019. [FREE Full text]
38. Zhang Y, Jin R, Zhou Z. Understanding bag-of-words model: a statistical framework. Int J Mach Learn Cyber 2010 Aug 28;1(1-4):43-52. [doi: 10.1007/s13042-010-0001-0]
39. Debortoli S, Müller O, Junglas I, vom Brocke J. Text Mining for Information Systems Researchers: An Annotated Topic Modeling Tutorial. CAIS 2016;39:110-135. [doi: 10.17705/1cais.03907]
40. Hall D, Jurafsky D, Manning C. Studying the history of ideas using topic models. In: World Health Organization. 2008 07 Presented at: Conference on Empirical Methods in Natural Language Processing; 25–27 October 2008; Honolulu, HI p. -164. [doi: 10.3115/1613715.1613763]
41. Pablo ZC, Oco N, Roldan MDG, Cheng C, Roxas RE. Toward an enriched understanding of factors influencing Filipino behavior during elections through the analysis of Twitter data. Philippine Political Science Journal 2014 Nov 27;35(2):203-224. [doi: 10.1080/01154451.2014.964794]
42. Role of community engagement in situations of extensive community transmission of COVID-19. World Health Organization: Western Pacific Region. 2020 May 05. URL: https://www.who.int/publications/i/item/WPR-DSE-2020-016 [accessed 2021-03-25]
43. Marston C, Renedo A, Miles S. Community participation is crucial in a pandemic. Lancet 2020 May 30;395(10238):1676-1678 [FREE Full text] [doi: 10.1016/S0140-6736(20)31054-0] [Medline: 32380042]
44. Gillespie AM, Obregon R, El Asawi R, Richey C, Manoncourt E, Joshi K, et al. Social Mobilization and Community Engagement Central to the Ebola Response in West Africa: Lessons for Future Public Health Emergencies. Glob Health Sci Pract 2016 Dec 23;4(4):626-646 [FREE Full text] [doi: 10.9745/GHSP-D-16-00226] [Medline: 28031301]
45. Ngai CSB, Singh RG, Lu W, Koon AC. Grappling With the COVID-19 Health Crisis: Content Analysis of Communication Strategies and Their Effects on Public Engagement on Social Media. J Med Internet Res 2020 Aug 24;22(8):e21360 [FREE Full text] [doi: 10.2196/21360] [Medline: 32750013]
46. Liao Q, Yuan J, Dong M, Yang L, Fielding R, Lam WWT. Public Engagement and Government Responsiveness in the Communications About COVID-19 During the Early Epidemic Stage in China: Infodemiology Study on Social Media Data. J Med Internet Res 2020 May 26;22(5):e18796 [FREE Full text] [doi: 10.2196/18796] [Medline: 32412414]
47. Hu G, Qiu W. From guidance to practice: Promoting risk communication and community engagement for prevention and control of coronavirus disease (COVID-19) outbreak in China. J Evid Based Med 2020 May;13(2):168-172 [FREE Full text] [doi: 10.1111/jebm.12387] [Medline: 32445287]
48. Gan Y, Ma J, Wu J, Chen Y, Zhu H, Hall BJ. Immediate and delayed psychological effects of province-wide lockdown and personal quarantine during the COVID-19 outbreak in China. Psychol Med 2020 Aug 13:1-12 [FREE Full text] [doi: 10.1017/S0033291720003116] [Medline: 32787981]
49. Sturniolo S, Waites W, Colbourn T, Manheim D, Panovska-Griffiths J. Testing, tracing and isolation in compartmental models. medRxiv. Preprint posted online December 1, 2020. [FREE Full text] [doi: 10.1101/2020.05.14.20101808]
50. The National Economy Maintained the Momentum of Recovery in the First Two Months. National Bureau of Statistics of China. URL: http://www.stats.gov.cn/tjsj/zxfb/202010/t20201020_1794939.html [accessed 2021-10-26]
51. Wang A, Zhang J. Topic discovery method based on topic model combined with hierarchical clustering. 2020 Presented at: IEEE 5th Information Technology and Mechatronics Engineering Conference; 12-14 June 2020; Chongqing, China. [doi: 10.1109/itoec49072.2020.9141892]

Abbreviations

LDA: latent Dirichlet allocation

Edited by C Basch; submitted 15.12.20; peer-reviewed by MF Kabir, C Shelley, L Cilar; comments to author 01.03.21; revised version received 11.03.21; accepted 11.03.21; published 05.04.21.

Please cite as:
Zhou X, Song Y, Jiang H, Wang Q, Qu Z, Zhou X, Jit M, Hou Z, Lin L
Comparison of Public Responses to Containment Measures During the Initial Outbreak and Resurgence of COVID-19 in China: Infodemiology Study
J Med Internet Res 2021;23(4):e26518
URL: https://www.jmir.org/2021/4/e26518
doi:10.2196/26518
PMID:33750739

Original Paper

Artificial Intelligence–Enabled Analysis of Public Attitudes on Facebook and Twitter Toward COVID-19 Vaccines in the United Kingdom and the United States: Observational Study

Amir Hussain[1], PhD, BSc; Ahsen Tahir[1,2], PhD, BSc; Zain Hussain[3], BSc, MBChB; Zakariya Sheikh[3]; Mandar Gogate[1], BSc, PhD; Kia Dashtipour[1], BSc, PhD; Azhar Ali[4,5], BSc, MBA, MBBS; Aziz Sheikh[6], BSc, MBBS

[1]School of Computing, Edinburgh Napier University, Edinburgh, United Kingdom

[2]Department of Electrical Engineering, University of Engineering and Technology, Lahore, Pakistan

[3]Edinburgh Medical School, College of Medicine and Veterinary Medicine, University of Edinburgh, Edinburgh, United Kingdom

[4]NHS Forth Medical Group, Grangemouth, United Kingdom

[5]Harvard TH Chan School of Public Health, Harvard University, Boston, MA, United States

[6]Usher Institute, Edinburgh Medical School, University of Edinburgh, Edinburgh, United Kingdom

Corresponding Author:
Amir Hussain, PhD, BSc
School of Computing
Edinburgh Napier University
10 Colinton Road
Edinburgh, EH10 5DT
United Kingdom
Phone: 44 0845 260 6040
Email: a.hussain@napier.ac.uk

Abstract

Background: Global efforts toward the development and deployment of a vaccine for COVID-19 are rapidly advancing. To achieve herd immunity, widespread administration of vaccines is required, which necessitates significant cooperation from the general public. As such, it is crucial that governments and public health agencies understand public sentiments toward vaccines, which can help guide educational campaigns and other targeted policy interventions.

Objective: The aim of this study was to develop and apply an artificial intelligence–based approach to analyze public sentiments on social media in the United Kingdom and the United States toward COVID-19 vaccines to better understand the public attitude and concerns regarding COVID-19 vaccines.

Methods: Over 300,000 social media posts related to COVID-19 vaccines were extracted, including 23,571 Facebook posts from the United Kingdom and 144,864 from the United States, along with 40,268 tweets from the United Kingdom and 98,385 from the United States from March 1 to November 22, 2020. We used natural language processing and deep learning–based techniques to predict average sentiments, sentiment trends, and topics of discussion. These factors were analyzed longitudinally and geospatially, and manual reading of randomly selected posts on points of interest helped identify underlying themes and validated insights from the analysis.

Results: Overall averaged positive, negative, and neutral sentiments were at 58%, 22%, and 17% in the United Kingdom, compared to 56%, 24%, and 18% in the United States, respectively. Public optimism over vaccine development, effectiveness, and trials as well as concerns over their safety, economic viability, and corporation control were identified. We compared our findings to those of nationwide surveys in both countries and found them to correlate broadly.

Conclusions: Artificial intelligence–enabled social media analysis should be considered for adoption by institutions and governments alongside surveys and other conventional methods of assessing public attitude. Such analyses could enable real-time assessment, at scale, of public confidence and trust in COVID-19 vaccines, help address the concerns of vaccine sceptics, and help develop more effective policies and communication strategies to maximize uptake.

(J Med Internet Res 2021;23(4):e26627) doi:10.2196/26627

KEYWORDS

artificial intelligence; COVID-19; deep learning; Facebook; health informatics; natural language processing; public health; sentiment analysis; social media; Twitter; infodemiology; vaccination

Introduction

The imminent availability of COVID-19 vaccines poses a pressing need to continually monitor and better understand public sentiments in order to develop baseline levels of confidence in them among the general public and enable the identification of early warning signals of loss in confidence [1]. This will help address the concerns of vaccine sceptics [2-4] and develop the required public trust in immunization [5,6] to realize the goal of generating herd immunity [7].

Traditionally, governments use surveys to understand public attitude; however, these typically have limitations including small sample sizes, closed questions, and limited spatiotemporal granularity. In order to overcome these limitations, we argue that social media data can be used to obtain more, real-time insights into public sentiments and attitudes with considerable spatiotemporal granularity. Over half of the worldwide population, including approximately 70% the populations of the United Kingdom and the United States, are active social media users, and social media usage has significantly increased during the pandemic; for instance, Facebook usage increased by 37%. Since social media data are largely unstructured, they are amenable to the application of established artificial intelligence (AI) techniques such as machine learning, deep learning (DL) [8], and natural language processing (NLP) [9] to extract topics and sentiments from social media posts.

Sentiment analysis involves categorizing subjective opinions from text, audio, and video sources [9] to determine polarities (eg, positive, negative, and neutral), emotions (eg, anger, sadness, and happiness), or states of mind (eg, interest vs disinterest) toward target topics, themes, or aspects of interest [10]. A complementary approach, termed stance detection [11], assigns a stance label (favorable, against, and none) to a post on a specific predetermined target, which in itself may not be referred to or be the target of opinion in the post. Such approaches are currently underutilized in health care research. In particular, there is significant untapped potential in drawing on AI-enabled social media analysis to inform public policy research.

Methods

Ethics

Since the data analyzed in this study were completely in the public domain, no ethics review was necessary. We conducted a thorough assessment of the privacy risk that our study posed to individuals, in accordance with previous reports [12,13], to ensure compliance with relevant sections of the General Data Protection Regulation. We strived to comply with best practices for user protection [14,15], ensuring that nonpublic material is not included in our data set. Further, to comply with privacy laws and social network policies in accordance with the General Data Protection Regulation to collect data from Twitter [16] and Facebook CrowdTangle [17] platforms, we have not shared or published direct tweets or posts by individuals, quotes from individuals, or names or locations of users who are not public organizations or entities.

Data Sources

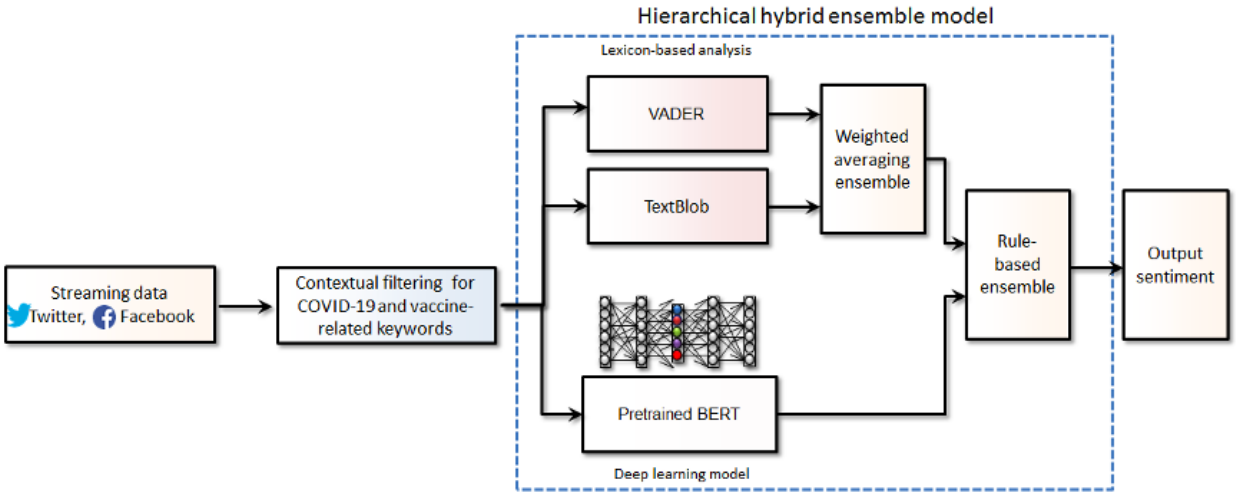
We used data from both Facebook and Twitter, two of the most popular and representative social media platforms [16]. We used Facebook posts and tweets that were posted in English in the United Kingdom and the United States from March 1 to November 22, 2020. Facebook posts were obtained through the CrowdTangle platform [17] and Twitter posts from a publicly available Twitter API. We used hydrated tweets from the global COVID-19 data set [18], which collects up to 4,400,000 tweets per day (including retweets) and up to 1,100,000 cleaned tweets without retweets. The total number of tweets hydrated and used in this study was >158,000,000. Facebook posts and tweets were thematically filtered for both COVID-19– and vaccine-related keywords and then geographically filtered for the United Kingdom and the United States. The first step in filtering COVID-19–related keywords involved widely used terms from the data set of Banda et al [18] (Multimedia Appendix 1). Vaccine-related terms used for second step filtering were selected by our team: “vaccine,” “vaccination,” “immunise,” “immunize,” “immunisation,” and “immunization.” A 2-step thematic filtering process was applied using these keywords before processing and analysis.

Analysis

The filtered data set was initially preprocessed (eg, removing links, hashtags, and stop words) and a new hierarchical hybrid ensemble–based AI model was developed for thematic sentiment analysis. This utilized an average weighting ensemble [19] of 2 lexicon-based methods: Valence Aware Dictionary for Sentiment Reasoning (VADER) [20] and TextBlob [21]. These were combined with a pretrained DL-based model, Bidirectional Encoder Representations from Transformers (BERT) [22], using a rule-based ensemble method (Figure 1).

A random 10% sample of Facebook posts and tweets was then manually annotated by the team and screened against our hybrid ensemble AI model’s sentiment classifications for refinement and validation. The hybrid ensemble model was optimized on the basis of the validation results, with sensitivity and specificity analysis revealing that the lexicon-based methods provided generally better accuracy for positive sentiments, and the BERT model generally provided better accuracy for neutral and negative sentiments, as illustrated using normalized confusion matrices (Multimedia Appendix 1).

Figure 1. Hierarchical hybrid-ensemble–based artificial intelligence model and data pipeline for thematic sentiment analysis. BERT: Bidirectional Encoder Representations from Transformers, VADER: Valence Aware Dictionary for Sentiment Reasoning.



VADER and TextBlob were combined though weighted averaging (VADERx0.45 + TextBlobx0.55), with TextBlob assigned a marginally higher weight of 0.55 owing to its performance. The weighted averaged output from the lexicons was combined with the output of the BERT model, using a final rule–based ensemble. The final output was combined through "If" and "Else" statements on the basis of the model's output sentiments.

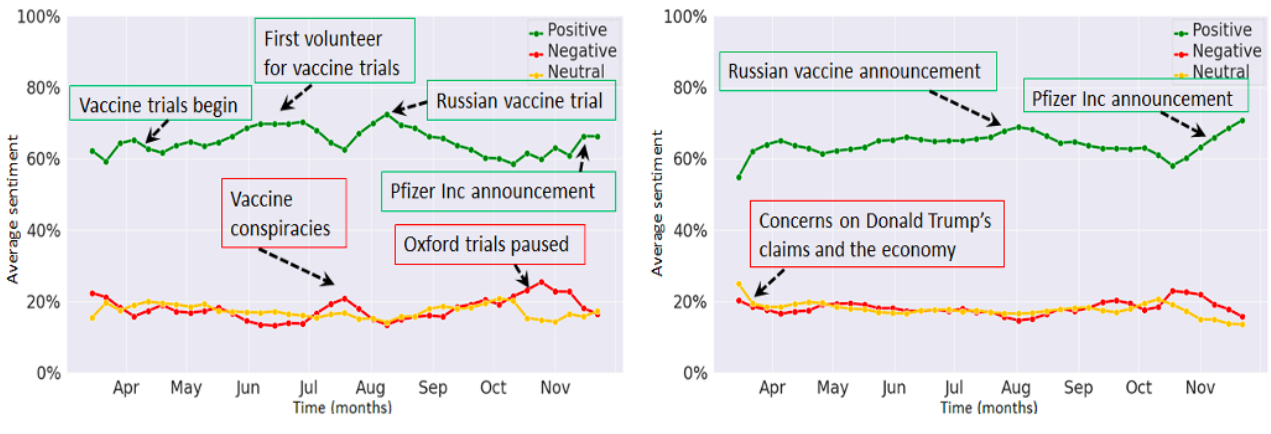
A number of established NLP techniques were used to analyze the processed data (Multimedia Appendix 1). Specifically, in addition to analyzing averaged sentiment trends and their geospatial mappings in the United Kingdom and the United States, we statistically analyzed the trends with Pearson correlation coefficient (r) and compared the findings with those of independent surveys. Sentiment word cloud and N-gram analyses were performed for specific periods of interest, around points of inflexion on sentiment trend graphs, to identify topics of discussion and to gain insight into the positive and negative content of online discourses. The analysis was also carried out over the entire study period to identify underlying themes and topics. Findings were validated, and further insights were obtained through manual reading of randomly selected posts around target points of interest by our team. Relevant social media data sets and outputs were anonymized, and statistical aggregates were made openly accessible for transparency and reproducibility (additionally through a publicly available dashboard [23]).

Results

Temporal Sentiment Trends

Monthly volume trends of the filtered Facebook posts and tweets in the United Kingdom and the United States for the target study period are shown in Multimedia Appendix 1. Figure 2 shows the averaged (weekly) positive, negative, and neutral Facebook sentiments in March-November 2020 in the United Kingdom and the United States. We identified topics of discussion on points of interest in the graphs. These are referred to in our descriptive analysis of the graphs below, and some are highlighted in Figures 2 and 3. It was interesting to note that the difference between the averaged positive and negative sentiment trends was more pronounced on Facebook than on Twitter.

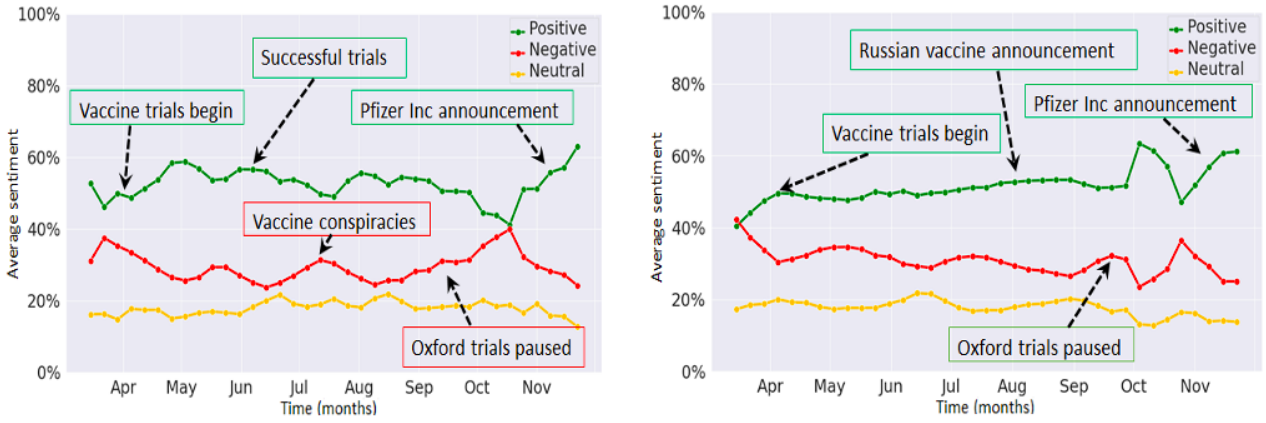
Figure 2. Averaged weekly trends in Facebook sentiments for (A) the United Kingdom and (B) the United States.



(A) The United Kingdom

(B) The United States

Figure 3. Average weekly trends in Twitter sentiments for (A) the United Kingdom and (B) the United States.



(A) The United Kingdom

(B) The United States

For the United Kingdom, positive sentiments on Facebook displayed the most prominent trend (Figure 2A), showing a steady increasing trend since May 2020, corresponding to the initiation of vaccine trials and the recruitment of the first trial volunteer. We observed a peak in mid-August 2020, which was potentially associated with news on vaccine development in the United Kingdom and Russia. Negative sentiments on Facebook displayed an inverse trend to that of positive sentiments, and discourse was centered around vaccine conspiracies and halting of trials. For the United States, positive sentiments on Facebook displayed the most prominent trend, showing a small peak in August 2020, which was associated with posts relating to research on the COVID-19 vaccine in Russia. Moreover, negative sentiments on Facebook displayed a slight increase in mid-September 2020, which was associated with posts relating to the accelerated development of the COVID-19 vaccine (Figure 2B). More recently, from mid-October 2020, the trend of positive sentiments on Facebook in the United Kingdom and the United states increased, partly because of announcements from Pfizer Inc and Moderna Inc on successful vaccine trials [24].

Figure 3 illustrates the positive, negative, and neutral sentiments on Twitter from March to November 2020 for the United Kingdom and the United States. Figure 3A shows that in the United Kingdom, positive sentiments on Twitter displayed the most prominent trend, showing a small peak at the end of April and July of 2020, the former related to the first human vaccine trial. The negative sentiment trend on Twitter displayed peaks in July and October of 2020, simultaneously with the United Kingdom opting out of the European Union vaccination scheme and halting of the phase III vaccine trials at the University of Oxford owing to safety concerns [25]. Figure 3B shows that in the United States, positive sentiments on Twitter displayed the most prominent trend, showing major peaks from end-September to end-November of 2020, which was related to claims by ex-President Donald Trump regarding a vaccine being ready in a few weeks and an increase in Twitter discourse due to his reference to the "herd mentality." We observed a small peak in the negative trend graph in mid-September 2020, which was related to halting of the phase III vaccine trial at the University of Oxford.

For both the United Kingdom and the United States, we observed a marked increase in the positive sentiment trend, since end-October 2020, which was related to recent breakthrough announcements by Pfizer Inc and Moderna Inc. Analysis of social media conversations indicated public optimism, with trial results being hailed as "good" and "amazing" and with "hope" prevailing for the "new year" (Multimedia Appendix 1). A notable peak in the negative sentiment trends for both countries, in approximately mid-October 2020, was associated with the growing antivaccination movement and with concerns regarding "fake news" and "misinformation."

Statistical Analysis of Sentiment Trends

Statistical analysis (results detailed in Multimedia Appendix 1) involved the assessment of the strength of the association between the predicted sentiment in the trend graph and the accuracy of the labeled data. Overall, regarding COVID-19 vaccines, we observed stronger sentiments on Twitter for the United States, with both positive and negative sentiments displaying stronger increasing and decreasing trends, respectively, compared to the United Kingdom. Public sentiments on Facebook reflected a reduction in positive sentiments and an increase in neutral sentiments in both the United Kingdom and the United States, with positive sentiments displaying a slightly stronger decreasing trend in the United Kingdom than in the United States.

Sentiment Word Clouds and Text N-Gram Analysis

We performed sentiment word cloud and text N-gram analyses for the entire study period to identify and analyze notable events that were of interest to social media users, and the findings are summarized in Multimedia Appendix 1 (some of these were also identified in the aforementioned analysis, on the sentiment trend graphs).

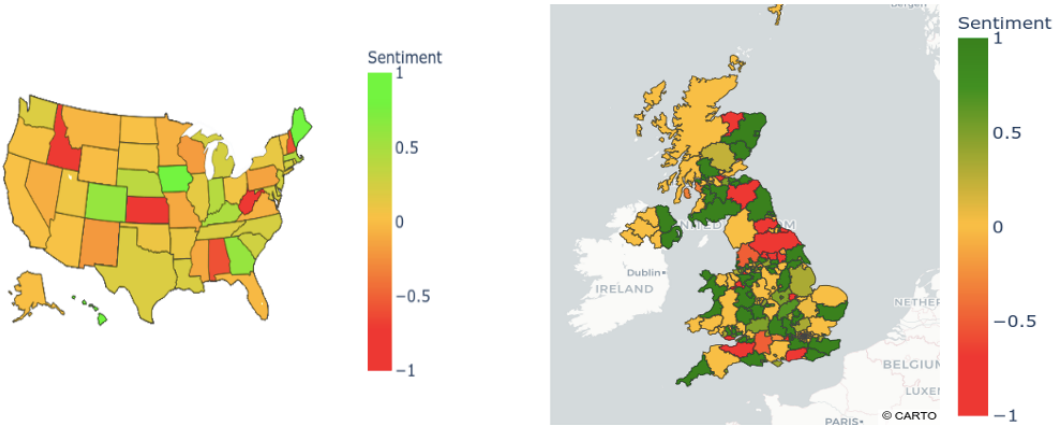
Geospatial Sentiment Analysis

A geospatial map of overall (averaged) sentiments at the state level in the United States is shown in Figure 4 (left), and it indicates that most states had a negative sentiment. The states with an overall negative sentiment toward COVID-19 vaccines were concentrated in the West and Midwest regions,

namely Idaho, Kansas, New Hampshire, West Virginia, and Alabama. The states with an overall positive sentiment were in the East, namely Maine, Colorado, Georgia, and Hawaii.

A geospatial map of averaged sentiments toward COVID-19 vaccines at the county level in the United Kingdom is shown in Figure 4 (right). In contrast with the United States, most counties in the United Kingdom had an overall positive sentiment toward COVID-19 vaccines. The counties with the most positive sentiments included Cornwall, Kent, East Sussex, Surrey, and Dorset in England and Aberdeenshire, Angus, and Stirlingshire in Scotland. Furthermore, the counties with the most negative sentiments were West Sussex, Somerset, North Yorkshire, and Durham in England.

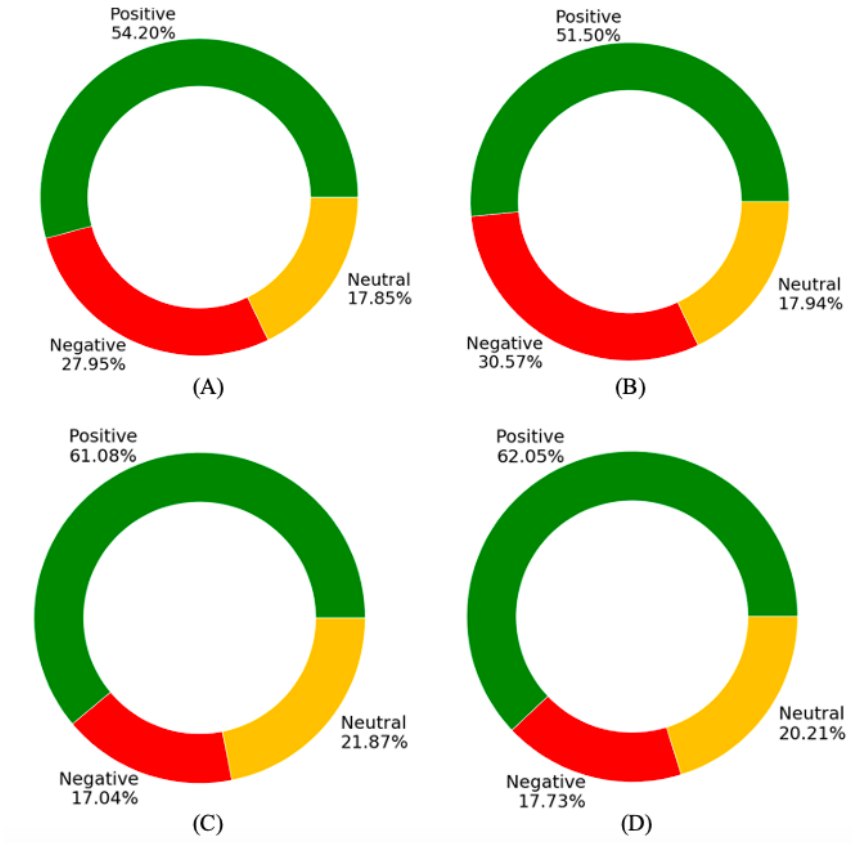
Figure 4. Geospatial mapping of averaged social media public sentiments in the United States (left) and the United Kingdom (right) toward COVID-19 vaccines (1 or green: positive sentiments, 0: neutral sentiments, and -1 or red: negative sentiments).



Overall Averaged Sentiments

Overall averaged sentiments in the United Kingdom and the United States on Facebook and Twitter are shown in Figure 5 and in Multimedia Appendix 1.

Figure 5. Overall averaged sentiments: (A) Twitter sentiments in the United Kingdom, (B) Twitter sentiments in the United States, (C) Facebook sentiments in the United Kingdom, and (D) Facebook sentiments in the United States.



Discussion

Principal Findings

We analyzed temporal variations in public sentiments toward COVID-19 vaccines in the United Kingdom and the United States. We identified, evaluated, and mapped the key events impacting positive, negative, and neutral sentiments to the temporal trends. We also mapped spatial variations in public sentiment to regions in the United Kingdom and states in the United States. Our geospatial maps can help identify areas with more negative sentiments toward COVID-19 vaccines, which can be further studied for potential interventions, to allay the underlying public fears and concerns.

Our findings indicate that online public discourse on Facebook and Twitter across the United Kingdom and the United States is evolving, with both complementary and contrasting insights obtained from the 2 popular platforms. Comparative analysis revealed that over the 9-month study period, averaged public sentiment toward COVID-19 vaccines has been mostly positive and similar in both the United Kingdom and the United States across both platforms (57.70% average across both platforms for the United Kingdom vs 56.80% for the United States). Positive sentiments were related to public opinions on vaccine development, related trials, and news related to vaccine availability.

On both platforms, overall averaged negative sentiments were found to be similar for the United Kingdom (22.50%) and the United States (24.10%). It is interesting to note that Twitter sentiments appeared more negatively biased, with the proportion of negative sentiments being almost 2-fold those on Facebook, for both the United Kingdom (27.95% vs 17.04%) and the United States (30.57% vs 17.73%), which potentially reflects their respective user demographics. This finding appears consistent with those of Waterloo et al [16], who reported that public opinions were often more negatively biased on Twitter than on Facebook, with public opinions being more positively biased on Facebook. Negative sentiments in our study were related to public apprehensions and concerns regarding delays or pauses in vaccine trials, vaccine safety, corporations, and governments influencing vaccine availability and rights exclusivity for economic benefits.

A comparative analysis with independent surveys was carried out. Our findings related to trends of averaged positive and negative sentiment across the United Kingdom and the United States were found to correlate broadly. In the United States, during the early stages of the pandemic, polling indicated that a significant minority had low trust in a vaccine; for example, a Yahoo News/YouGov survey in May 2020 [26] reported that only 55% of people in the United States intended to get vaccinated against COVID-19, while almost 1 in 5 (19%) individuals would not get vaccinated. A similar survey in July 2020 [27] reported that 42% of people in the United States would get vaccinated (27% would not get vaccinated), while a survey in September 2020 [28] reported that only 36% of people in the United States were certain they would get vaccinated (32% would not get vaccinated). More recently, an Axios-Ipsos survey in November 2020 reported, consistent with the findings of our social media analysis, a marked increase in the proportion of individuals who are likely to get vaccinated (51% were "very" or "somewhat" likely to take the first-generation vaccine; this proportion would increase to 70% if the vaccine was proven safe and effective by public health officials) [29].

In the United Kingdom, a YouGov survey in June 2020 [30] reported that 41% of respondents would "probably" or "definitely" get vaccinated, while 1 in 6 (16%) respondents would "definitely" or "probably" not get vaccinated. The survey also reported that individuals who used social media more than traditional media as their source of news were 9% less likely to be in favor of being vaccinated. A more recent YouGov survey in the United Kingdom in November 2020 [31], related to the Pfizer COVID-19 vaccine, reported that 67% of individuals were "very" or "fairly" likely to take the vaccine when available, and approximately 1 in 5 (21%) individuals were unlikely to take it. While there has been a slight increase in the proportion of individuals unlikely to take the vaccine, the proportion of those likely to get vaccinated has increased, indicating a reduction in the number of individuals who were previously unsure. This could be attributed in part to the recent announcements by vaccine manufacturers, and our results corroborate this finding with a marked increase in positive sentiments since mid-October 2020, in both the United Kingdom and the United States, across the 2 social media platforms. Further studies on changes in sentiments could further the current understanding of factors that have contributed to this, with particular focus on the impact of government education programs.

Limitations

It is important to consider the limitations of our data sources and techniques and the related challenges and opportunities they present for future research. While we attempted to gauge nationwide public sentiments in the United Kingdom and the United States by analyzing posts in English on both Facebook and Twitter, our data may not be representative of the broader population of both countries. Users are known to differ in their usage and preferences regarding social media platforms on the basis of their sociodemographics (eg, age, socioeconomic status, and political affiliation). Vaccines are likely to be preferentially targeted at older populations and possibly ethnic minorities, communities with historically lower rates of vaccine uptake [32,33]. Further exploration is therefore imperative to increase our understanding of the public perception toward vaccines and their underlying behavioral determinants [34]. Social network analysis [35,36] can be performed in conjunction with DL methods to effectively identify sources of fake news or misinformation and their social networks to help deal with infodemic challenges [37,38]. Demographic data including age, gender, race, and geographic origin can also be inferred from social media profiles of users by using AI techniques [39]. This can help categorize distinct groups and inform the development of demographic-level engagement and tailored communication strategies to promote diversity and inclusion in vaccination campaigns. These can also effectively account for the fact that there are genuine knowledge voids being filled by misinformation [34].

The technical limitations of our approach include challenges in determining the geographic location of users and issues relating to the accuracy of the AI techniques (eg, interpreting sarcasm and implicit context). Alternative deep neural networks [40-42] and fuzzy-based approaches [43,44] can be explored and potentially integrated as part of our ensemble model in an attempt to further refine the present findings. The 2-step keyword-based thematic filtering process and the use of geotagged posts in this study resulted in relatively small sample sizes. This could be improved by using more sensitive filtering and data-driven search mechanisms, network metadata (such as likes and retweets), and additional social media and web-based platforms. On account of the current study limitations, our approach should only be used in conjunction with other techniques for understanding public sentiments, such as focus groups, input from civil society organizations, surveys, and public consultations.

Future studies could consider conducting periodic public surveys over the period of interest being explored through social media analysis. This would ensure that both methodologies were informed by each other over the course of the study to enable more granular spatiotemporal analysis, thus allowing more robust comparisons from reciprocal findings and deeper insights for policymakers. These could also complement other qualitative methods, such as in-depth interviews and ethnographic studies, as part of mixed-study approaches. Manual annotation or labeling of datasets is imperative when training AI models for NLP tasks to ensure accuracy and generalizability. These can be affected by the skill of annotators and the proportion of the data set that is labeled. Confounding factors, such as political affiliations, should also be included in future studies, by applying further filters to screen strategies and through targeted demographic analysis, to further the current understanding of the underlying determinants of public sentiments. Attitudes toward different vaccine manufacturers could also be explored to identify and assess effective public engagement strategies to build support for ethical principles and maximize the uptake of the imminently available vaccines.

Conclusions

One of the main threats to the resilience of vaccination programs globally is the rapid and global spread of misinformation. Public confidence in COVID-19 vaccines can be exacerbated by unproven concerns regarding vaccine safety, which seed doubt and distrust. Furthermore, there have been cases where vaccine debates have been purposefully polarized, thus exploiting the doubting public and system weaknesses for political purposes, while waning vaccine confidence elsewhere may be influenced by a general distrust in the government and scientific elites. Recent surveys and polls in the United Kingdom and the United States have indicated the fragility of support for vaccination, which furthers the requirement for a better understanding of underlying public concerns and attitudes, both at scale and in real time. Retrospective analysis of 2 popular and most representative social media platforms in this study demonstrates the potential of AI-enabled real-time social media monitoring of public sentiments and attitudes to help detect and prevent such fears and also to enable policymakers to understand the reasons why some social groups may be reluctant to be vaccinated against COVID-19. This can inform more effective policy-making and promote participatory dialogue on complex vaccine deployment issues, under conditions of uncertainty, including decisions on prioritization and equitability, to help maximize the uptake of imminently available vaccines.

Acknowledgments

This study was funded by the Chief Scientist Office of the Scottish Government under its COVID-19 priority research program (COV/NAP/20/07) and is jointly led by co-Principal Investigators AH and AS. AS was supported by BREATHE - The Health Data Research Hub for Respiratory Health [MC_PC_19004], which is funded by the UK Research and Innovation Industrial Strategy Challenge Fund and delivered through Health Data Research UK. AH was supported by the UK Government's Engineering and Physical Sciences Research Council grants (EP/T021063/1, EP/T024917/1). This study does not represent the views of the Scottish or UK Governments.

Conflicts of Interest

AS is a member of the Chief Medical Officer's COVID-19 Advisory Group of the Scottish Government and the UK Government's New and Emerging Respiratory Virus Threats Risk Stratification Subgroup.

Multimedia Appendix 1
Supplementary Material as quoted in the main article.
[DOCX File , 2893 KB - jmir_v23i4e26627_app1.docx]

References

1. de Figueiredo A, Simas C, Karafillakis E, Paterson P, Larson HJ. Mapping global trends in vaccine confidence and investigating barriers to vaccine uptake: a large-scale retrospective temporal modelling study. Lancet 2020 Sep 26;396(10255):898-908 [FREE Full text] [doi: 10.1016/S0140-6736(20)31558-0] [Medline: 32919524]
2. Gellin B. Why vaccine rumours stick—and getting them unstuck. The Lancet 2020 Aug;396(10247):303-304. [doi: 10.1016/s0140-6736(20)31640-8]

3. Horder J. Toll of vaccine hesitancy. Nat Hum Behav 2020 Apr;4(4):335. [doi: 10.1038/s41562-020-0874-1] [Medline: 32303721]
4. Ten threats to global health in 2019. World Health Organization. URL: https://www.who.int/news-room/spotlight/ten-threats-to-global-health-in-2019, [accessed 2021-10-26]
5. Lazarus J, Ratzan SC, Palayew A, Gostin LO, Larson HJ, Rabin K, et al. A global survey of potential acceptance of a COVID-19 vaccine. Nat Med 2021 Feb;27(2):225-228 [FREE Full text] [doi: 10.1038/s41591-020-1124-9] [Medline: 33082575]
6. Bloom BR, Nowak GJ, Orenstein W. "When Will Have a Vaccine?" - Understanding Questions and Answers about Covid-19 Vaccination. N Engl J Med 2020 Dec 03;383(23):2202-2204. [doi: 10.1056/NEJMp2025331] [Medline: 32897660]
7. Fontanet A, Cauchemez S. COVID-19 herd immunity: where are we? Nat Rev Immunol 2020 Oct;20(10):583-584 [FREE Full text] [doi: 10.1038/s41577-020-00451-5] [Medline: 32908300]
8. Perez S. Report: WhatsApp has seen a 40% increase in usage due to COVID-19 pandemic. TechCrunch. 2020 Mar 26. URL: https://techcrunch.com/2020/03/26/report-whatsapp-has-seen-a-40-increase-in-usage-due-to-covid-19-pandemic/ [accessed 2021-03-17]
9. Adeel A, Gogate M, Hussain A. Contextual deep learning-based audio-visual switching for speech enhancement in real-world environments. Information Fusion 2020 Jul;59:163-170. [doi: 10.1016/j.inffus.2019.08.008]
10. Dashtipour K, Gogate M, Li J, Jiang F, Kong B, Hussain A. A hybrid Persian sentiment analysis framework: Integrating dependency grammar based rules and deep neural networks. Neurocomputing 2020 Mar;380:1-10. [doi: 10.1016/j.neucom.2019.10.009]
11. Majumder N, Bhardwaj R, Poria S, Gelbukh A, Hussain A. Improving aspect-level sentiment analysis with aspect extraction. Neural Comput & Applic 2020 Aug 17:1-11 [FREE Full text] [doi: 10.1007/s00521-020-05287-7]
12. Al-Ghadir AI, Azmi AM, Hussain A. A novel approach to stance detection in social media tweets by fusing ranked lists and sentiments. Information Fusion 2021 Mar;67:29-40. [doi: 10.1016/j.inffus.2020.10.003]
13. Zimmer M. “But the data is already public”: on the ethics of research in Facebook. Ethics Inf Technol 2010 Jun 4;12(4):313-325. [doi: 10.1007/s10676-010-9227-5]
14. Williams ML, Burnap P, Sloan L. Towards an Ethical Framework for Publishing Twitter Data in Social Research: Taking into Account Users' Views, Online Context and Algorithmic Estimation. Sociology 2017 Dec;51(6):1149-1168 [FREE Full text] [doi: 10.1177/0038038517708140] [Medline: 29276313]
15. Franzke AS, Bechmann A, Zimmer M, Ess CM. Internet Research: Ethical Guidelines 3.0. Association of Internet Researchers. 2020. URL: https://aoir.org/reports/ethics3.pdf [accessed 2021-03-17]
16. ESRC Research Data Policy. Economic and Social Research Council. URL: https://esrc.ukri.org/files/about-us/policies-and-standards/esrc-research-data-policy/ [accessed 2021-03-17]
17. Waterloo SF, Baumgartner SE, Peter J, Valkenburg PM. Norms of online expressions of emotion: Comparing Facebook, Twitter, Instagram, and WhatsApp. New Media Soc 2018 May;20(5):1813-1831 [FREE Full text] [doi: 10.1177/1461444817707349] [Medline: 30581358]
18. Garmur M, King G, Mukerjee Z, Persily N, Silverman B. CrowdTangle Platform and API. Harvard Dataverse. 2019. URL: https://doi.org/10.7910/DVN/SCCQYD, [accessed 2021-03-17]
19. Banda J, Tekumalla R, Wang G, Yu J, Liu T, Ding Y, et al. A large-scale COVID-19 Twitter chatter dataset for open scientific research -- an international collaboration. arXiv. Preprint posted online November 13, 2020 [FREE Full text]
20. Yoo BH, Kim J, Lee B, Hoogenboom G, Kim KS. A surrogate weighted mean ensemble method to reduce the uncertainty at a regional scale for the calculation of potential evapotranspiration. Sci Rep 2020 Jan 21;10(1):870 [FREE Full text] [doi: 10.1038/s41598-020-57466-0] [Medline: 31964919]
21. Hutto CJ, Gilbert E. VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. In: Proceedings of the International AAAI Conference on Web and Social Media. 2014 Presented at: The 8th International Conference on Weblogs and Social Media; 2014; Ann Arbor, MI URL: http://comp.social.gatech.edu/papers/icwsm14.vader.hutto.pdf,
22. Loria S, Keen P, Honnibal M, Yankovsky R, Karesh D, Dempsey E. TextBlob: simplified text processing. Secondary TextBlob: simplified text processing. Textblob. URL: https://textblob.readthedocs.io/en/dev/#:~:text=TextBlob%20is%20a%20Python%20(2,classification%2C%20translation%2C%20and%20more [accessed 2021-03-17]
23. Devlin J, Chang MW, Lee K, Toutanova K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv. Preprint posted online May 24, 2019 [FREE Full text]
24. COVID-19 AI-powered dashboard for sentiment and opinion mining in social media platforms. Edinburgh Napier University. URL: http://covidtracker.cloud [accessed 2021-03-17]
25. Doshi P. Peter Doshi: Pfizer and Moderna’s “95% effective” vaccines—let’s be cautious and first see the full data. BMJ Opinion. 2020 Nov 26. URL: https://blogs.bmj.com/bmj/2020/11/26/peter-doshi-pfizer-and-modernas-95-effective-vaccines-lets-be-cautious-and-first-see-the-full-data/ [accessed 2021-03-17]
26. Mahase E. Covid-19: Oxford researchers halt vaccine trial while adverse reaction is investigated. BMJ 2020 Sep 09;370:m3525. [doi: 10.1136/bmj.m3525] [Medline: 32907856]

27. Sanders L. Who will get a COVID-19 vaccine when it arrives. YouGov. 2020 May 11. URL: https://today.yougov.com/topics/health/articles-reports/2020/05/11/who-will-get-covid-vaccine [accessed 2021-03-17]
28. Bruce G. Fewer Americans say they'll get vaccinated for COVID-19. YouGov. 2020 Aug 05. URL: https://today.yougov.com/topics/politics/articles-reports/2020/08/05/vaccinated-covid-19 [accessed 2021-03-17]
29. Jaimungal C. Concerns over fast-tracked COVID-19 vaccine has Americans unsure about vaccination. YouGov. 2020 Sep 15. URL: https://today.yougov.com/topics/health/articles-reports/2020/09/15/concerns-over-fast-tracked-covid-19-vaccine-has-am [accessed 2021-03-17]
30. Jackson C, Newall M, Yi J. COVID-19 one year later: American reemergence? Ipsos. 2021 Mar 09. URL: https://www.ipsos.com/en-us/news-polls/axios-ipsos-coronavirus-index [accessed 2021-03-17]
31. YouGov/CCDH Survey Results. YouGov. URL: https://docs.cdn.yougov.com/5mkju0kxbj/CCDH_RESULTS_062620_PR%20%28002%29.pdf [accessed 2021-03-17]
32. McDonnell A. How many Britons are willing to take a coronavirus vaccine? YouGov. 2020 Nov 16. URL: https://yougov.co.uk/topics/health/articles-reports/2020/11/16/how-many-britons-are-willing-take-coronavirus-vacc [accessed 2021-03-17]
33. Forster AS, Rockliffe L, Chorley AJ, Marlow LAV, Bedford H, Smith SG, et al. Ethnicity-specific factors influencing childhood immunisation decisions among Black and Asian Minority Ethnic groups in the UK: a systematic review of qualitative research. J Epidemiol Community Health 2017 Jun;71(6):544-549 [FREE Full text] [doi: 10.1136/jech-2016-207366] [Medline: 27531844]
34. Malik AA, McFadden SM, Elharake J, Omer SB. Determinants of COVID-19 vaccine acceptance in the US. EClinicalMedicine 2020 Sep;26:100495 [FREE Full text] [doi: 10.1016/j.eclinm.2020.100495] [Medline: 32838242]
35. COVID-19 vaccine deployment: Behaviour, ethics, misinformation and policy strategies. The Royal Society. 2020 Oct 21. URL: https://royalsociety.org/-/media/policy/projects/set-c/set-c-vaccine-deployment.pdf [accessed 2021-03-17]
36. Block P, Hoffman M, Raabe IJ, Dowd JB, Rahal C, Kashyap R, et al. Social network-based distancing strategies to flatten the COVID-19 curve in a post-lockdown world. Nat Hum Behav 2020 Jun;4(6):588-596. [doi: 10.1038/s41562-020-0898-6] [Medline: 32499576]
37. Hung M, Lauren E, Hon ES, Birmingham WC, Xu J, Su S, et al. Social Network Analysis of COVID-19 Sentiments: Application of Artificial Intelligence. J Med Internet Res 2020 Aug 18;22(8):e22590 [FREE Full text] [doi: 10.2196/22590] [Medline: 32750001]
38. Gallotti R, Valle F, Castaldo N, Sacco P, De Domenico M. Assessing the risks of 'infodemics' in response to COVID-19 epidemics. Nat Hum Behav 2020 Dec;4(12):1285-1293. [doi: 10.1038/s41562-020-00994-6] [Medline: 33122812]
39. Abd-Alrazaq A, Alhuwail D, Househ M, Hamdi M, Shah Z. Top Concerns of Tweeters During the COVID-19 Pandemic: Infoveillance Study. J Med Internet Res 2020 Apr 21;22(4):e19016 [FREE Full text] [doi: 10.2196/19016] [Medline: 32287039]
40. Brandt J, Buckingham K, Buntain C, Anderson W, Ray S, Pool J, et al. Identifying social media user demographics and topic diversity with computational social science: a case study of a major international policy forum. J Comput Soc Sc 2020 Jan 07;3(1):167-188. [doi: 10.1007/s42001-019-00061-9]
41. Ieracitano C, Pantó F, Mammone N, Paviglianiti A, Frontera P, Morabito FC. Toward an Automatic Classification of SEM Images of Nanomaterials via a Deep Learning Approach. In: Esposito A, Faundez-Zanuy M, Morabito FC, Pasero E, editors. Neural Approaches to Dynamics of Signal Exchanges. Singapore: Springer; Sep 19, 2019:61-72.
42. Mahmud M, Kaiser MS, Hussain A, Vassanelli S. Applications of Deep Learning and Reinforcement Learning to Biological Data. IEEE Trans Neural Netw Learn Syst 2018 Jun;29(6):2063-2079. [doi: 10.1109/TNNLS.2018.2790388] [Medline: 29771663]
43. Zhang X, Zhou M, Liu H, Hussain A. A Cognitively Inspired System Architecture for the Mengshi Cognitive Vehicle. Cogn Comput 2019 Nov 15;12(1):140-149. [doi: 10.1007/s12559-019-09692-6]
44. Ali A, Li J, Kanwal S, Yang G, Hussain A, Jane O'Shea S. A Novel Fuzzy Multilayer Perceptron (F-MLP) for the Detection of Irregularity in Skin Lesion Border Using Dermoscopic Images. Front Med (Lausanne) 2020;7:297 [FREE Full text] [doi: 10.3389/fmed.2020.00297] [Medline: 32733903]

Abbreviations

AI: artificial intelligence
BERT: Bidirectional Encoder Representations from Transformers
DL: deep learning
NLP: natural language processing
VADER: Valence Aware Dictionary for Sentiment Reasoning

Edited by C Basch; submitted 18.12.20; peer-reviewed by E Elyan, C Ieracitano; comments to author 28.12.20; revised version received 04.01.21; accepted 31.01.21; published 05.04.21.

Please cite as:
Hussain A, Tahir A, Hussain Z, Sheikh Z, Gogate M, Dashtipour K, Ali A, Sheikh A
Artificial Intelligence–Enabled Analysis of Public Attitudes on Facebook and Twitter Toward COVID-19 Vaccines in the United Kingdom and the United States: Observational Study
J Med Internet Res 2021;23(4):e26627
URL: https://www.jmir.org/2021/4/e26627
doi:10.2196/26627
PMID:33724919

Original Paper

Prediction Models for the Clinical Severity of Patients With COVID-19 in Korea: Retrospective Multicenter Cohort Study

Bumjo Oh[1*], MD, PhD; Suhyun Hwangbo[2*], MS; Taeyeong Jung[2], BS; Kyungha Min[1], MD; Chanhee Lee[2], BS; Catherine Apio[2], BS; Hyejin Lee[3], MD, PhD; Seungyeoun Lee[4], PhD; Min Kyong Moon[5], MD, PhD; Shin-Woo Kim[6], MD, PhD; Taesung Park[7], PhD

[1]Department of Family Medicine, Seoul Metropolitan Government Seoul National University Boramae Medical Center, Seoul, Republic of Korea
[2]Interdisciplinary Program in Bioinformatics, Seoul National University, Seoul, Republic of Korea
[3]Department of Family Medicine, Seoul National University Bundang Hospital, Gyeonggi-do, Republic of Korea
[4]Department of Mathematics and Statistics, Sejong University, Seoul, Republic of Korea
[5]Department of Internal Medicine, Seoul Metropolitan Government Seoul National University Boramae Medical Center, Seoul, Republic of Korea
[6]Department of Internal Medicine, Kyungpook National University, Daegu, Republic of Korea
[7]Department of Statistics, Seoul National University, Seoul, Republic of Korea
[*]these authors contributed equally

Corresponding Author:
Taesung Park, PhD
Department of Statistics
Seoul National University
1 Gwanak-ro, Gwanak-gu
Seoul, 08826
Republic of Korea
Phone: 82 2 880 8924
Email: tspark@stats.snu.ac.kr

Abstract

Background: Limited information is available about the present characteristics and dynamic clinical changes that occur in patients with COVID-19 during the early phase of the illness.

Objective: This study aimed to develop and validate machine learning models based on clinical features to assess the risk of severe disease and triage for COVID-19 patients upon hospital admission.

Methods: This retrospective multicenter cohort study included patients with COVID-19 who were released from quarantine until April 30, 2020, in Korea. A total of 5628 patients were included in the training and testing cohorts to train and validate the models that predict clinical severity and the duration of hospitalization, and the clinical severity score was defined at four levels: mild, moderate, severe, and critical.

Results: Out of a total of 5601 patients, 4455 (79.5%), 330 (5.9%), 512 (9.1%), and 301 (5.4%) were included in the mild, moderate, severe, and critical levels, respectively. As risk factors for predicting critical patients, we selected older age, shortness of breath, a high white blood cell count, low hemoglobin levels, a low lymphocyte count, and a low platelet count. We developed 3 prediction models to classify clinical severity levels. For example, the prediction model with 6 variables yielded a predictive power of >0.93 for the area under the receiver operating characteristic curve. We developed a web-based nomogram, using these models.

Conclusions: Our prediction models, along with the web-based nomogram, are expected to be useful for the assessment of the onset of severe and critical illness among patients with COVID-19 and triage patients upon hospital admission.

(J Med Internet Res 2021;23(4):e25852) doi:10.2196/25852

KEYWORDS

clinical decision support system; clinical characteristics; COVID-19; SARS-CoV-2; prognostic tool; severity

Introduction

COVID-19, an infectious disease, is currently spreading at an unprecedented pace. The World Health Organization declared COVID-19 a public health emergency of worldwide concern on January 30, 2020, and subsequently a pandemic on March 11, 2020. The COVID-19 pandemic has posed challenges to public health systems worldwide [1,2].

The clinical spectrum of SARS-CoV-2 infection ranges from asymptomatic to fatal, requiring mechanical ventilation [3]. According to initial data from China, the clinical spectrum of COVID-19 is broad, with most infected individuals experiencing only mild or subclinical illnesses, especially in the early phase of the disease [4]. However, a recent study reported that approximately 14%-30% of hospitalized patients diagnosed with COVID-19 develop a severe respiratory failure that requires intensive care [5-7]. The wide range of outcomes observed, ranging from subpopulations that are mainly asymptomatic to those with substantial fatality rates, calls for risk stratification.

Although dexamethasone and remdesivir have recently been considered a preferred treatment strategy, it is still difficult to use them universally for all patients with COVID-19 [8]; hence, supportive treatments to protect multiorgan functions are a major resource for reducing mortality [9,10]. Several promising innovative drugs and treatment strategies are under investigation; however, until they become commercially available, the capacity of the medical system remains limited, prompting the need for making rationing decisions [10,11]. We argue that early identification of patients at the risk of severe respiratory failure would facilitate better resource planning and help set up effective organizational and clinical interventions, including early pharmacotherapy to prevent admission to the intensive care unit.

Since COVID-19 is a pandemic, many studies have assessed regional clinical features among patients. Pandemic preparedness and strategies differ among countries, and the clinical characteristics of patients admitted to medical facilities seem to vary in different cohorts.

We obtained data on 5628 confirmed patients with COVID-19 admitted to hospitals in Korea and analyzed their clinical features and clinical findings upon admission. Therefore, the objectives of this study are to (1) develop models that predict which individuals are at a high risk of severe disease and their duration of hospitalization in a cohort of hospitalized patients with a confirmed diagnosis of COVID-19 and (b) generate a web-based nomogram based on these models. Our results are expected to provide clinicians with a better understanding of the clinical course of COVID-19 and a guideline for critical care rationing.

Methods

Data Source and Study Design

This is a retrospective, multicenter cohort study conducted in Korea. The data used in this study were public data provided by the Korea Disease Control and Prevention Agency (KDCA) in Korea. Data were collected by the KDCA from physicians at multiple centers. The study cohort included 5628 patients with COVID-19 confirmed through the RT–PCR test and hospitalized or released from quarantine upon recovery by April 30, 2020.

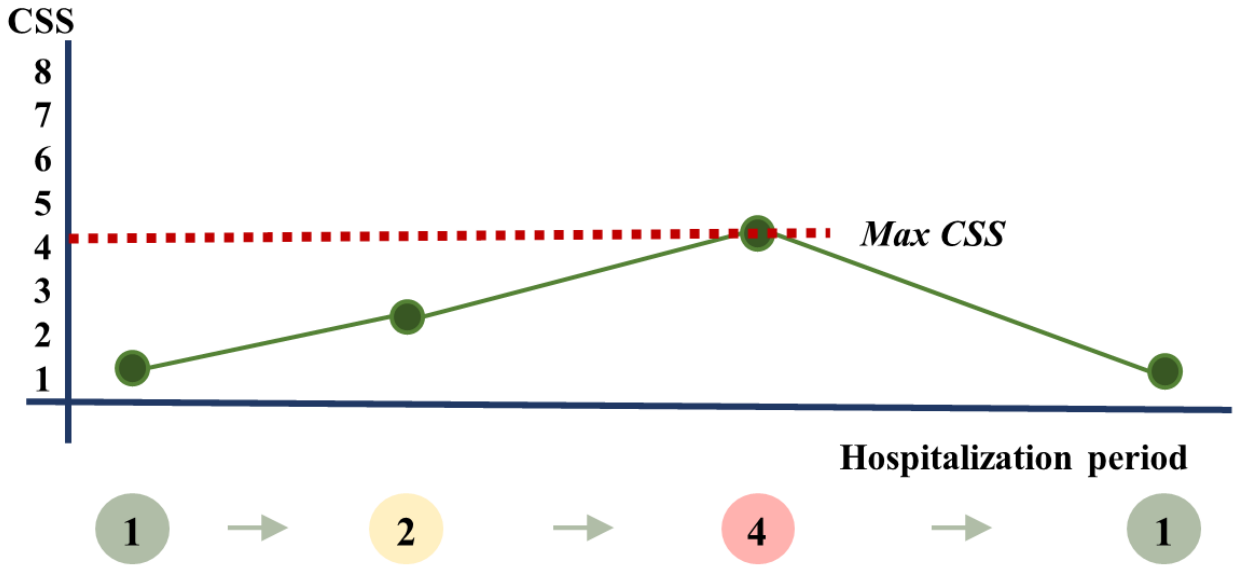
A total of 41 variables were recorded for each patient. These 41 variables are classified into 7 types (Multimedia Appendix 1, Table S1). Among the 41 variables provided by KDCA, 35 were used as predictors, including demographics, physical measurements, initial vital signs, comorbidities, and laboratory findings collected upon admission. We excluded 6 pregnancy-related variables because they were applicable only to women. This study was approved by the institutional review board of Seoul National University (protocol# E2008/003-004).

Definitions of the Primary and Secondary Outcomes

In this study, the primary outcome of interest is the maximum clinical severity score (CSS). The original CSSs provided by the KDCA have 8 levels (Multimedia Appendix 1, Table S2). The CSSs contain ordered information about the clinical severity of patients with COVID-19. For example, the lowest level (ie, level 1) represents no activity restrictions and the highest level (ie, level 8) represents death. As shown in Figure 1, each patient may go through different CSS levels during the course of hospitalization. For each patient, the “max CSS” was defined as the maximum level of CSS reported through their hospital duration (Figure 1). Instead of the original 8 levels, the severity was reclassified into 4 levels depending on the patient’s condition to determine the appropriate treatment in our study. Accordingly, the modified CSS (mCSS) was defined as mild, moderate, severe, and critical (Multimedia Appendix 1, Table S2). The mild group included patients with no activity restrictions, which corresponded to 1 in the original CSS levels. The moderate group displayed limited activity but did not require oxygen therapy. This group corresponded to the original CSS level of 2. Patients who received oxygen therapy were classified under the severe group and those who received ventilation or extracorporeal membrane oxygenation or those who died were classified under the critical group. The severe group corresponded to original CSS levels of 3 and 4, and the critical group corresponded to original CSS levels of 5, 6, 7, and 8.

The secondary outcome was the total duration of hospitalization from the time of admission to discharge. In Korea, once a patient tests positive for COVID-19 on the RT–PCR test, he/she would be admitted to hospital or an isolation facility immediately. Our data set contains data on only the hospitalized patients with COVID-19 having clinical findings such as blood test results.

Figure 1. Diagrammatic representation of the definition of the maximum clinical severity score. CSS: clinical severity score.



Data Preprocessing

Among 35 predictor variables, 7 variables including body temperature, heart rate, and 5 laboratory results were continuous variables, while all the other variables were categorical variables. Among the 7 continuous variables, body temperature and heart rate were recategorized. Specifically, body temperature was divided into 2 categories with 37.5°C considered the threshold, and the heart rate was divided into 3 groups of <60 beats/min, 60-100 beats/min, and ≥100 beats/min. Among the 28 original categorical variables, age, body mass index (BMI), systolic blood pressure (SBP), and diastolic blood pressure (DBP) were recategorized. Age was originally grouped into 10-year-old intervals: <10 years, 10-19 years, 20-29 years, 30-39 years, 40-49 years, 50-59 years, 60-69 years, 70-79 years, and ≥80 years. Of these groups, the values of the age groups of 0-9 years and 10-19 years were merged into 1 group. For BMI, 5 groups were formed: <18.5 kg/m^2, 18.5-22.9 kg/m^2, 23-24.9 kg/m^2, 25-29.9 kg/m^2, and ≥30 kg/m^2. Of these groups, values ranging 25-29.9 kg/m^2 and ≥30 kg/m^2 were merged into 1 group. For SBP, 5 groups were initially formed: <120 mmHg, 120-129 mmHg, 130-139 mmHg, 140-159 mmHg, and ≥160 mmHg. For DBP, 4 groups were initially formed: <80 mmHg, 80-89 mmHg, 90-99 mmHg, and ≥100 mmHg. SBP and DBP were divided into 2 groups based on the values of 140 mmHg and 90 mmHg, respectively.

To analyze the primary outcome (ie, mCSS), 5601 samples were used, excluding missing observations. To analyze the secondary outcome (ie, the duration of hospitalization), 5387 samples were used after excluding patients who died through the course of hospitalization. The median duration of hospitalization was 24 days. Accordingly, we classified the duration of hospitalization into two treatment groups: short-term and long-term.

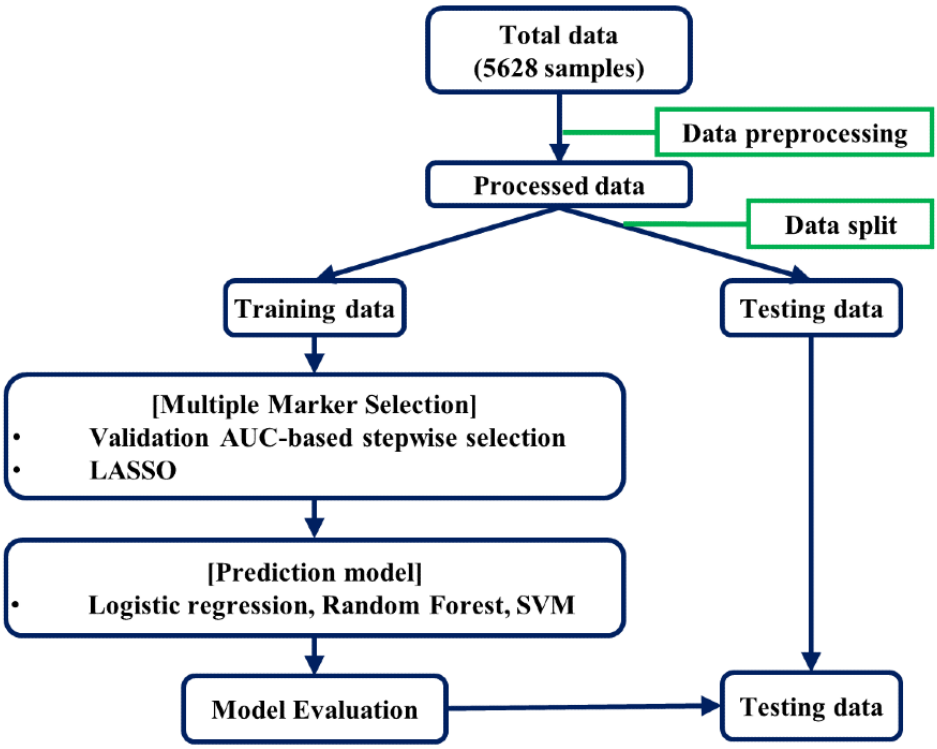
Predictive Marker Selection Through Univariate Analysis

To identify candidate predictive markers related to the primary and secondary outcomes, univariate analysis was first performed. On univariate analysis, mCSS was considered a continuous variable. We performed correlation analysis between mCSS and continuous predictors using the Pearson, Spearman, and Kendall rank correlation tests [12,13], two-tailed *t* test for binary predictors, and analysis of variance for multilevel categorical predictors. Furthermore, we performed the Cochran–Armitage Trend test [14] to identify predictors with a linear trend of mCSS. For the duration of hospitalization, we used a Cox proportional hazards (CoxPH) model to identify candidate predictive markers [15].

Development and Evaluation of the Prediction Model

Figure 2 shows the workflow for model development and evaluation. To avoid overfitting, we evaluated testing errors by splitting the total data set into training and testing data sets in a ratio of 2:1 in a stratified manner, by considering the ratio of the max CSS 4 levels and the long- or short-term group. To maintain the same scale for predictor variables, we standardized each predictor variable.

Figure 2. Workflow for model building and evaluation. AUC: area under the curve, LASSO: least absolute shrinkage and selection operator, SVM: support vector machine.



In order to develop models that predict the max CSS, the 4-level mCSS was combined into two levels in three ways such as (1) y_1: mild (mCSS=1) vs above moderate (mCSS≥2), (2) y_2: below moderate (mCSS≤2) vs above severe (mCSS≥3), and (3) y_3: below severe (mCSS≤3) vs critical (mCSS=4). We fit 3 logistic regression models for binary responses. For multiple marker selection, stepwise variable selection was performed on the basis of the area under the receiver operating characteristic curve (AUC) [16], and we used the least absolute shrinkage and selection operator (LASSO) regression method [17,18]. For both stepwise and LASSO variable selections, 5-fold cross-validation was performed. For prediction models, we considered logistic regression, random forest (RF) classification, and a support vector regression machine [19,20]. Each prediction model was fit using markers selected through stepwise and LASSO regression analyses. The performance of each model was evaluated on the basis of the AUC, sensitivity, and specificity. The optimal threshold for sensitivity and specificity was selected as the threshold value with the maximum balanced accuracy. All analyses were implemented in the R package (version 3.6.1, The R Foundation).

Results

Demographics and Clinical Characteristics

The demographics and clinical characteristics of the 5628 patients, with particular focus on the risk predictors for mCSS or the duration of hospitalization, are presented in Table 1. A complete list of all predictors is provided in Multimedia Appendix 1, Table S3. Among them, 1785 (31.8%) patients were aged over 60 years, and 2320 (41.2%) were male. In total, 1299 (29.4%) patients were overweight or obese by Asia-Pacific BMI criteria. At the time of initial admission, 1936 (35.3%) patients had an SBP of ≥140 mmHg, and 887 (15.9%) had a body temperature of ≥37.5°C. At the time of diagnosis, the patients experienced the following symptoms: fever (n=1305, 23.2%), sputum production (n=1619, 28.8%), shortness of breath (SOB) (n=666, 11.8%), and altered consciousness or confusion (ACC) (n=35, 0.6%).

The patients had the following underlying comorbidities: diabetes mellitus (DM) (n=691, 12.3%), hypertension (HTN) (n=1201, 21.4%), heart failure (HF) (n=59, 1.0%), asthma (n=128, 2.3%), and chronic obstructive pulmonary disease (COPD) (n=40, 0.7%). Initial mean laboratory values were 13.3 (SD 1.8) g/dL for hemoglobin, 39.2% (SD 5%) for hematocrit, 29.1% (SD 11.7%) for the proportion of lymphocytes, 236,733/μL (SD 82,921/μL) for the platelet count, and 6126/μL (SD 2824/μL) for the white blood cell (WBC) count.

Table 1. Demographics and clinical characteristics of the study participants with COVID-19[a] (N=5624).

Variables	Value	*P* value for differences in the modified clinical severity score	*P* value for differences in the duration of hospitalization
Age (years), n (%)		<.001	<.001
0-19	272 (4.8)		
20-29	1119 (19.9)		
30-39	564 (10.0)		
40-49	742 (13.2)		
50-59	1146 (20.4)		
60-69	916 (16.3)		
70-79	545 (9.7)		
≥80	324 (5.8)		
Sex, n (%)		<.001	N/A[b]
Male	2320 (41.2)		
Female	3308 (58.8)		
BMI (kg/m^2), n (%)		.002	N/A
<18.5	260 (5.9)		
18.5-22.9	1867 (42.2)		
23.0-24.9	1039 (23.5)		
≥25	260 (5.9)		
Systolic blood pressure (mmHg), n (%)		<.001	.005
<140	3550 (64.7)		
≥140	1936 (35.3)		
Heart rate (beats/min), n (%)		.003	N/A
<60	108 (2.0)		
60-100	4563 (83.0)		
≥100	828 (15.1)		
Body temperature, (°C), n (%)		<.001	<.001
<37.5	4699 (84.1)		
≥37.5	887 (15.9)		
Fever, n (%)		<.001	<.001
No	4319 (76.8)		
Yes	1305 (23.2)		
Cough, n (%)		.06	<.001
No	3283 (58.4)		
Yes	2341 (41.6)		
Sputum, n (%)		.002	<.001
No	4005 (71.2)		
Yes	1619 (28.8)		
Sore throat, n (%)		<.001	N/A
No	4743 (84.3)		
Yes	881 (15.7)		
Runny nose or rhinorrhea, n (%)		<.001	N/A
No	5003 (89.0)		

Variables	Value	*P* value for differences in the modified clinical severity score	*P* value for differences in the duration of hospitalization
Yes	621 (11.0)		
Muscle aches or myalgia, n (%)		N/A	.001
No	4698 (83.5)		
Yes	926 (16.5)		
Fatigue or malaise, n (%)		<.001	.09
No	5390 (95.8)		
Yes	234 (4.2)		
Shortness of breath or dyspnea, n (%)		<.001	<.001
No	4958 (88.2)		
Yes	666 (11.8)		
Headache, n (%)		<.001	N/A
No	4657 (82.8)		
Yes	967 (17.2)		
Altered consciousness or confusion, n (%)		<.001	.04
No	5589 (99.4)		
Yes	35 (0.6)		
Vomiting or nausea, n (%)		<.001	<.001
No	5380 (95.7)		
Yes	244 (4.3)		
Diabetes mellitus, n (%)		<.001	<.001
No	4934 (87.7)		
Yes	691 (12.3)		
Hypertension, n (%)		<.001	<.001
No	4424 (78.6)		
Yes	1201 (21.4)		
Heart failure, n (%)		<.001	.03
No	5566 (99.0)		
Yes	59 (1.0)		
Chronic cardiovascular disease (except heart failure), n (%)		<.001	.006
No	5430 (96.8)		
Yes	179 (3.2)		
Asthma, n (%)		.003	N/A
No	5497 (97.7)		
Yes	128 (2.3)		
Chronic obstructive pulmonary disease, n (%)		<.001	.02
No	5585 (99.3)		
Yes	40 (0.7)		
Chronic kidney disease, n (%)		<.001	N/A
No	5570 (99.0)		
Yes	55 (1.0)		
Cancer, n (%)		<.001	.07
No	5479 (97.4)		

Variables	Value	*P* value for differences in the modified clinical severity score	*P* value for differences in the duration of hospitalization
Yes	145 (2.6)		
Chronic liver disease, n (%)		.004	N/A
No	5219 (98.4)		
Yes	83 (1.6)		
Dementia, n (%)		<.001	.002
No	5075 (95.8)		
Yes	38 (0.7)		
Hemoglobin (g/dL), mean (SD)	13.3 (1.8)	<.001	<.001
Hematocrit (%), mean (SD)	39.2 (5.0)	<.001	<.001
Lymphocytes (%), mean (SD)	29.1 (11.7)	<.001	<.001
Platelet count (/μL), mean (SD)	236734 (82921)	<.001	<.001
White blood cell count (/μL), mean (SD)	6126 (2824)	<.001	N/A

[a]*P* values were obtained through Pearson correlation analysis for the modified clinical severity score and with the Cox proportional hazards model for the duration of hospitalization.

[b]N/A: not applicable.

The Severity of COVID-19

Based on the severity of COVID-19, determined from the mCSS, patients were divided into four levels: mild (n=4455, 79.5%), moderate (n=330, 5.9%), severe (n=512, 9.1%), and critical (n=304, 5.4%). Among patients aged >60 years, 1157 (64.8%) 1567 (87.8%) belonged to the severe and critical levels, respectively. Specifically, patients in the severe and critical levels and aged ≥60 years accounted for 135 (26.4%) and 58 (19.1%), respectively, of the mCSS cohort. Patients in the severe and critical levels and aged ≥70 years accounted for 125 (24.4%) and 89 (29.3%), and those aged ≥80 years accounted for 72 (14.1%) and 120 (39.5%), respectively. Furthermore, with respect to the duration of hospitalization, patients aged >60 years were more frequently found in the long-term treatment group than in the short-term treatment group.

Univariate Analysis

Table 2 shows the association between the 30 key prediction markers and the mCSS, determined through univariate analysis at a 5% significance level. Patients with an older age; high BMI; SBP of ≥140 mmHg; high heart rate; body temperature of ≥37.5°C; 6 subjective clinical findings including fever, sputum, fatigue or malaise, SOB, ACC, and vomiting or nausea (VN); or 10 comorbidities including DM, HTN, HF, chronic cardiovascular disease (CCD), asthma, COPD, chronic kidney disease, cancer, chronic liver disease, and dementia were likely to have a higher risk of severe disease. Men were found to be at a higher risk of having a high mCSS than women (*P*<.001). Furthermore, patients with a high WBC count or low values of 4 laboratory findings (hemoglobin, hematocrit, lymphocytes, and platelets) tended to be at a higher the risk of severe disease (*P*<.001).

Table 2. Significant markers associated with the modified clinical severity score of the study participants with COVID-19[a].

Variable	*t* test (or ANOVA)		Cochran–Armitage trend test		Pearson correlation analysis		Spearman correlation analysis		Kendall rank correlation analysis	
	t (or *F*)	*P* value	*T*	*P* value	*r*	*P* value	ρ	*P* value	*T*	*P* value
Qualitative										
Age	N/A[b]	<.001	N/A	N/A	0.41	<.001	0.38	<.001	0.32	<.001
Sex	–0.08	<.001	(1)[c]	N/A	–0.05	<.001	–0.03	.01	–0.03	.01
BMI	N/A	<.001	N/A	N/A	0.05 (.002)	.002	0.04	.007	0.04	.007
Systolic blood pressure	0.11	<.001	N/A	<.001	0.06 (*P*<.001)	<.001	0.05	<.001	0.05	<.001
Heart rate	N/A	<.001	N/A	N/A	0.04	.003	0.03 (.03)	N/A	0.03	.03
Temperature	0.43	<.001	N/A	<.001	0.18	<.001	0.18	<.001	0.18	<.001
Fever	0.39	<.001	N/A	<.001	0.19	<.001	0.19	<.001	0.19	<.001
Sputum	0.08	.002	N/A	<.001	0.04	.002	0.04	.006	0.04	.006
Sore throat	–0.17	<.001	1	N/A	–0.07	<.001	–0.06	<.001	–0.06	<.001
Runny nose or rhinorrhea	–0.19	<.001	1	N/A	–0.07	<.001	–0.07	<.001	–0.06	<.001
Fatigue or malaise	0.24	<.001	N/A	<.001	0.06	<.001	0.05	<.001	0.05	<.001
Shortness of breath	0.95	<.001	N/A	<.001	0.36	<.001	0.31	<.001	0.30	<.001
Headache	–0.12	<.001	1	N/A	–0.05	<.001	–0.05	<.001	–0.05	<.001
Altered consciousness or confusion	1.98	<.001	N/A	<.001	0.18	<.001	0.14	<.001	0.14	<.001
Vomiting or nausea	0.26	<.001	N/A	<.001	0.06	<.001	0.06	<.001	0.06	<.001
Diabetes mellitus	0.56	<.001	N/A	<.001	0.21	<.001	0.19	<.001	0.19	<.001
Hypertension	0.57	<.001	N/A	<.001	0.27	<.001	0.25	<.001	0.24	<.001
Heart failure	1.22	<.001	N/A	<.001	0.14	<.001	0.13	<.001	0.13	<.001
Chronic cardiovascular disease	0.6	<.001	N/A	<.001	0.12	<.001	0.12	<.001	0.11	<.001
Asthma	0.23	.01	N/A	.001	0.04	.003	0.04	.005	0.04	<.001
Chronic obstructive pulmonary disease	0.93	<.001	N/A	<.001	0.09	<.001	0.08	<.001	0.08	<.001
Chronic kidney disease	1.19	<.001	N/A	<.001	0.14	<.001	0.12	<.001	0.12	<.001
Cancer	0.43	<.001	N/A	<.001	0.08	<.001	0.07	<.001	0.07	<.001
Chronic liver disease	0.28	.02	N/A	.002	0.04	.004	0.04	.005	0.04	.005
Dementia	1.26	<.001	N/A	<.001	0.29	<.001	0.29	<.001	0.28	<.001
Quantitative										
Hemoglobin	N/A	N/A	N/A	N/A	–0.22	<.001	–0.19	<.001	–0.15	<.001
Hematocrit	N/A	N/A	N/A	N/A	–0.24	<.001	–0.21	<.001	–0.17	<.001
Lymphocytes	N/A	N/A	N/A	N/A	–0.38	<.001	–0.35	<.001	–0.28	<.001
Platelets	N/A	N/A	N/A	N/A	–0.19	<.001	–0.22	<.001	–0.17	<.001
White blood cells	N/A	N/A	N/A	N/A	0.12	<.001	0.04	.02	0.03	.02

[a]For each test, a variable with positive coefficient represents the predictor positively associated with an increase in clinical severity.

[b]N/A: not applicable.

[c]Sex: Female=1, Male=0; clinical findings or comorbidities; Yes=1, No=0.

The results of univariate analysis for the duration of hospitalization are shown in Multimedia Appendix 1, Table S4. We identified 20 key prediction markers associated with the duration of hospitalization. Patients with an older age; SBP of ≥140 mmHg; body temperature of ≥37.5°C; 7 subjective clinical findings including fever, cough, sputum, muscle aches or myalgia, SOB, ACC, and VN; 6 comorbidities including DM, HTN, HF, CCD, COPD, and dementia; or low values of blood parameters including hemoglobin, hematocrit, lymphocytes, and platelets tended to have a long duration of hospitalization.

Development and Evaluation of the Prediction Model

To develop prediction models for mCSS and the duration of hospitalization, we selected multiple markers using AUC-based stepwise selection and the LASSO method. For the application of statistical and machine learning models, we defined three binary response variables by regrouping the 4 levels of mCSS into two levels as follows; (1) y_1: mild (mCSS=1) vs above moderate (mCSS≥2), (2) y_2: below moderate (mCSS≤2) vs above severe (mCSS≥3), and (3) y_3: below severe (mCSS≤3) vs critical (mCSS=4). Table 3 shows the results of variable selection and evaluation for each y. Details regarding the selected variables are provided in Multimedia Appendix 1, Table S5. For each case, we aimed to develop a parsimonious model with higher predictive power. Variables selected through the LASSO method were determined as the final model for each y. For y_1, predictors including older age, high body temperature, SOB, low lymphocyte value, and low platelet count were selected as risk factors. The prediction model with these 5 predictors had an AUC of ≥0.83 (ie, AUC=0.830, sensitivity=0.710, and specificity=0.843 for the RF model). For y_2, older age, high body temperature, SOB, low hematocrit and lymphocyte values, and low platelet count were selected as risk factors. The prediction model with these 6 predictors yielded an AUC of ≥0.865 (ie, AUC=0.865, sensitivity=0.772, and specificity=0.842 for the RF model). For y_3, older age, SOB, high WBC count, low hemoglobin and lymphocyte values, and low platelet count were selected as risk factors. The prediction model with these 6 predictors yielded an AUC of ≥0.933 (ie, AUC=0.933, sensitivity=0.895, and specificity=0.865 for the RF model).

Based on these 3 prediction models, we developed a prognostic nomogram to predict the mCSS for each patient. The nomogram is available on the internet for clinical use [21]. Figure 3 shows an example of the developed nomogram. The fitted results of the logistic model used to develop the nomogram are shown in Multimedia Appendix 1, Table S6. Based on the standardized β coefficients of the fitted results, we ranked the importance of the predictors for each model (Figure 4) [22]. In Figure 4, the x-axis represents the standardized β coefficient, and the relative importance of the predictors is shown in descending order for each model. In all 3 prediction models, the SOB ranked first. The temperature selected in the 2 prediction models ranked second in both models, y_1 and y_2. In all 3 prediction models, lymphocytes ranked third.

We performed similar analyses for the duration of hospitalization. The results of variable selection and evaluation are summarized in Multimedia Appendix 1, Table S7. The prediction model selected 13 predictors through stepwise selection, including age, hematocrit, cough, FM, platelets, muscle aches or myalgia, dementia, asthma, VN, lymphocytes, WBC count, diarrhea, and body temperature. This model yielded an AUC of ≥0.601. With the LASSO method, only age was selected, and the prediction model yielded a performance of up to 0.571.

Table 3. Prediction model and performance for the modified clinical severity score.

Response	Variable selection method	Variables, n	Sample size		Model	Training			Testing		
			Training	Testing		Area under the curve	Sensitivity	Specificity	Area under the curve	Sensitivity	Specificity
y_1[a]	Stepwise	16	2643	1331	Logistic regression	0.865	0.831	0.745	0.853	0.775	0.797
					Random forest	0.958	0.888	0.888	0.841	0.765	0.792
					Support vector machine	0.87	0.783	0.806	0.856	0.83	0.75
y_1	Least absolute shrinkage and selection operator	5	2686	1354	Logistic regression	0.85	0.755	0.794	0.847	0.745	0.812
					Random forest	0.905	0.767	0.847	0.83	0.71	0.843
					Support vector machine	0.854	0.77	0.787	0.848	0.739	0.84
y_2[b]	Stepwise	14	2931	1459	Logistic regression	0.891	0.803	0.807	0.864	0.82	0.765
					Random forest	0.901	0.832	0.85	0.834	0.757	0.837
					Support vector machine	0.887	0.834	0.771	0.854	0.868	0.687
y_2	Least absolute shrinkage and selection operator	6	2683	1348	Logistic regression	0.881	0.757	0.832	0.877	0.812	0.793
					Random forest	0.943	0.87	0.841	0.865	0.772	0.842
					Support vector machine	0.886	0.835	0.768	0.879	0.812	0.816
y_3[c]	Stepwise	11	2931	1460	Logistic regression	0.939	0.952	0.795	0.94	0.982	0.766
					Random forest	0.931	0.871	0.925	0.863	0.807	0.91
					Support vector machine	0.933	0.944	0.789	0.935	0.842	0.895
y_3	Least absolute shrinkage and selection operator	6	2691	1357	Logistic regression	0.923	0.86	0.873	0.944	0.884	0.88
					Random forest	0.991	0.984	0.949	0.933	0.895	0.865
					Support vector machine	0.918	0.812	0.918	0.943	0.874	0.906

[a]y_1: mild vs above moderate.

[b]y_2: below oderate vs above severe.

[c]y_3: below severe vs critical.

Figure 3. An example of our web-based nomogram.

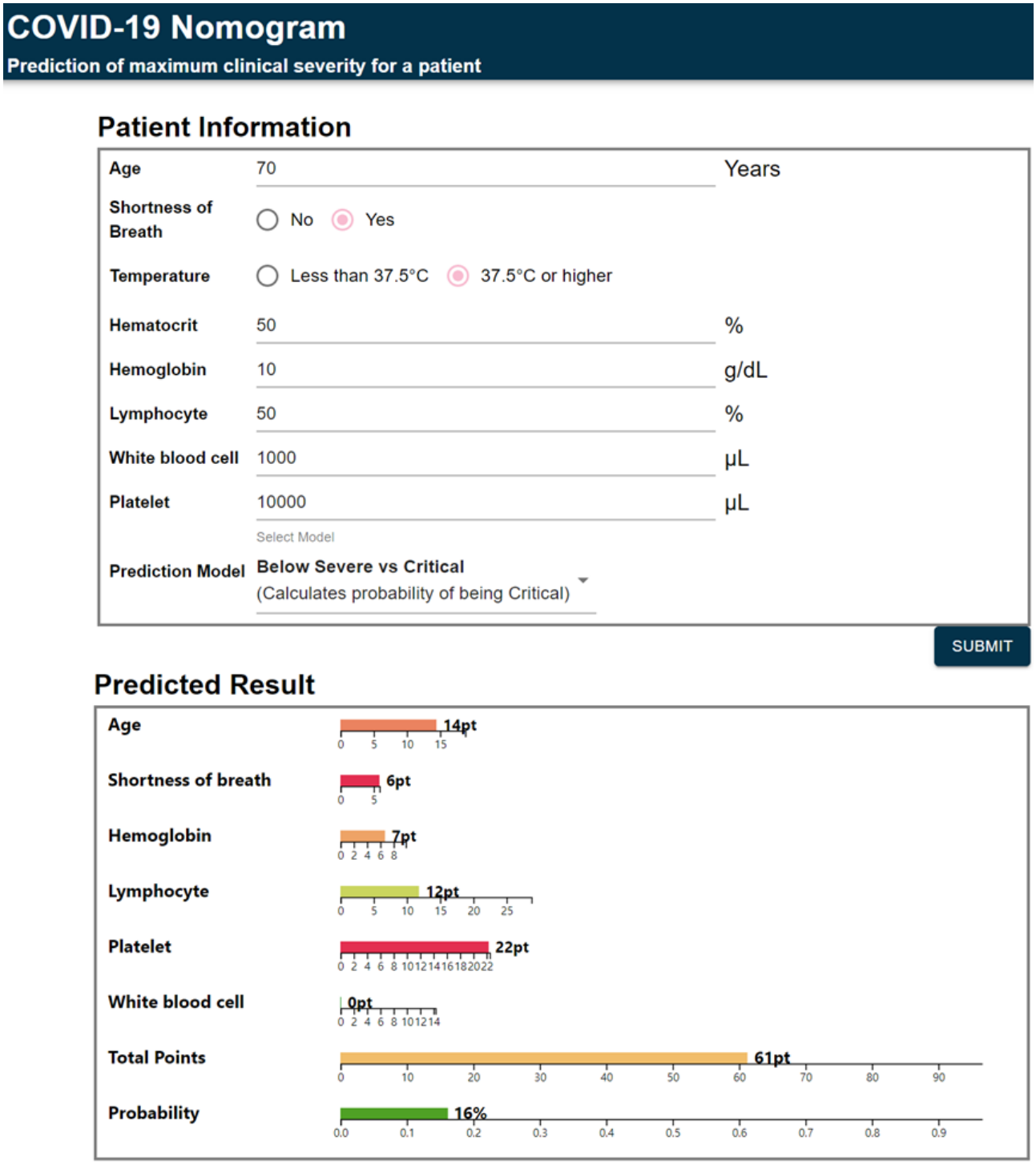
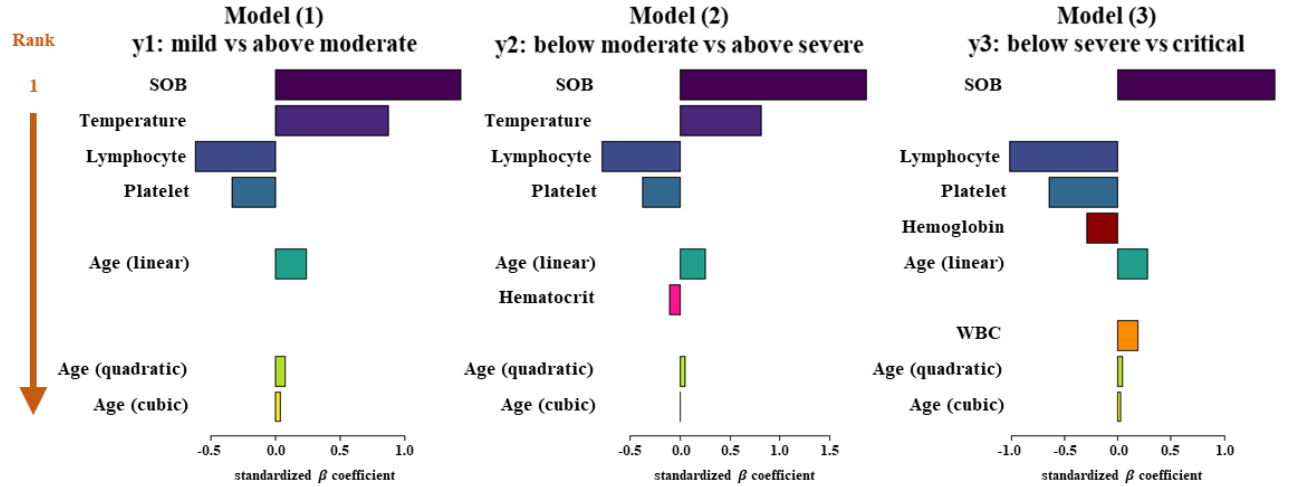


Figure 4. Importance of the predictors. SOB: shortness of breath, WBC: white blood cell.



Discussion

Principal Findings

In this study, we retrospectively assessed the characteristics of 5628 patients with COVID-19 from multiple hospitals in Korea and identified the risk factors for predicting the maximum clinical severity and duration of hospitalization. Older patients aged >60 years accounted for 31.8% of the total, and patients with mild disease accounted for 79.5% of the total. Through univariate analysis for each outcome, we identified 30 risk factors for mCSS and 20 risk factors for the duration of hospitalization. Common risk factors between mCSS and the duration of hospitalization included age, SBP, body temperature, fever, sputum, SOB, ACC, VN, DM, HTN, HF, CCD, COPD, dementia, hemoglobin, hematocrit, lymphocytes, and platelets.

We successfully developed 3 prediction models for mCSS by combining mCSS with 4 levels into 2 levels and developed a web-based nomogram [21] by using these models. Our results indicate that age, body temperature, SOB, lymphopenia, a low hematocrit, low hemoglobin, a low platelet count, and a high WBC count were risk factors positively associated with the maximum clinical severity of COVID-19. These 8 variables have been reported as important predictor variables in the medical literature [23-28]. Specifically, age, shortness of breath, body temperature, lymphocytes, and hemoglobin have been reported as variables for predicting admission to the intensive care unit [23,25], critical illness [24,29], or severe disease [26,27,30]. In particular, Wu et al [26] reported that the severe group had a significantly lower platelet and higher WBC counts than the nonsevere group. Furthermore, Zhang et al [28] reported that hematocrit was significantly lower in the severe group than in the nonsevere group. Our study provides a list of useful risk predictors that can be widely used in a large health care organization during the pandemic.

With an increase in the number of confirmed patients, the number of severely symptomatic patients is also increasing, thus posing a challenge to the management of severe patients during COVID-19 outbreaks. The wide range of outcomes observed, ranging from subpopulations that are mainly asymptomatic to those with markedly high fatality rates, calls for risk stratification. Timely identification of patients at a high risk of developing acute respiratory distress syndrome or multiple organ failure and performing risk stratification management can facilitate more personalized treatment plans and optimized use of medical resources and help prevent further deterioration. To define identify individuals at a high risk of severe disease, the Centers for Disease Control and Prevention defined the following criteria for a high risk of severe disease: age ≥65 years, living in nursing homes, and having at least one underlying comorbidity including chronic lung disease, serious heart conditions, severe obesity, diabetes, chronic kidney disease, liver disease, or an immunocompromised status.

Age and the male gender identified as risk factors of severe COVID-19 in our study have been previously confirmed as risk factors in other countries [31,32]. An elevation in the body temperature is the result of the progression of the infection; hence, if the body temperature is high (≥37.5°C), the prognosis is likely to be poor. In addition, shortness of breath can be considered a symptom that occurs in the course of the disease, since COVID-19 is a type of respiratory disease [33,34]. Among the hematologic abnormalities we observed, we shall consider 2 variables: lymphocytes and platelets. Because lymphopenia and immune dysregulation may impact disease severity, especially because SARS-CoV-2 can directly infect T-lymphocytes, which may be the underlying mechanism of lymphopenia [35]. Regarding the finding of platelet abnormalities, it can be explained that the development of autoimmune antibodies or immune complexes induced by viral infection may play an important role in inducing thrombocytopenia. In addition, SARS-CoV-2 can also directly infect hematopoietic stem or progenitor cells, megakaryocytes, and platelets to inhibit growth and induce apoptosis; furthermore, increased platelet consumption or decreased platelet production in damaged lungs is a potential alternative mechanism that may contribute to thrombocytopenia in severe critical pulmonary conditions [36].

Limitations

Wynants et al [34] reviewed 50 COVID-19 prediction models and reported that most of the models have a high risk of bias when evaluated with the prediction model risk of bias assessment tool [37]. They found that 2 common causes of bias

in prediction models for COVID-19 were the lack of external validation and selection bias. Our study also has these limitations. Since the cohort of patients with COVID-19 in this study includes those whose clinical course has not yet been completed and those who may still potentially develop severe disease, there is a chance that discharged patients without any indication of severe disease during hospitalization would later develop severe disease outside of hospital. In addition, our model was not validated with an external cohort (including foreign data), even though we divided the cohort into a training and testing set to evaluate the predictive power of the developed models. This limitation is mainly due to the limited research environment and the time provided by KDCA to prevent data leakage. Another study limitation is that the data did not include smoking status, which is a very important aspect of an individual's lifestyle, and medication history, especially their history of taking corticosteroids, was not identified. This is an important factor that is closely associated with the exacerbation of the clinical course of COVID-19. The KDCA did not provide information on the smoking status of these individuals because these data were largely missing. In the future, it is expected that more variables from a larger set of patients with COVID-19 be included in the data set to increase the accuracy of the analysis [38,39].

Recently, these prediction tools have been presented in various ways worldwide, but variables with predictive power are identified slightly differently depending on the characteristics of the study population, including nationality and race. In addition, these prediction models can be updated in the current situation where the number of patients continues to rise. Therefore, to develop a model with higher predictive power, it is necessary to constantly compare and validate the results of various studies.

Conclusions

In this study, we developed models that predict the clinical severity of patients with COVID-19. Compared to previous studies that focused on predicting admission to the intensive care unit [23,25], critical illness [24,29], or severe disease [26,27,30], our model used the largest cohort and showed higher performances, even with a limited number of laboratory variables. Specifically, in the case of the model for predicting the critical group, the predictive power was >0.93. Furthermore, we developed a web-based nomogram [21] that can be easily applied visually.

These models are expected to be used as decision supporting tools at the initial stage of treatment; that is, they can be used to predict patients who might need intensive care owing to deterioration among most patients hospitalized with mild or asymptomatic conditions. They can also help hospitals that manage in-patients acquire and use facilities such as negative pressure beds, mechanical ventilation systems, and extracorporeal membrane oxygenation equipment that must be provided to patients with severe symptoms. If further validated through a prospective study, our prediction model might serve for both rationing decisions at health care levels and selecting patients for randomized controlled trials on new treatment options.

Acknowledgments

We thank the members of the Korea Centers for Disease Control and Prevention who provided valuable COVID-19 data during the study. This study was supported by a grant for COVID-19 from the Korea Centers for Disease Control and Prevention (grant# 2020061277D).

Authors' Contributions

BO, SH, TJ, and TP conceived and designed the study. SH and TJ contributed to data analysis and generated the tables and figures. CL developed the web-based nomogram. BO and SH drafted the manuscript and contributed to the literature search. BO, SH, TJ, SK, and TP interpreted the data. All authors critically reviewed and approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary tables.
[DOC File , 249 KB - jmir_v23i4e25852_app1.doc]

References

1. Mahase E. Coronavirus covid-19 has killed more people than SARS and MERS combined, despite lower case fatality rate. BMJ 2020 Feb 18;368:m641. [doi: 10.1136/bmj.m641] [Medline: 32071063]
2. Ranney M, Griffeth V, Jha A. Critical Supply Shortages - The Need for Ventilators and Personal Protective Equipment during the Covid-19 Pandemic. N Engl J Med 2020 Apr 30;382(18):e41. [doi: 10.1056/NEJMp2006141] [Medline: 32212516]
3. Grasselli G, Zangrillo A, Zanella A, Antonelli M, Cabrini L, Castelli A, COVID-19 Lombardy ICU Network. Baseline Characteristics and Outcomes of 1591 Patients Infected With SARS-CoV-2 Admitted to ICUs of the Lombardy Region, Italy. JAMA 2020 Apr 28;323(16):1574-1581 [FREE Full text] [doi: 10.1001/jama.2020.5394] [Medline: 32250385]

4. Wu Z, McGoogan J. Characteristics of and Important Lessons From the Coronavirus Disease 2019 (COVID-19) Outbreak in China: Summary of a Report of 72 314 Cases From the Chinese Center for Disease Control and Prevention. JAMA 2020 Apr 07;323(13):1239-1242. [doi: 10.1001/jama.2020.2648] [Medline: 32091533]
5. Grasselli G, Pesenti A, Cecconi M. Critical Care Utilization for the COVID-19 Outbreak in Lombardy, Italy: Early Experience and Forecast During an Emergency Response. JAMA 2020 Apr 28;323(16):1545-1546. [doi: 10.1001/jama.2020.4031] [Medline: 32167538]
6. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet 2020 Feb 15;395(10223):497-506 [FREE Full text] [doi: 10.1016/S0140-6736(20)30183-5] [Medline: 31986264]
7. Richardson S, Hirsch J, Narasimhan M, Crawford JM, McGinn T, Davidson KW, the Northwell COVID-19 Research Consortium, et al. Presenting Characteristics, Comorbidities, and Outcomes Among 5700 Patients Hospitalized With COVID-19 in the New York City Area. JAMA 2020 May 26;323(20):2052-2059 [FREE Full text] [doi: 10.1001/jama.2020.6775] [Medline: 32320003]
8. Vetter P, Kaiser L, Calmy A, Agoritsas T, Huttner A. Dexamethasone and remdesivir: finding method in the COVID-19 madness. The Lancet Microbe 2020 Dec;1(8):e309-e310. [doi: 10.1016/S2666-5247(20)30173-7]
9. Alhazzani W, Møller MH, Arabi Y, Loeb M, Gong MN, Fan E, et al. Surviving Sepsis Campaign: Guidelines on the Management of Critically Ill Adults with Coronavirus Disease 2019 (COVID-19). Crit Care Med 2020 Jun;48(6):e440-e469 [FREE Full text] [doi: 10.1097/CCM.0000000000004363] [Medline: 32224769]
10. White D, Lo B. A Framework for Rationing Ventilators and Critical Care Beds During the COVID-19 Pandemic. JAMA 2020 May 12;323(18):1773-1774. [doi: 10.1001/jama.2020.5046] [Medline: 32219367]
11. Sanders J, Monogue M, Jodlowski T, Cutrell J. Pharmacologic Treatments for Coronavirus Disease 2019 (COVID-19): A Review. JAMA 2020 May 12;323(18):1824-1836. [doi: 10.1001/jama.2020.6019] [Medline: 32282022]
12. Spearman C. "General Intelligence," Objectively Determined and Measured. Am J Psychol 1904 Apr;15(2):1961.
13. Kendall MG. A New Measure of Rank Correlation. Biometrika 1938;30:81-93.
14. Armitage P. Tests for Linear Trends in Proportions and Frequencies. Biometrics 1955;11(3):375-386.
15. Cox DR. Regression Models and Life-Tables. J R Stat Soc Series B 1972;34(2):187-202.
16. Kim S, Song M, Hwangbo S, Lee S, Cho U, Kim JH, et al. Development of Web-Based Nomograms to Predict Treatment Response and Prognosis of Epithelial Ovarian Cancer. Cancer Res Treat 2019 Jul;51(3):1144-1155 [FREE Full text] [doi: 10.4143/crt.2018.508] [Medline: 30453728]
17. Meier L, Van de Geer S, Bühlmann P. The group lasso for logistic regression. J R Statist Soc B 2008;70(1):53-71 [FREE Full text]
18. Tibshirani R. Regression shrinkage and selection via the lasso: a retrospective. J R Statist Soc B 1996;58(1):267-288.
19. Liaw A, Wiener M. Classification and Regression by randomForest. R news 2002;2(3):18-22 [FREE Full text]
20. Drucker H, Burges C, Kaufman L, Smola A, Vapnik V. Support Vector Regression Machines. In: Proceedings of the 9th International Conference on Neural Information Processing Systems. 1996 Dec Presented at: 9th International Conference on Neural Information Processing Systems p. 155-161.
21. Lee C. COVID-19 Nomogram: Prediction of maximum clinical severity for a patient. Bioinformatics and Biostatistics Laboratory, Seoul National University. 2020 Oct 15. URL: http://statgen.snu.ac.kr/covid19/nomogram/maxcss/ [accessed 2021-04-12]
22. Hong C, Ryu H. Information Theoretic Standardized Logistic Regression Coefficients with Various Coefficients of Determination. Commun Stat Appl Methods 2006;13(1):49-60 [FREE Full text]
23. Zhao Z, Chen A, Hou W, Graham JM, Li H, Richman PS, et al. Prediction model and risk scores of ICU admission and mortality in COVID-19. PLoS One 2020;15(7):e0236618 [FREE Full text] [doi: 10.1371/journal.pone.0236618] [Medline: 32730358]
24. Liang W, Liang H, Ou L, Chen B, Chen A, Li C, China Medical Treatment Expert Group for COVID-19. Development and Validation of a Clinical Risk Score to Predict the Occurrence of Critical Illness in Hospitalized Patients With COVID-19. JAMA Intern Med 2020 Aug 01;180(8):1081-1089 [FREE Full text] [doi: 10.1001/jamainternmed.2020.2033] [Medline: 32396163]
25. Kim H, Han D, Kim JH, Kim D, Ha B, Seog W, et al. An Easy-to-Use Machine Learning Model to Predict the Prognosis of Patients With COVID-19: Retrospective Cohort Study. J Med Internet Res 2020 Nov 09;22(11):e24225 [FREE Full text] [doi: 10.2196/24225] [Medline: 33108316]
26. Wu G, Yang P, Xie Y, Woodruff HC, Rao X, Guiot J, et al. Development of a clinical decision support system for severity risk prediction and triage of COVID-19 patients at hospital admission: an international multicentre study. Eur Respir J 2020 Aug;56(2):2001104 [FREE Full text] [doi: 10.1183/13993003.01104-2020] [Medline: 32616597]
27. Zhang W, Zhang Z, Ye Y, Luo Y, Pan S, Qi H, et al. Lymphocyte percentage and hemoglobin as a joint parameter for the prediction of severe and nonsevere COVID-19: a preliminary study. Ann Transl Med 2020 Oct;8(19):1231 [FREE Full text] [doi: 10.21037/atm-20-6001] [Medline: 33178763]
28. Zhang H, Wang X, Fu Z, Luo M, Zhang Z, Zhang K, et al. Potential Factors for Prediction of Disease Severity of COVID-19 Patients. medRxiv. Preprint posted online March 23, 2020 2020 Mar 23.

29. Schalekamp S, Huisman M, van Dijk RA, Boomsma MF, Freire Jorge PJ, de Boer WS, et al. Model-based Prediction of Critical Illness in Hospitalized Patients with COVID-19. Radiology 2021 Jan;298(1):E46-E54 [FREE Full text] [doi: 10.1148/radiol.2020202723] [Medline: 32787701]
30. Gao Y, Li T, Han M, Li X, Wu D, Xu Y, et al. Diagnostic utility of clinical laboratory data determinations for patients with the severe COVID-19. J Med Virol 2020 Jul;92(7):791-796 [FREE Full text] [doi: 10.1002/jmv.25770] [Medline: 32181911]
31. Liu F, Zhang Q, Huang C, Shi C, Wang L, Shi N, et al. CT quantification of pneumonia lesions in early days predicts progression to severe illness in a cohort of COVID-19 patients. Theranostics 2020;10(12):5613-5622 [FREE Full text] [doi: 10.7150/thno.45985] [Medline: 32373235]
32. Mo J, Liu J, Wu S, Lü A, Xiao L, Chen D, et al. Predictive role of clinical features in patients with coronavirus disease 2019 for severe disease. Zhong Nan Da Xue Xue Bao Yi Xue Ban 2020 May 28;45(5):536-541 [FREE Full text] [doi: 10.11817/j.issn.1672-7347.2020.200384] [Medline: 32879103]
33. Ji D, Zhang D, Xu J, Chen Z, Yang T, Zhao P, et al. Prediction for Progression Risk in Patients With COVID-19 Pneumonia: The CALL Score. Clin Infect Dis 2020 Sep 12;71(6):1393-1399 [FREE Full text] [doi: 10.1093/cid/ciaa414] [Medline: 32271369]
34. Wynants L, Van Calster B, Collins G, Riley RD, Heinze G, Schuit E, et al. Prediction models for diagnosis and prognosis of covid-19 infection: systematic review and critical appraisal. BMJ 2020 Apr 07;369:m1328 [FREE Full text] [doi: 10.1136/bmj.m1328] [Medline: 32265220]
35. Henry B, Benoit S, de Oliveira MHS, Hsieh WC, Benoit J, Ballout RA, et al. Laboratory abnormalities in children with mild and severe coronavirus disease 2019 (COVID-19): A pooled analysis and review. Clin Biochem 2020 Jul;81:1-8 [FREE Full text] [doi: 10.1016/j.clinbiochem.2020.05.012] [Medline: 32473151]
36. Yang M, Ng M, Li C. Thrombocytopenia in patients with severe acute respiratory syndrome (review). Hematology 2005;10(2):101-105. [Medline: 16019455]
37. Wolff R, Moons K, Riley R, Whiting PF, Westwood M, Collins GS, PROBAST Group†. PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies. Ann Intern Med 2019 Jan 01;170(1):51-58. [doi: 10.7326/M18-1376] [Medline: 30596875]
38. Polverino F. Cigarette Smoking and COVID-19: A Complex Interaction. Am J Respir Crit Care Med 2020 Aug 01;202(3):471-472 [FREE Full text] [doi: 10.1164/rccm.202005-1646LE] [Medline: 32530714]
39. Figliozzi S, Masci P, Ahmadi N, Tondi L, Koutli E, Aimo A, et al. Predictors of adverse prognosis in COVID-19: A systematic review and meta-analysis. Eur J Clin Invest 2020 Oct;50(10):e13362. [doi: 10.1111/eci.13362] [Medline: 32726868]

Abbreviations

ACC: altered consciousness or confusion
AUC: area under the receiver operating characteristic curve
CCD: chronic cardiovascular disease
COPD: chronic obstructive pulmonary disease
CoxPH: Cox proportional hazards
CSS: clinical severity score
DBP: diastolic blood pressure
DM: diabetes mellitus
HF: heart failure
HTN: hypertension
KDCA: Korea Disease Control and Prevention Agency
LASSO: least absolute shrinkage and selection operator
mCSS: modified CSS
RF: random forest
SBP: systolic blood pressure
SOB: shortness of breath
VN: vomiting or nausea
WBC: white blood cell

Edited by C Basch; submitted 18.11.20; peer-reviewed by W Han, M Basit, S Sabarguna; comments to author 14.01.21; revised version received 04.02.21; accepted 18.03.21; published 16.04.21.

Please cite as:
Oh B, Hwangbo S, Jung T, Min K, Lee C, Apio C, Lee H, Lee S, Moon MK, Kim SW, Park T
Prediction Models for the Clinical Severity of Patients With COVID-19 in Korea: Retrospective Multicenter Cohort Study
J Med Internet Res 2021;23(4):e25852
URL: https://www.jmir.org/2021/4/e25852
doi:10.2196/25852
PMID:33822738

Original Paper

Adaptive Susceptible-Infectious-Removed Model for Continuous Estimation of the COVID-19 Infection Rate and Reproduction Number in the United States: Modeling Study

Mark B Shapiro[1], PhD; Fazle Karim[1], PhD; Guido Muscioni[1], MSc; Abel Saju Augustine[1], MSc

Anthem, Inc, Indianapolis, IN, United States

Corresponding Author:
Mark B Shapiro, PhD
Anthem, Inc
220 Virginia Avenue
Indianapolis, IN, 46204
United States
Phone: 1 708 295 8150
Email: mark.shapiro@anthem.com

Abstract

Background: The dynamics of the COVID-19 pandemic vary owing to local population density and policy measures. During decision-making, policymakers consider an estimate of the effective reproduction number R_t, which is the expected number of secondary infections spread by a single infected individual.

Objective: We propose a simple method for estimating the time-varying infection rate and the R_t.

Methods: We used a sliding window approach with a Susceptible-Infectious-Removed (SIR) model. We estimated the infection rate from the reported cases over a 7-day window to obtain a continuous estimation of R_t. A proposed adaptive SIR (aSIR) model was applied to analyze the data at the state and county levels.

Results: The aSIR model showed an excellent fit for the number of reported COVID-19 cases, and the 1-day forecast mean absolute prediction error was <2.6% across all states. However, the 7-day forecast mean absolute prediction error approached 16.2% and strongly overestimated the number of cases when the R_t was rapidly decreasing. The maximal R_t displayed a wide range of 2.0 to 4.5 across all states, with the highest values for New York (4.4) and Michigan (4.5). We found that the aSIR model can rapidly adapt to an increase in the number of tests and an associated increase in the reported cases of infection. Our results also suggest that intensive testing may be an effective method of reducing R_t.

Conclusions: The aSIR model provides a simple and accurate computational tool for continuous R_t estimation and evaluation of the efficacy of mitigation measures.

(J Med Internet Res 2021;23(4):e24389) doi:10.2196/24389

KEYWORDS

compartmental models; COVID-19; decision-making; estimate; infection rate; infectious disease; modeling; pandemic; prediction; reproduction number; SARS-CoV-2; United States

Introduction

The COVID-19 pandemic is currently underway. As of September 2, 2020, over 6,000,000 individuals in the United States have been reported positive for COVID-19. Modeling studies are key to understanding the factors that drive the spread of the disease and for developing mitigation strategies. Early modeling efforts forecasted very large numbers of infected individuals, which would overwhelm health care systems in many countries [1-3]. These forecasts served as a call to action for policymakers to introduce policy measures including social distancing, travel restrictions, and eventually lockdowns to avoid the predicted catastrophe [4-6]. The mitigating policy measures have been successful in changing the dynamics of the pandemic and in "flattening the curve," such that fewer people have needed to seek treatment at any given time, and this has prevented the health care system from getting overwhelmed.

One of the most fundamental metrics that describes the pandemic's dynamics is the reproduction number R_t, which is the expected number of secondary infections spread by a single

infectious individual [7]. In 1906, Hamer [8] speculated that the course of an epidemic is determined by the rate of contact between susceptible and infectious individuals. Later, Kermack and McKendrick [9] reported that epidemics end not when there are no susceptible individuals left, but rather when each infectious individual can infect, on average, <1 more individual. The R_t depends on three factors: (1) the likelihood of infection per contact, (2) the period during which infectious individuals freely interact with susceptible individuals and spread the disease, and (3) the rate of contact. The likelihood of infection per contact (factor 1) is determined on the basis of pathogen virulence and protective measures such as social distancing or wearing masks. Free interactions between infectious and susceptible individuals (factor 2) occur until the infectious individual is self-quarantined or hospitalized, either when the individual tests positive or experiences severe symptoms. Finally, the rate of contact (factor 3) is strongly affected by public health measures to mitigate risk [10], such as lockdowns during the COVID-19 pandemic. Thus, R_t is determined on the basis of the biological properties of the pathogen and multiple aspects of social behavior. When $R_t>1$, the number of cases is expected to increase exponentially. The pandemic is considered to have been contained when R_t decreases and remains at <1. Real-time R_t estimation is critical for determining the effect of implemented mitigation measures and future planning.

We propose a method for continuous estimation of the infection rate and R_t to investigate the effect of mitigation measures and immunity acquired by those who recover from the disease. We estimated R_t with a Susceptible-Infectious-Removed (SIR) model [9] that describes the dynamics of population compartments as follows: individuals are initially "susceptible," contract the viral infection and become "infectious," and are then moved to the "removed" compartment once they are quarantined or hospitalized, recover, or die. The SIR model is one of the simplest epidemiological models that still captures the main properties of an epidemic [11,12], and it has been widely used in epidemic modeling studies. In most SIR modeling studies, the model parameters were constant. An SIR model with constant parameters, however, cannot be applied for the COVID-19 pandemic because various mitigating measures were introduced during pandemic progression. The effect of policy changes on COVID-19 dynamics has been modeled using the combination of an SIR model and Bayesian inference [13,14]. In these modeling studies, the rate of infection spread was assumed to be piece-wise linear among the 3 dates of the implementation of policy changes. In another approach, continuous estimation of R_t and an assessment of the effect of mitigation measures were carried out on the basis of estimates of the distribution of the serial intervals between symptom onset in the primary and secondary cases [15-17]. Bayesian inference and methods based on estimations of the serial interval include multiple parameters whose values are not estimated from the data. In contrast, we propose an adaptive SIR (aSIR) model in which only one parameter—the removal rate—is determined from the literature, while the second parameter—the infection rate—is continuously estimated from the data through a sliding window approach. A continuous R_t estimate is then obtained using the infection rate estimate. The SIR model is described as a system of differential equations, and the key idea in our proposed method is that the initial conditions for each window are considered as values estimated for the previous window. The only additional hyperparameter is the length of the sliding window. The proposed method retains the conceptual and computational simplicity of SIR-type models and can be easily extended through the introduction of additional compartments supported by data.

Methods

Data

Data on daily and cumulative confirmed cases between February 29 and September 2, 2020, were obtained from John Hopkins University (JHU), and the dates of interventions by state (eg, state of emergency and stay-at-home orders) were obtained from Wikipedia. The JHU data were available at 2 levels of aggregation: county and state. JHU considers many sources for reporting these data; county-level information was extracted from the websites of the states' departments of health, and state-level data were extracted directly from the website of the Centers for Disease Control and Prevention.

Model

The SIR model is a system of ordinary differential equations:

Here, S is the number of susceptible individuals, I is the number of infectious individuals who freely interact with others and can transmit the infection, R is the number of individuals excluded from the other 2 compartments because they are quarantined or hospitalized, have recovered and acquired immunity, or have died. Several sources of government data on COVID-19 provide the daily number of newly confirmed cases and a cumulative number of confirmed cases. Careful consideration is required to determine whether these numbers should be attributed to the I or R compartment. In the United States, once an individual has been confirmed positive for COVID-19, he/she is expected to be either self-isolated or hospitalized. Therefore, we assigned the data on confirmed cases to the R compartment, and we fit the model to the cumulative number of confirmed cases.

The infection rate is determined as follows:

$$\beta = p \times c \qquad \textbf{(2)}$$

where p is the probability of being infected upon contact with an infectious individual, and c is the average number of contacts per day. We have no data that would allow us to estimate p and c separately; hence, we directly estimated β, as is usually performed when using SIR models.

The removal rate γ determines the rate at which infected individuals are moved from the I to the R compartments. In the context of the COVID-19 pandemic, γ is determined from the time taken for the appearance of severe symptoms, such that the individual can be tested and is self-quarantined or hospitalized, as required. Therefore, we assumed the duration of the infectious period as the average time taken for the infected

individual to be isolated, not the overall time for recovery. We assumed that an individual is infectious from the day he/she contracts the infection before symptom onset [18-20]. The average time to symptom onset is 5-6 days [21-23]. We assumed that the infectious period before the development of severe symptoms is 6 days; hence, $\gamma=1/6$.

Time-Variant Parameter Estimation

The aSIR model contains two parameters, β and γ, with $\gamma=1/6$ obtained from the literature, and β estimated from the reported data for each region of interest. The time-variant $\beta(t)$ was estimated using a sliding window of $\tau=7$ days and step of $s=1$ day, with the estimated values for S and I obtained from the previous window used as the initial conditions for the next window.

The reproduction number was calculated as follows:

$$R_t(t) = \beta(t)/\gamma \quad \textbf{(3)}$$

For the first window, we determined the date when the number of confirmed cases began to increase exponentially. This is important because for many states or counties, very few confirmed cases were initially reported for a number of days or even weeks, which suggests that either the epidemic had not started or the true number of infected individuals was unknown. It is not reasonable to apply an SIR model for this initial period. We considered the onset of the pandemic as the first of the 4 consecutive days in which the number of reported confirmed cases increased in at least 3 days. The initial conditions for system (1) for window 0 were as follows:

$$S_0(0) = N \quad \textbf{(4)}$$

where N is the population in the region of interest, $I_0(0)=1$, and $R_0(0)=0$. The infection rate β_i and $S(t)$, $I(t)$ for t $[0, \tau—1]$ were estimated from the initial conditions and actual R.

The window was slid by $s=1$ point. For the new $i+1$ window, the initial conditions were considered as the estimated values from the previous window $S_{i+1}(0)=S_i(s)$, $I_{i+1}(0)=I_i(s)$, and actual $R_{i+1}(0)=R(s)$. The actual values of $R(t)$ were used, and the infection rates β_{i+1} and $S_{i+1}(t)$, and $I_{i+1}(t)$ were estimated.

For each window, the $R_{t.i}$ was determined as follows:

$$R_{t.i} = \beta_i/\gamma \quad \textbf{(5)}$$

The $R_{t.i}$ was assigned to the last time point of the window. To obtain a smooth estimate of R_t, we used a rolling average of 5 points.

Results

We fit the model for each state and county in the United States. Model performance was evaluated by calculating the quality of fit as the root mean squared error between the actual and fitted R data for all windows concatenated (wRMSE). The fit was excellent with wRMSE<6 across all states. Furthermore, we calculated 1-day, 3-day, and 7-day forecasts of R after each window (Figure 1A). The mean absolute prediction error for the forecasts is provided in Table 1. The 1-day forecast error did not exceed 2.6% across all states, while the 7-day forecast error was large and approached 16.2% for New York. In particular, the 7-day forecast strongly overestimated the number of cases when R_t was rapidly decreasing (Figure 1).

Figure 1. (A) Estimated Infectious and forecast Removed. (B) Estimated reproduction number R_t. The shaded region indicates the dates of the lockdown. While the 1-day and 3-day forecasts are accurate, the 7-day forecast exhibits marked errors when R_t>1 and is rapidly decreasing.

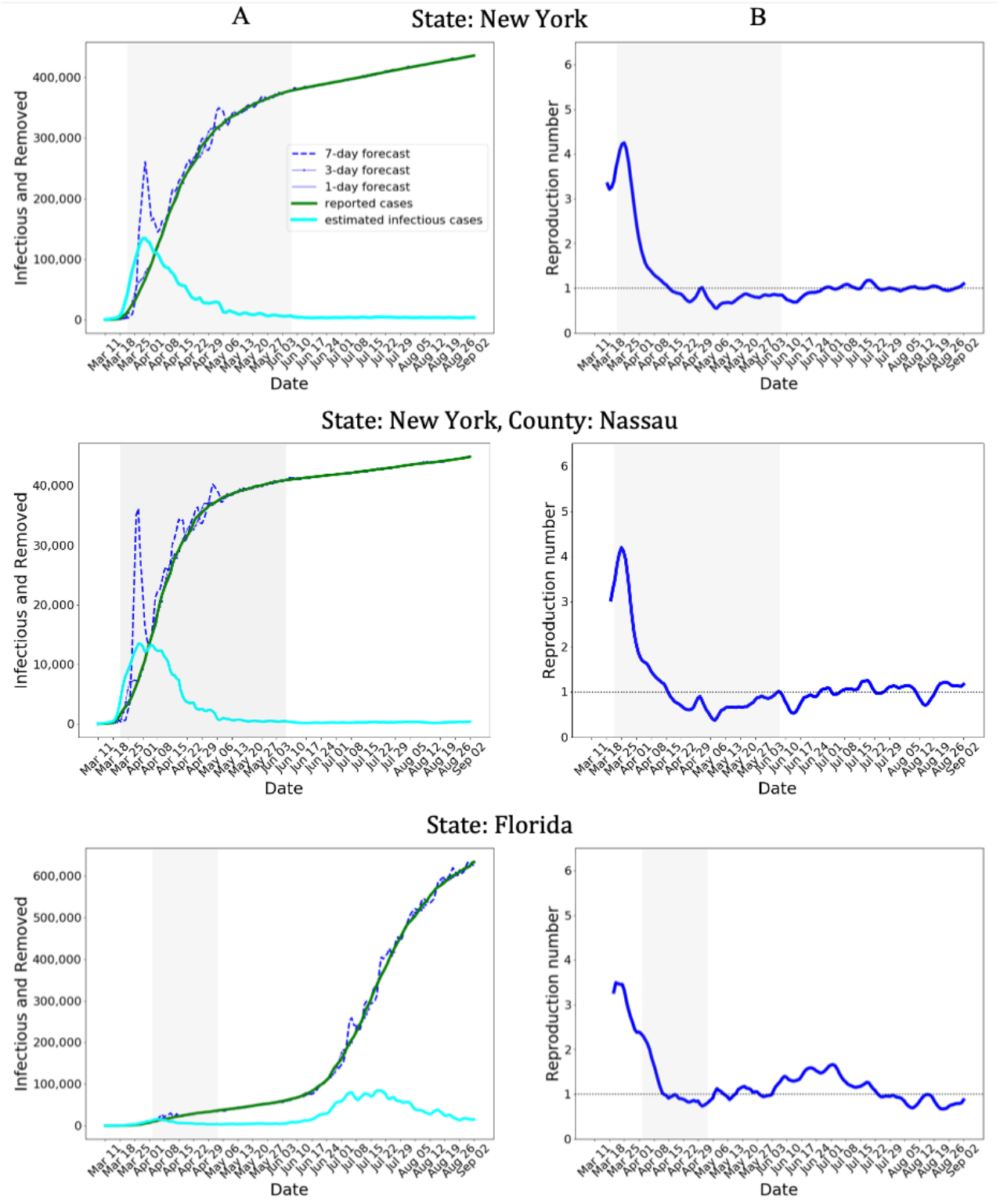


Table 1. Reproduction numbers and forecast accuracy for 50 US states.

State	R_t[a] max	MAPE[b] (1-day forecast), %	MAPE (3-day forecast), %	MAPE (7-day forecast), %
Alabama	2.9	1.5	4.2	10.0
Alaska	2.8	1.6	3.7	11.3
Arizona	3.3	1.3	2.9	10.1
Arkansas	2.8	1.5	4.0	12.6
California	2.5	1.7	2.9	6.3
Colorado	2.6	1.1	2.9	7.3
Connecticut	4.1	2.0	3.1	9.3
Delaware	2.4	1.7	2.9	7.9
District of Columbia	2.1	0.8	1.8	4.4
Florida	3.6	2.0	4.4	9.3
Georgia	3.0	1.8	3.7	7.4
Hawaii	2.7	2.0	3.6	9.7
Idaho	3.4	2.4	4.8	13.6
Illinois	4.0	1.3	2.5	8.5
Indiana	3.8	1.4	4.0	10.4
Iowa	2.8	1.8	3.6	8.0
Kansas	3.0	1.6	3.5	8.6
Kentucky	3.0	2.6	4.9	11.2
Louisiana	3.7	1.8	4.0	12.1
Maine	2.0	1.2	2.8	6.7
Maryland	3.3	1.2	2.8	6.2
Massachusetts	3.4	1.3	3.6	9.7
Michigan	4.5	1.6	3.5	12.8
Minnesota	2.7	1.3	2.9	8.0
Mississippi	2.9	1.2	3.0	9.3
Missouri	3.6	1.8	3.3	11.4
Montana	3.3	1.6	3.7	11.9
Nebraska	2.5	2.0	4.1	9.7
Nevada	2.9	2.3	3.8	10.0
New Hampshire	2.3	1.7	3.3	8.5
New Jersey	4.1	1.5	2.3	7.8
New Mexico	2.3	2.2	3.4	7.3
New York	4.4	1.5	4.2	16.2
North Carolina	3.2	1.3	2.4	7.2
North Dakota	2.4	1.8	4.8	12.6
Ohio	3.3	1.2	3.4	9.8
Oklahoma	3.1	1.4	3.5	10.2
Oregon	2.5	1.3	2.8	6.6
Pennsylvania	3.2	1.7	2.9	6.3
Rhode Island	2.4	1.5	3.1	6.8
South Carolina	3.5	2.1	4.3	10.6
South Dakota	2.1	1.3	3.2	8.7

State	R_t[a] max	MAPE[b] (1-day forecast), %	MAPE (3-day forecast), %	MAPE (7-day forecast), %
Tennessee	3.5	2.2	4.8	12.5
Texas	3.6	2.0	3.9	9.3
Utah	3.2	1.4	3.1	8.3
Vermont	2.9	0.8	2.4	7.7
Virginia	2.5	1.1	2.1	5.1
Washington	3.0	2.0	4.8	8.6
West Virginia	3.5	1.6	3.9	14.0
Wisconsin	3.6	1.5	3.2	10.0
Wyoming	2.9	1.9	4.7	14.2

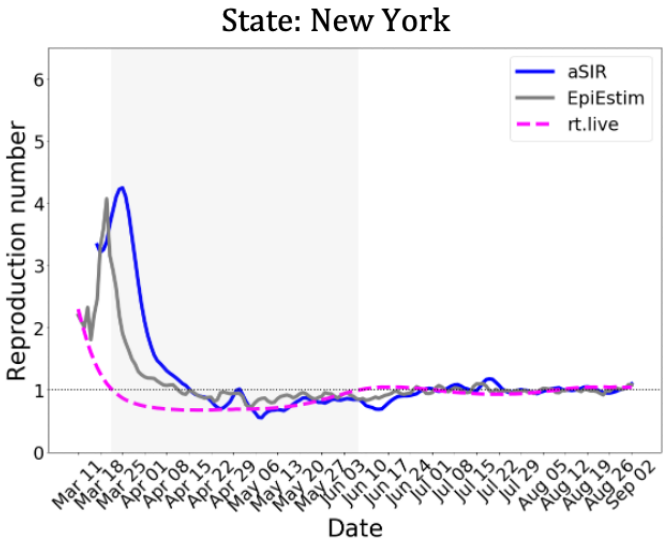
[a]R_t: reproduction number.

[b]MAPE: mean absolute prediction error.

The estimated time course of R_t for New York and Nassau county, one of the most affected counties since the beginning of the COVID-19 pandemic, are shown in Figure 1. The estimated daily number of infectious individuals rapidly increased and then gradually declined after the lockdown was implemented on March 22, 2020 (Figure 1A). The estimated R_t also declined upon implementation of the lockdown (Figure 1B). The time course of R_t exhibits weekly seasonality, which likely reflects the effect of social interactions and possibly the effect of fluctuations in case reporting on weekdays vs weekends. For New York and Nassau county, R_t initially increased, which may reflect the fact that the pandemic in New York was continuously seeded by travelers arriving at John F Kennedy International Airport until a ban on international travel was implemented on March 12, 2020. This may also reflect the fact that not all severe cases were initially recognized and reported as COVID-19 cases. In Florida, R_t decreased to almost 1 by mid-April but then began increasing at the end of May (Figure 1B). In June 2020, Florida authorities introduced more stringent measures to control the pandemic, which is reflected in the reduction in R_t in the second half of July 2020. The opening of multiple states since June 2020 has been accompanied by an increase in R_t beyond 1 (data not shown), and close monitoring of R_t is needed to contain another wave of the pandemic.

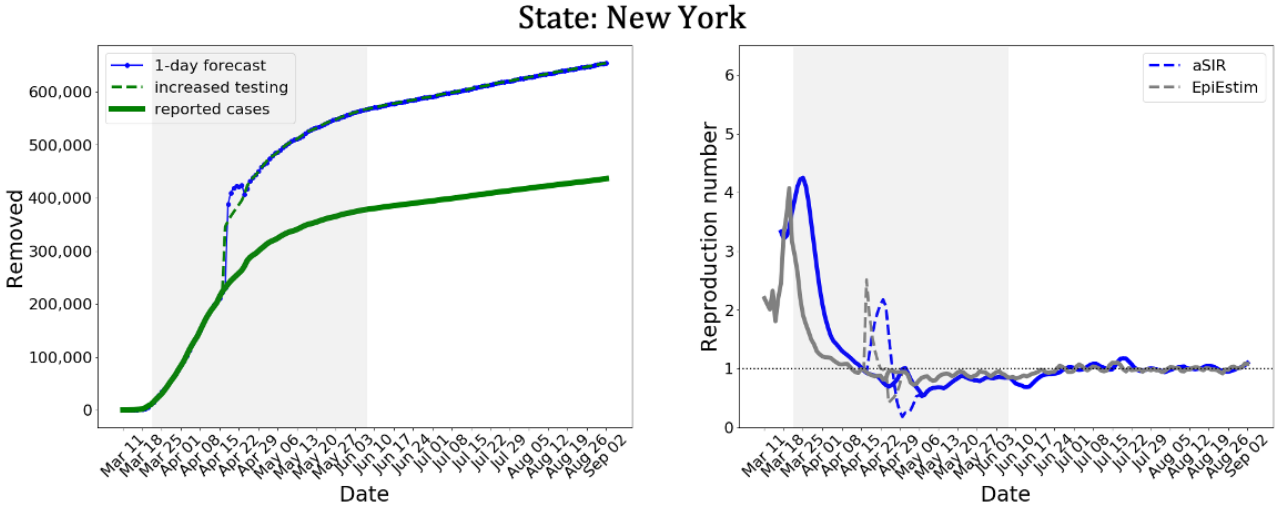
Next, we compared aSIR with the model developed by Cori et al [15], implemented as R package EpiEstim, and a model implemented by Systrom, Vladeck, and Krieger in rt.live [24] (Figure 2). In EpiEstim, we assumed an equal probability of infection within the infectious period of 6 days, the R_t estimate was smoothed with a 7-point rolling average window, same as that in aSIR. While all 3 models show similar estimates when R_t approaches 1, their estimates differ considerably in the beginning of the pandemic. In particular, the rt.live model [24] returned a lower maximum R_t than the other 2 models and estimated that R_t already decreased to 1 by the time the lockdown was announced in New York on March 22, 2020 (Figure 2, shaded region). The EpiEstim and aSIR models estimated similar peak values of R_t, and both models estimated that R_t decreased and approached 1 in the first week of April 2020. Although both models show a rapid reduction in R_t in March, the aSIR model shows a lagged change. However, we are not aware of the ground truth data to determine which model yields a more accurate estimate.

Figure 2. Comparison of models that generate continuous R_t estimates. The three R_t estimates differ widely in the beginning of the COVID-19 pandemic. In particular, the R_t estimated using the rt.live model of Systrom, Vladeck, and Krieger [24] decreased to 1 at the onset at the lockdown on March 22, 2020 (shaded region). aSIR: adaptive Susceptible-Infectious-Removed.



Finally, we investigated the effect of an abrupt increase in testing on the estimated R_t (Figure 3). We assumed a step-wise 50% increase in testing, which persisted after April 12, 2020 (Figure 3, left panel). Both aSIR and EpiEstim models exhibited a spike in R_t. However, an increase in testing would help identify and quarantine infectious individuals sooner, resulting in a shorter infectious period and larger removal rate γ, in turn decreasing R_t. We did not model a potential increase in γ. Instead, we assumed that the underlying dynamics of the pandemic did not change, and within 2 weeks both models returned to the R_t time course estimated without an increase in testing.

Figure 3. Effect of a step-wise 50% increase in testing (left panel, dashed line). The 1-day forecast by the aSIR model adapts within a week. For the Rt estimate, both EpiEstim and our aSIR models produced a spike, followed by a reduction (right panel, dashed lines) before returning to the unperturbed Rt time course (solid lines). aSIR: adaptive Susceptible-Infectious-Removed.



Discussion

Principal Findings

We developed a simple approach to adaptively estimate the time-varying parameters of the SIR model, using reported data on the number of confirmed COVID-19 cases. This approach adds to the already large literature on COVID-19 modeling in 2 ways. First, we estimate the parameters of the SIR model with a sliding window of a limited duration (7 days) to account for rapid changes in transmissibility and contact patterns in response to changes in social behavior and government mitigation measures. The window duration is a hyperparameter that can be changed as needed, the trade-off being the accuracy of the parameter estimates versus the rapid reaction to changes in the underlying pandemic. Because the proposed model is so simple, a number of scenarios can be explored as needed.

Second, we attribute the data on reported cases to the Removed compartment rather than the Infectious compartment. This modeling decision is based on the realities of the COVID-19 pandemic in the United States, where individuals with confirmed COVID-19 are supposed to self-isolate or be hospitalized. Although these individuals remain infectious and can infect

other family members or caregivers even when self-isolated or hospitalized, they would not freely interact with the susceptible population, as would be required to attribute them to the *I* compartment. The addition of a new X compartment in the SIR model has been proposed to model symptomatic quarantined infectious individuals [25]. However, we have no data to independently estimate this additional parameter of quarantine rate. For the same reason, we did not use the Susceptible-Exposed-Infected-Removed (SEIR) model because we are not aware of reliable data on the duration of the exposure period during which an infected person is not yet infectious. Moreover, it has been reported that the SIR model performed better than the SEIR model in representing the information contained in the confirmed-case data on COVID-19 [26].

The reported number of positive COVID-19 cases represents a fraction of infected individuals because of the limited testing capacity in March and April 2020; consequently, only those who developed severe symptoms were tested. Up to 80% of infected individuals may have been asymptomatic or may have experienced mild symptoms [27] and were not tested; hence, for that period, our model applies only to the small subpopulation with severe symptoms. However, this subpopulation is of particular interest because it represents those who are at the greatest risk, and R_t estimated from these limited data can be used to guide policy decisions aimed at protecting the most vulnerable population [28]. As the number of the tested individuals increases, the short sliding window approach makes our model adaptable to an increasing proportion of the population (Figure 3).

Across all US states, the maximal R_t values were estimated for New York (4.4) and Michigan (4.5) (Table 1), which is similar to the mean value of 4.34 estimated for Italy [29] but higher than that obtained with a stochastic transmission model [30,31]. The wide range of maximal values of R_t of 2.0-4.5 (Table 1) likely reflects the differences in contact rates owing to the population density [32,33]. Increased social distancing is required to contain the spread of the pandemic [34,35], with more stringent mitigation measures, including lockdown, considered necessary to decrease the contact rate in high-density states and counties. Another measure to decrease R_t is to increase the removal rate γ through intensive testing and quarantining of individuals who test positive. This targeted intervention would strongly decrease the interaction between infectious and susceptible individuals and maintain an R_t of <1 until a vaccine is available and while vaccination efforts are ramping up. Intensive testing combined with social distancing and mask wearing, followed by the isolation of individuals confirmed with COVID-19, are key features of reopening strategies for schools and universities [36-38]. Our model allows researchers and policymakers to monitor R_t in different geographic regions of the United States, better understand the effect of government policies on the dynamics of the pandemic, and develop further mitigation strategies as we continue to battle COVID-19 [39,40].

Limitations

The SIR model is perhaps the simplest model that captures the dynamics of a pandemic. It is based on several assumptions that are valid only to some degree as we consider real-life scenarios. The 2 main limitations of the original SIR model are that it has constant parameters and it is deterministic. Our proposed aSIR model allows us to estimate time-varying parameters and thus removes the first limitation. The other limitation remains, however. It is assumed that infectious individuals freely interact with the susceptible population. The infection rate β encompasses both the probability of transmission and the average number of contacts per day. The SIR model does not reflect interaction dynamics that are stochastic in nature and are described by stochastic epidemiologic models [15-17,41]. In its simplest form, the SIR model does not reflect the heterogeneity of viral transmission reflected in overdispersion or superspreading where few numbers of infected individuals infect a large number of susceptible individuals [42-44]. The removal rate γ is an average number of days until an infectious individual is excluded and does not reflect the variability of this interval, nor does it allow one to model a possibility of a subsequent, albeit reduced transmission to caregivers or other susceptible individuals as may happen in a real-life scenario. Consequently, R_t, which is calculated using constants β and γ rather than their distributions, does not reflect the stochastic nature of the dynamics of the pandemic. Parameter distributions can be obtained by applying Bayesian methods to SIR modeling [45]. Moreover, a single value of R_t estimated for a large population does not reflect differences in subpopulations, such as age groups, which is especially relevant for COVID-19 [46-48]. The generation of a bank of aSIR models for each subpopulation or region can provide a more realistic insight into the dynamics of the pandemic across a larger population [49,50]. Another critical assumption is that once an individual is infected and recovers, he/she is no longer susceptible to repeated infection; however, this assumption does not appear strongly violated. Although some cases of repeated infection with SARS-CoV-2 have been reported [51-53], the risk of re-infection is considered low [54,55]. Overall, the SIR model is a trade-off between the computational simplicity and veracity of describing the real-life complexities of a pandemic.

Conclusions

SIR type models, particularly the proposed time-variant aSIR model, have an advantage over more complex models in the initial stages of a pandemic when critical public policy decisions need to be made while the empirical data on interaction dynamics, transmission rates, and the disease progression and contagiousness from the moment of infection are not yet readily available. Our model provides a simple and efficient method to assess the efficacy of interventions as the pandemic progresses.

Conflicts of Interest

None declared.

References

1. Arenas A, Cota W, Gomez-Gardenes J, Gómez S, Granell C, Matamalas J, et al. A mathematical model for the spatiotemporal epidemic spreading of COVID19. medRxiv. 2020 Mar 23 Preprint posted online March 23, 2020. [doi: 10.1101/2020.03.21.20040022]
2. Gandhi KRR, Murthy KVR, Rao SSP, Casella F. Non-Pharmaceutical Interventions (NPIs) to Reduce COVID-19 Mortality. SSRN. 2020 Mar 25 Preprint posted online March 25, 2020. [doi: 10.2139/ssrn.3560688]
3. Mitjà O, Arenas À, Rodó X, Tobias A, Brew J, Benlloch JM, 62 signatories. Experts' request to the Spanish Government: move Spain towards complete lockdown. Lancet 2020 Apr 11;395(10231):1193-1194 [FREE Full text] [doi: 10.1016/S0140-6736(20)30753-4] [Medline: 32224297]
4. Adam D. Special report: The simulations driving the world's response to COVID-19. Nature 2020 Apr;580(7803):316-318. [doi: 10.1038/d41586-020-01003-6] [Medline: 32242115]
5. Anderson RM, Heesterbeek H, Klinkenberg D, Hollingsworth TD. How will country-based mitigation measures influence the course of the COVID-19 epidemic? Lancet 2020 Mar 21;395(10228):931-934 [FREE Full text] [doi: 10.1016/S0140-6736(20)30567-5] [Medline: 32164834]
6. Enserink M, Kupferschmidt K. With COVID-19, modeling takes on life and death importance. Science 2020 Mar 27;367(6485):1414-1415. [doi: 10.1126/science.367.6485.1414-b] [Medline: 32217707]
7. Delamater PL, Street EJ, Leslie TF, Yang YT, Jacobsen KH. Complexity of the Basic Reproduction Number (R). Emerg Infect Dis 2019 Jan;25(1):1-4 [FREE Full text] [doi: 10.3201/eid2501.171901] [Medline: 30560777]
8. Hamer WH. The Milroy Lectures ON EPIDEMIC DISEASE IN ENGLAND—THE EVIDENCE OF VARIABILITY AND OF PERSISTENCY OF TYPE. The Lancet 1906 Mar;167(4305):569-574. [doi: 10.1016/S0140-6736(01)80187-2]
9. Kermack WO, McKendrick AG. A contribution to the mathematical theory of epidemics. Proc. R. Soc. Lond. A 1997 Jan;115(772):700-721. [doi: 10.1098/rspa.1927.0118]
10. Non-pharmaceutical public health measures for mitigating the risk and impact of epidemic and pandemic influenza. World Health Organization. 2019 Oct. URL: http://www.who.int/influenza/publications/public_health_measures/publication/en/ [accessed 2020-07-07]
11. Anderson RM. Discussion: the Kermack-McKendrick epidemic threshold theorem. Bull Math Biol 1991;53(1-2):3-32. [doi: 10.1016/s0092-8240(05)80039-4] [Medline: 2059740]
12. Anderson R, May R. Infectious diseases of humans: dynamics and control. Oxford, UK: Oxford University Press; 1992.
13. Dehning J, Zierenberg J, Spitzner F, Wibral M, Neto J, Wilczek M, et al. Inferring change points in the spread of COVID-19 reveals the effectiveness of interventions. medRxiv. 2020 May 06 Preprint posted online May 6, 2020. [doi: 10.1101/2020.04.02.20050922]
14. Karnakov P, Arampatzis G, Kičić I, Wermelinger F, Wälchli D, Papadimitriou C, et al. Data driven inference of the reproduction number (R0) for COVID-19 before and after interventions for 51 European countries. medRxiv. 2020 May 23 Preprint posted online May 23, 2020. [doi: 10.1101/2020.05.21.20109314]
15. Cori A, Ferguson NM, Fraser C, Cauchemez S. A new framework and software to estimate time-varying reproduction numbers during epidemics. Am J Epidemiol 2013 Nov 01;178(9):1505-1512 [FREE Full text] [doi: 10.1093/aje/kwt133] [Medline: 24043437]
16. Wallinga J, Lipsitch M. How generation intervals shape the relationship between growth rates and reproductive numbers. Proc Biol Sci 2007 Mar 22;274(1609):599-604 [FREE Full text] [doi: 10.1098/rspb.2006.3754] [Medline: 17476782]
17. Wallinga J, Teunis P. Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures. Am J Epidemiol 2004 Sep 15;160(6):509-516 [FREE Full text] [doi: 10.1093/aje/kwh255] [Medline: 15353409]
18. Arons MM, Hatfield KM, Reddy SC, Kimball A, James A, Jacobs JR, Public Health–SeattleKing CountyCDC COVID-19 Investigation Team. Presymptomatic SARS-CoV-2 Infections and Transmission in a Skilled Nursing Facility. N Engl J Med 2020 May 28;382(22):2081-2090 [FREE Full text] [doi: 10.1056/NEJMoa2008457] [Medline: 32329971]
19. Hu Z, Song C, Xu C, Jin G, Chen Y, Xu X, et al. Clinical characteristics of 24 asymptomatic infections with COVID-19 screened among close contacts in Nanjing, China. Sci China Life Sci 2020 May;63(5):706-711 [FREE Full text] [doi: 10.1007/s11427-020-1661-4] [Medline: 32146694]
20. Tindale LC, Stockdale JE, Coombe M, Garlock ES, Lau WYV, Saraswat M, et al. Evidence for transmission of COVID-19 prior to symptom onset. Elife 2020 Jun 22;9:e57149 [FREE Full text] [doi: 10.7554/eLife.57149] [Medline: 32568070]
21. Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia. N Engl J Med 2020 Mar 26;382(13):1199-1207 [FREE Full text] [doi: 10.1056/NEJMoa2001316] [Medline: 31995857]
22. Linton N, Kobayashi T, Yang Y, Hayashi K, Akhmetzhanov A, Jung S, et al. Incubation Period and Other Epidemiological Characteristics of 2019 Novel Coronavirus Infections with Right Truncation: A Statistical Analysis of Publicly Available Case Data. medRxiv. 2020 Feb 18 Preprint posted online February 18, 2020. [doi: 10.1101/2020.01.26.20018754]
23. Coronavirus disease (COVID-19). World Health Organization. 2020 Oct 12. URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/question-and-answers-hub/q-a-detail/q-a-coronaviruses [accessed 2021-03-29]
24. Systrom K, Vladeck T, Krieger M. Rt.live. 2020. URL: https://github.com/rtcovidlive/covid-model

25. Maier BF, Brockmann D. Effective containment explains subexponential growth in recent confirmed COVID-19 cases in China. Science 2020 May 15;368(6492):742-746 [FREE Full text] [doi: 10.1126/science.abb4557] [Medline: 32269067]
26. Roda WC, Varughese MB, Han D, Li MY. Why is it difficult to accurately predict the COVID-19 epidemic? Infect Dis Model 2020;5:271-281 [FREE Full text] [doi: 10.1016/j.idm.2020.03.001] [Medline: 32289100]
27. Wu Z, McGoogan JM. Characteristics of and Important Lessons From the Coronavirus Disease 2019 (COVID-19) Outbreak in China: Summary of a Report of 72 314 Cases From the Chinese Center for Disease Control and Prevention. JAMA 2020 Feb 24:1239-1242. [doi: 10.1001/jama.2020.2648] [Medline: 32091533]
28. Lourenço J, Paton R, Ghafari M, Kraemer M, Thompson C, Simmonds P, et al. Fundamental principles of epidemic spread highlight the immediate need for large-scale serological surveys to assess the stage of the SARS-CoV-2 epidemic. medRxiv. 2020 Dec 22 Preprint posted online December 22, 2020. [doi: 10.1101/2020.03.24.20042291]
29. Wangping J, Ke H, Yang S, Wenzhe C, Shengshu W, Shanshan Y, et al. Extended SIR Prediction of the Epidemics Trend of COVID-19 in Italy and Compared With Hunan, China. Front Med (Lausanne) 2020;7:169 [FREE Full text] [doi: 10.3389/fmed.2020.00169] [Medline: 32435645]
30. Abbott S, Hellewell J, Thompson RN, Sherratt K, Gibbs HP, Bosse NI, et al. Estimating the time-varying reproduction number of SARS-CoV-2 using national and subnational case counts. Wellcome Open Res 2020 Jun 1;5:112. [doi: 10.12688/wellcomeopenres.16006.1]
31. Kucharski AJ, Russell TW, Diamond C, Liu Y, Edmunds J, Funk S, Centre for Mathematical Modelling of Infectious Diseases COVID-19 working group. Early dynamics of transmission and control of COVID-19: a mathematical modelling study. Lancet Infect Dis 2020 May;20(5):553-558 [FREE Full text] [doi: 10.1016/S1473-3099(20)30144-4] [Medline: 32171059]
32. Hu H, Nigmatulina K, Eckhoff P. The scaling of contact rates with population density for the infectious disease models. Math Biosci 2013 Aug;244(2):125-134 [FREE Full text] [doi: 10.1016/j.mbs.2013.04.013] [Medline: 23665296]
33. Sy K, White L, Nichols B. Population density and basic reproductive number of COVID-19 across United States counties. medRxiv. 2020 Jun 13 Preprint posted online June 13, 2020. [doi: 10.1101/2020.06.12.20130021]
34. Prem K, Liu Y, Russell TW, Kucharski AJ, Eggo RM, Davies N, Centre for the Mathematical Modelling of Infectious Diseases COVID-19 Working Group, et al. The effect of control strategies to reduce social mixing on outcomes of the COVID-19 epidemic in Wuhan, China: a modelling study. Lancet Public Health 2020 May;5(5):e261-e270 [FREE Full text] [doi: 10.1016/S2468-2667(20)30073-6] [Medline: 32220655]
35. Delen D, Eryarsoy E, Davazdahemami B. No Place Like Home: Cross-National Data Analysis of the Efficacy of Social Distancing During the COVID-19 Pandemic. JMIR Public Health Surveill 2020 May 28;6(2):e19862 [FREE Full text] [doi: 10.2196/19862] [Medline: 32434145]
36. Mukherjee UK, Bose S, Ivanov A, Souyris S, Seshadri S, Sridhar P, et al. Evaluation of reopening strategies for educational institutions during COVID-19 through agent based simulation. Sci Rep 2021 Mar 17;11(1):6264 [FREE Full text] [doi: 10.1038/s41598-021-84192-y] [Medline: 33731722]
37. Paltiel AD, Zheng A, Walensky RP. Assessment of SARS-CoV-2 Screening Strategies to Permit the Safe Reopening of College Campuses in the United States. JAMA Netw Open 2020 Jul 01;3(7):e2016818 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.16818] [Medline: 32735339]
38. Panovska-Griffiths J, Kerr CC, Stuart RM, Mistry D, Klein DJ, Viner RM, et al. Determining the optimal strategy for reopening schools, the impact of test and trace interventions, and the risk of occurrence of a second COVID-19 epidemic wave in the UK: a modelling study. Lancet Child Adolesc Health 2020 Nov;4(11):817-827 [FREE Full text] [doi: 10.1016/S2352-4642(20)30250-9] [Medline: 32758453]
39. Friedman E, Friedman J, Johnson S, Landsberg A. Transitioning Out of the Coronavirus Lockdown: A Framework for Evaluating Zone-Based Social Distancing. Front Public Health 2020;8:266 [FREE Full text] [doi: 10.3389/fpubh.2020.00266] [Medline: 32587845]
40. Inglesby TV. Public Health Measures and the Reproduction Number of SARS-CoV-2. JAMA 2020 Jun 02;323(21):2186-2187. [doi: 10.1001/jama.2020.7878] [Medline: 32356869]
41. Allen LJS. A primer on stochastic epidemic models: Formulation, numerical simulation, and analysis. Infect Dis Model 2017 May;2(2):128-142 [FREE Full text] [doi: 10.1016/j.idm.2017.03.001] [Medline: 29928733]
42. Zhang Y, Li Y, Wang L, Li M, Zhou X. Evaluating Transmission Heterogeneity and Super-Spreading Event of COVID-19 in a Metropolis of China. Int J Environ Res Public Health 2020 May 24;17(10):3705 [FREE Full text] [doi: 10.3390/ijerph17103705] [Medline: 32456346]
43. Lau MSY, Grenfell B, Thomas M, Bryan M, Nelson K, Lopman B. Characterizing superspreading events and age-specific infectiousness of SARS-CoV-2 transmission in Georgia, USA. Proc Natl Acad Sci U S A 2020 Sep 08;117(36):22430-22435 [FREE Full text] [doi: 10.1073/pnas.2011802117] [Medline: 32820074]
44. Adam DC, Wu P, Wong JY, Lau EHY, Tsang TK, Cauchemez S, et al. Clustering and superspreading potential of SARS-CoV-2 infections in Hong Kong. Nat Med 2020 Nov;26(11):1714-1719. [doi: 10.1038/s41591-020-1092-0] [Medline: 32943787]
45. Lawson AB, Kim J. Space-time covid-19 Bayesian SIR modeling in South Carolina. PLoS One 2021;16(3):e0242777 [FREE Full text] [doi: 10.1371/journal.pone.0242777] [Medline: 33730035]

46. Goldstein E, Lipsitch M, Cevik M. On the effect of age on the transmission of SARS-CoV-2 in households, schools and the community. medRxiv. 2020 Jul 28 Preprint posted online July 28, 2020. [doi: 10.1101/2020.07.19.20157362]
47. Davies NG, Klepac P, Liu Y, Prem K, Jit M, CMMID COVID-19 working group, et al. Age-dependent effects in the transmission and control of COVID-19 epidemics. Nat Med 2020 Aug;26(8):1205-1211. [doi: 10.1038/s41591-020-0962-9] [Medline: 32546824]
48. Liu Y, Gu Z, Xia S, Shi B, Zhou X, Shi Y, et al. What are the underlying transmission patterns of COVID-19 outbreak? An age-specific social contact characterization. EClinicalMedicine 2020 May;22:100354 [FREE Full text] [doi: 10.1016/j.eclinm.2020.100354] [Medline: 32313879]
49. Turk PJ, Chou S, Kowalkowski MA, Palmer PP, Priem JS, Spencer MD, et al. Modeling COVID-19 Latent Prevalence to Assess a Public Health Intervention at a State and Regional Scale: Retrospective Cohort Study. JMIR Public Health Surveill 2020 Jun 19;6(2):e19353 [FREE Full text] [doi: 10.2196/19353] [Medline: 32427104]
50. Mehta M, Julaiti J, Griffin P, Kumara S. Early Stage Machine Learning-Based Prediction of US County Vulnerability to the COVID-19 Pandemic: Machine Learning Approach. JMIR Public Health Surveill 2020 Sep 11;6(3):e19446 [FREE Full text] [doi: 10.2196/19446] [Medline: 32784193]
51. Murillo-Zamora E, Mendoza-Cano O, Delgado-Enciso I, Hernandez-Suarez C. Predictors of severe symptomatic laboratory-confirmed SARS-COV-2 reinfection. medRxiv. 2020 Oct 18 Preprint posted online October 18, 2020. [doi: 10.1101/2020.10.14.20212720]
52. Gousseff M, Penot P, Gallay L, Batisse D, Benech N, Bouiller K, in behalf of the COCOREC study group. Clinical recurrences of COVID-19 symptoms after recovery: Viral relapse, reinfection or inflammatory rebound? J Infect 2020 Nov;81(5):816-846 [FREE Full text] [doi: 10.1016/j.jinf.2020.06.073] [Medline: 32619697]
53. To K, Hung I, Ip J, Chu A, Chan W, Tam A, et al. COVID-19 re-infection by a phylogenetically distinct SARS-coronavirus-2 strain confirmed by whole genome sequencing. Clin Infect Dis 2020 Aug 25:ciaa1275 [FREE Full text] [doi: 10.1093/cid/ciaa1275] [Medline: 32840608]
54. Raddad L, Chemaitelly H, Malek J, Ahmed A, Mohamoud Y, Younuskunju S, et al. Assessment of the risk of SARS-CoV-2 reinfection in an intense re-exposure setting. medRxiv. 2020 Sep 28 Preprint posted online September 28, 2020. [doi: 10.1101/2020.08.24.20179457]
55. McMahon A, Robb NC. Reinfection with SARS-CoV-2: Discrete SIR (Susceptible, Infected, Recovered) Modeling Using Empirical Infection Data. JMIR Public Health Surveill 2020 Nov 16;6(4):e21168 [FREE Full text] [doi: 10.2196/21168] [Medline: 33052872]

Abbreviations

SIR: Susceptible-Infectious-Removed
aSIR: adaptive Susceptible-Infectious-Removed
JHU: Johns Hopkins University

Edited by C Basch; submitted 17.09.20; peer-reviewed by T Liu, S Pesälä; comments to author 01.03.21; revised version received 21.03.21; accepted 21.03.21; published 07.04.21.

Please cite as:
Shapiro MB, Karim F, Muscioni G, Augustine AS
Adaptive Susceptible-Infectious-Removed Model for Continuous Estimation of the COVID-19 Infection Rate and Reproduction Number in the United States: Modeling Study
J Med Internet Res 2021;23(4):e24389
URL: https://www.jmir.org/2021/4/e24389
doi:10.2196/24389
PMID:33755577

Original Paper

Classification Models for COVID-19 Test Prioritization in Brazil: Machine Learning Approach

Íris Viana dos Santos Santana[1]; Andressa CM da Silveira[2], MSc; Álvaro Sobrinho[1,3], PhD; Lenardo Chaves e Silva[4], PhD; Leandro Dias da Silva[3], PhD; Danilo F S Santos[2], PhD; Edmar C Gurjão[2], PhD; Angelo Perkusich[2], PhD

[1]Federal University of the Agreste of Pernambuco, Garanhuns, Brazil

[2]Federal University of Campina Grande, Campina Grande, Brazil

[3]Federal University of Alagoas, Maceió, Brazil

[4]Federal Rural University of the Semi-Arid, Pau dos Ferros, Brazil

Corresponding Author:
Álvaro Sobrinho, PhD
Federal University of the Agreste of Pernambuco
Av. Bom Pastor, s/n - Boa Vista
Garanhuns, 55292-270
Brazil
Phone: 55 87981493955
Email: alvaro.alvares@ufape.edu.br

Abstract

Background: Controlling the COVID-19 outbreak in Brazil is a challenge due to the population's size and urban density, inefficient maintenance of social distancing and testing strategies, and limited availability of testing resources.

Objective: The purpose of this study is to effectively prioritize patients who are symptomatic for testing to assist early COVID-19 detection in Brazil, addressing problems related to inefficient testing and control strategies.

Methods: Raw data from 55,676 Brazilians were preprocessed, and the chi-square test was used to confirm the relevance of the following features: *gender*, *health professional*, *fever*, *sore throat*, *dyspnea*, *olfactory disorders*, *cough*, *coryza*, *taste disorders*, and *headache*. Classification models were implemented relying on preprocessed data sets; supervised learning; and the algorithms multilayer perceptron (MLP), gradient boosting machine (GBM), decision tree (DT), random forest (RF), extreme gradient boosting (XGBoost), k-nearest neighbors (KNN), support vector machine (SVM), and logistic regression (LR). The models' performances were analyzed using 10-fold cross-validation, classification metrics, and the Friedman and Nemenyi statistical tests. The permutation feature importance method was applied for ranking the features used by the classification models with the highest performances.

Results: *Gender*, *fever*, and *dyspnea* were among the highest-ranked features used by the classification models. The comparative analysis presents MLP, GBM, DT, RF, XGBoost, and SVM as the highest performance models with similar results. KNN and LR were outperformed by the other algorithms. Applying the easy interpretability as an additional comparison criterion, the DT was considered the most suitable model.

Conclusions: The DT classification model can effectively (with a mean accuracy≥89.12%) assist COVID-19 test prioritization in Brazil. The model can be applied to recommend the prioritizing of a patient who is symptomatic for COVID-19 testing.

(J Med Internet Res 2021;23(4):e27293) doi:10.2196/27293

KEYWORDS

COVID-19; test prioritization; classification models; medical diagnosis

Introduction

Overview

In modern medical systems, health care professionals, managers, and governments use information and data analysis to make decisions [1]. Data is stored, enabling rapid access and sharing during the diagnosis, monitoring, and treatment of patients. Therefore, there are propositions of eHealth and mobile health (mHealth) systems to assist health care professionals and policy makers with decision making [2,3]. Such systems are relevant to provide decision support advice based on patients' data,

helping health care professionals and policy makers address problems related to inefficient COVID-19 testing and control strategies (eg, limited testing resources) in low- and middle–income countries [4]. For example, people who live in low- and middle–income settings, remote settings, and hard-to-reach settings are the most affected by precarious health care. Such a situation is even more critical in a pandemic scenario.

COVID-19 is a disease caused by SARS-CoV-2 [5]. In December 2019, the first cases of COVID-19 appeared in Wuhan, Hubei Province, China [6]. Due to the high growth of COVID-19 confirmed cases worldwide, on January 30, 2020, the World Health Organization considered the COVID-19 outbreak a Public Health Emergency of International Concern [7].

Motivation and Problem Statement

As the number of COVID-19 confirmed cases continuously increases, health care professionals and policy makers need to define guidelines to prevent the disease, delaying the transmission rates. Such guidelines are relevant due to the high probability of collapse in health services and shortages of medical supplies (eg, testing resources) [8]. Confirmation of the first COVID-19 case in Brazil was in March 2020, and since then, there has been an upward trend in confirmed cases and deaths. Unfortunately, the Brazilian government has reported more than 13 million cases, with more than 333,000 deaths. Currently, Brazil is one of the most affected countries by COVID-19, with insufficient control measure implementation. Controlling the COVID-19 outbreak in Brazil is a challenge of continental proportions due to the population's size and urban density, inefficient maintenance of social distancing and testing strategies, and limited availability of testing resources [9].

This study addresses the COVID-19 testing prioritization for patients who are symptomatic to assist early COVID-19 detection in Brazil. Addressing this problem is relevant due to the need for prioritization guidelines to improve testing and control strategies' efficiency. Therefore, the main research question (RQ) is can demographic characteristics and symptoms that do not require expensive exams effectively assist the test prioritization for early COVID-19 detection in Brazil? From the main RQ, the four secondary RQs are (1) what demographic characteristics are relevant to conduct the test prioritization? (2) what symptoms are suitable to drive the test prioritization? (3) what is the most suitable classification model for test prioritization? and (4) what are the impacts of the reduction of reported symptoms in the test prioritization?

Aim of the Study

The study relied on preprocessing a raw data set with information on 55,676 patients, aiming to provide a classification model that effectively recommends or not the prioritization of patients who are symptomatic for COVID-19 testing (ie, a binary classification problem). The implementation of classification models also relied on supervised learning and the algorithms multilayer perceptron (MLP), gradient boosting machine (GBM), decision tree (DT), random forest (RF), extreme gradient boosting (XGBoost), k-nearest neighbors (KNN), support vector machine (SVM), and logistic regression (LR). The algorithms were trained and tested using preprocessed data sets composed of demographic characteristics and reported symptoms that do not require expensive exams [10]. Use of such symptoms is a relevant strategy for COVID-19 test prioritization due to the majority of the Brazilian population's high poverty levels [11].

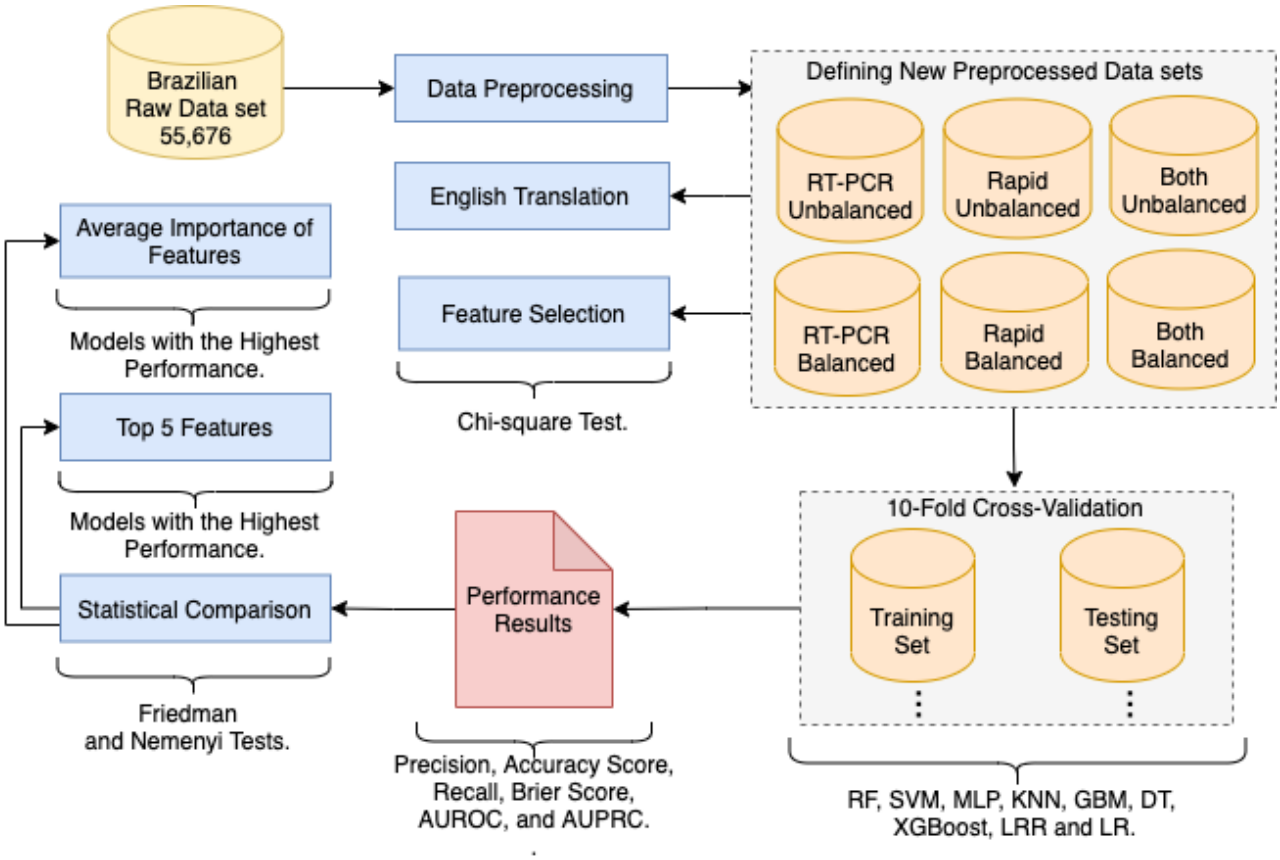
Our findings also provide insights for developers of eHealth and mHealth systems when choosing the most suitable classification model for COVID-19 testing prioritization. Such insights are also relevant for health care professionals and policy makers who envision applying a classification model to prioritize patients who are symptomatic for testing. The study enhances the state of the art by providing three main contributions: (1) the preprocessing of raw data from 55,676 Brazilians, with the availability of data related to patients who are symptomatic [10]; (2) the implementation of classification models, along with reports of feature ranking, to support COVID-19 test prioritization [12]; and (3) a comparative analysis of the classification models.

Methods

Overview

This study's research methodology consists of data preprocessing, the definition of new data sets, English translation, feature selection, 10-fold cross-validation, statistical comparisons, and feature ranking (Figure 1). The raw data from 55,676 Brazilians were preprocessed to define new data sets with information about patients who are symptomatic tested for COVID-19 using reverse transcriptase polymerase chain reaction (RT-PCR) and rapid tests (antibody and antigen). The textual descriptions of six preprocessed data sets (ie, *RT-PCR unbalanced*, *RT-PCR balanced*, *rapid unbalanced*, *rapid balanced*, *both unbalanced*, and *both balanced*) were translated from Portuguese into English for public data availability. The chi-square test was applied in the new data sets to support the feature selection with a $P<.01$, verifying the relevance of features for the classification task by dependence and independence relations [13]. The chi-square test for independence compared two variables in a contingency table to verify if they relate to each other.

Figure 1. Overview of the research methodology applied for the study. The methodological steps consist of data preprocessing, the definition of new data sets, English translation, feature selection, 10-fold cross-validation, statistical comparisons, and feature ranking. AUPR: area under the precision-recall curve; AUROC: area under the receiver operating characteristic curve; DT: decision tree; GBM: gradient boosting machine; KNN: k-nearest neighbors; LR: logistic regression (weak regularization); LRR: logistic regression (strong regularization); MLP: multilayer perceptron; RF: random forest; RT-PCR: reverse transcription polymerase chain reaction; SVM: support vector machine; XGBoost: extreme gradient boosting.



We applied the 10-fold cross-validation method, with five repetitions, to validate the MLP, GBM, DT, RF, XGBoost, KNN, SVM, and LR (weak/strong regularization) classification models using the six data sets. We selected such algorithms because they have different characteristics such as using neural layers, tree combinations, and calculating the distance between data. The mean results for classification metrics were also calculated: precision, accuracy score, recall, Brier Score, area under the receiver operating characteristic curve (AUROC), and area under the precision-recall curve (AUPRC). The recall results were further analyzed using the Friedman and Nemenyi statistical tests to improve the classification models' comparisons. We used the Friedman test to verify the differences between classification models. We applied the Nemenyi test to group classification models based on the verification of differences using multiple comparisons. Finally, we conducted features' ranking for each classification model with the highest performance using the permutation feature importance method, providing average importance and SD. The source code for replication is available in a GitHub repository [12].

Data Collection

The raw data from 55,676 Brazilians included information on tested patients in a spreadsheet format. However, the data collection is not a contribution of this study. The raw data was collected by the public health agency of the city of Campina Grande, Paraíba State in Northeast Brazil. Such a public agency is informed by all the COVID-19 exams performed in the city of Campina Grande. The health agency employees removed patient identification, and the data made available were reused to enable this study. The raw data set comprises categorical features such as *health professional*, *security professional*, *ethnicity*, *test type*, *fever*, *sore throat*, *dyspnea*, *olfactory disorders*, *cough*, *coryza*, *taste disorders*, *headache*, *additional symptoms*, *test result*, *comorbidities*, *test status*, and *symptoms description*.

Data Preprocessing

We conducted the data preprocessing using the Python programming language. The raw data set was preprocessed by applying string matching algorithms to correct inconsistencies. One example of inconsistency was the occurrence of empty columns of symptoms; however, the same symptoms were in a column for the general description of symptoms.

Furthermore, the following instances from the total 55,676 sample were removed due to our exclusion criteria: patients with uncompleted tests or undefined final classifications (n=12,929, 23.22%), duplicated instances (n=251, 0.45%), outliers related to input errors (n=10,408, 18.69%), test types that are not RT-PCR or rapid (n=771, 1.38%), undefined gender (n=27, 0.05%), and patients who were asymptomatic (n=11,269, 20.24%). Patients who were asymptomatic were removed because the inputs for the algorithms rely on demographic characteristics and symptoms.

Removing the feature related to the symptoms' descriptions provides dimensionality reduction in the raw data set feature space. For example, fatigue was removed because the symptom was reported by 228 (0.41%) of the 55,676 patients. Given the main focus on symptoms, the data sets did not include comorbidities and the remaining features (eg, *ethnicity*). As inclusion criteria, the most frequently reported symptoms (ie, fever, sore throat, dyspnea, olfactory disorders, cough, coryza, taste disorders, and headache) and relevant demographic characteristics (ie, gender and health professional) were selected as features of unbalanced and balanced data sets (Table 1). Health care professionals were considered relevant due to the frequency of exposure to SARS-CoV-2. However, for gender, there is no consensus if there is a difference in the proportions of males and females infected with SARS-CoV-2 (usually a relatively even distribution) [14,15].

Table 1. Demographic and symptoms from patients who are symptomatic of both test type data sets.

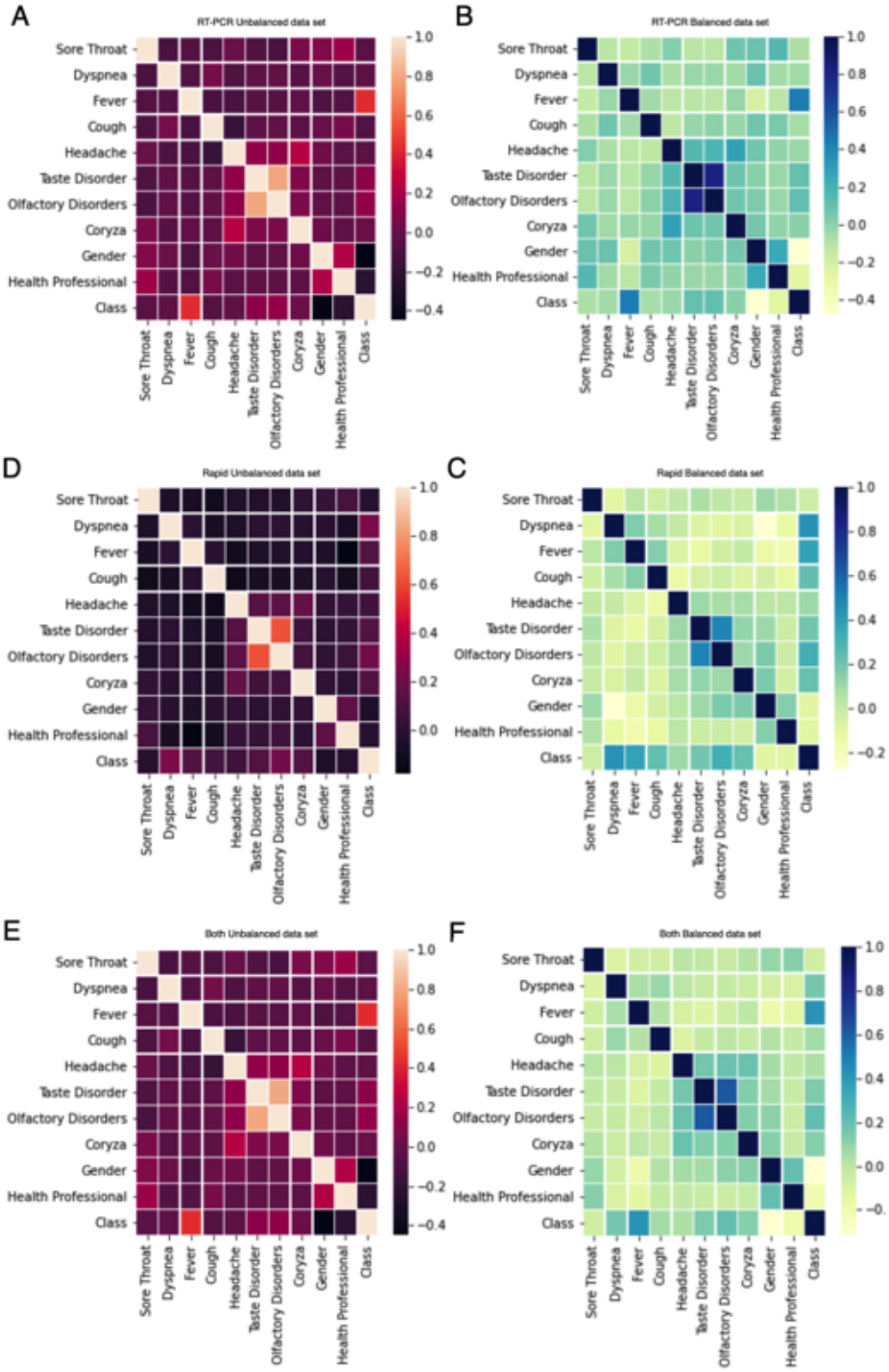
Features	Unbalanced (n=20,021)	Balanced (n=3128)
Demographic characteristics		
Gender: male, n (%)	8919 (44.55)	1639 (52.40)
Health professional, n (%)	2485 (12.41)	475 (15.19)
Symptoms		
Fever, n (%)	9169 (45.80)	1856 (59.34)
Sore throat, n (%)	5976 (29.85)	848 (27.11)
Dyspnea, n (%)	3704 (18.50)	1082 (34.59)
Olfactory disorders, n (%)	1967 (9.82)	522 (16.69)
Cough, n (%)	11,641 (58.14)	1944 (62.15)
Coryza, n (%)	1159 (5.79)	266 (8.50)
Taste disorders, n (%)	1596 (12.37)	387 (12.37)
Headache, n (%)	4034 (20.15)	577 (18.45)

The categorical data were converted into binary representations during the preprocessing. For the feature *gender*, the number 0 represents a female patient, and 1 represents a male. For the features *health professional*, *fever*, *sore throat*, *dyspnea*, *olfactory disorders*, *cough*, *coryza*, *taste disorders*, and *headache*, the number 0 represents a positive response, and 1 represents a negative response. For each data set, the *test result* was the class that can be labeled as 0 for recommending a patient who is symptomatic for COVID-19 test prioritization or 1 for not recommending such patient's prioritization.

The preprocessing included undersampling using the near-miss technique [16], considering COVID-19 positive and negative cases. Undersampling was applied instead of oversampling to prevent the use of synthetic data in training and testing sets. However, as stated, unbalanced data were also considered, without undersampling, to improve the experiments' representativity and to achieve a scenario closer to a real-world setting, with more negative than positive COVID-19 cases.

Using the chi-square test for the *both unbalanced* and *both balanced* data sets, the independence hypothesis was only confirmed for *headache*. For the *RT-PCR unbalanced* data set, the independence hypothesis was confirmed for *sore throat*, *dyspnea*, *headache*, and *coryza*. In the *rapid unbalanced* data set, the independence hypothesis was confirmed for *sore throat* and *health professionals* features. For the *RT-PCR balanced* data set, the independence hypothesis was confirmed for *dyspnea*, *cough*, *headache*, and *coryza*; while for the *rapid balanced* data set, the hypothesis was only confirmed for *sore throat*. Such information was used for feature selection during the experiments, presenting scenarios with different numbers of symptoms to implement classification models. Furthermore, we used a correlation matrix to analyze the correlation coefficients between the features for each data set (Figure 2). For example, *fever* was among the features with the highest correlation coefficients for all data sets.

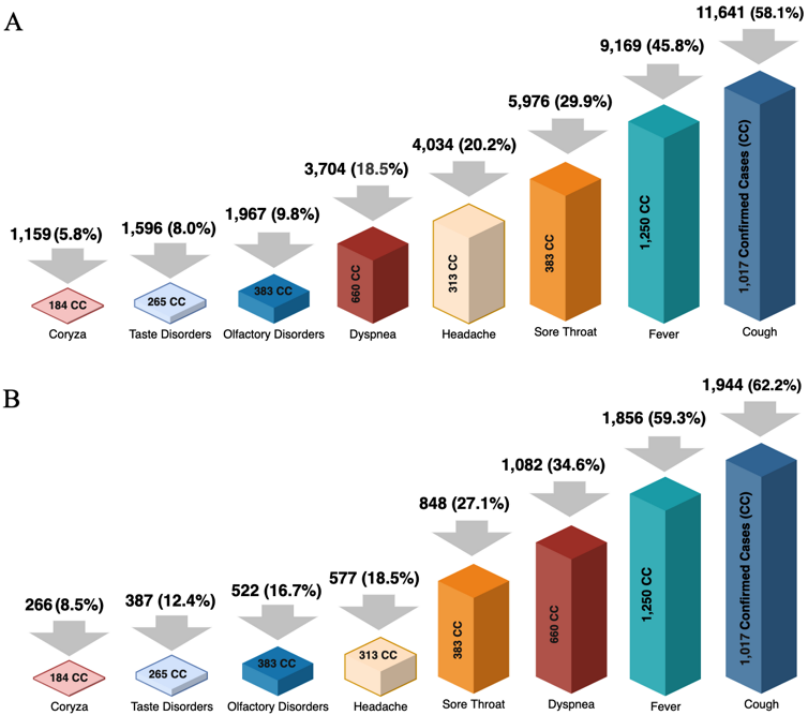
Figure 2. Correlation matrix for (A) RT-PCR unbalanced data set, (B) RT-PCR balanced data set, (C) rapid unbalanced data set, (D) rapid balanced data set, (E) both unbalanced data set, and (F) both balanced data set. RT-PCR: reverse transcription polymerase chain reaction.



The *both unbalanced* data set was composed of 20,021 patients tested by both RT-PCR and rapid tests. The reduction in the number of patients occurred due to the uncompleted tests, duplicated instances, outliers related to input errors, test type, and patients who were asymptomatic. The *both unbalanced* data set contained 1564 (7.81%) positive and 18,457 (92.19%) negative COVID-19 cases, while the balanced one included 1564 cases of each class. From the female patients, 496 (2.48%) were positive and 10,606 (52.97%) were negative cases. For male patients, 1068 (5.33%) were positive and 7851 (39.21%) were negative cases. Cough was the most frequent symptom (n=11,641, 58.1%). Fever was the second most common symptom (n=9169, 45.8%). The remaining symptoms were reported by at most 5976 (29.9%) patients who were symptomatic (Figure 3A).

The *both balanced* data set contained 3128 patients tested by RT-PCR and rapid tests. The near-miss technique reduced the number of negative cases to be equal to positive cases; 496 (15.86%) were positive and 993 (31.75%) were negative cases from the female patients. For males, 1068 (34.14%) were positive and 571 (18.25%) were negative cases. Cough and fever continued to be the first and second most frequently reported symptoms, respectively. The remaining symptoms were also reported by at most 1082 (34.6%) patients (Figure 3B).

Figure 3. (A) The frequency of symptoms for the 20,021 patients who were symptomatic of the both unbalanced data set and the number of CCs. Top values are frequencies; numbers on the geometric forms are the CC for frequency. (B) The frequency of symptoms for the 3128 patients who were symptomatic of the both balanced data set and the number of CCs.



Finally, the *RT-PCR unbalanced* data set included 916 (32.96%) positive and 1863 (67.04%) negative COVID-19 cases, while the balanced one included 916 cases of each class. The *rapid unbalanced* data set included 648 (3.76%) positive and 16,594 (96.24%) negative COVID-19 cases, while the balanced one included 648 of each class. The six scenarios' presentations aim to compare the classification models' results using various test types. Thus, there was no requirement to implement different clinical protocols or select patients with specific profiles for testing based on the results related to the six scenarios presented in this paper.

Algorithms

We implemented the classification models using supervised learning and the MLP, GBM, DT, RF, XGBoost, KNN, SVM, and LR algorithms. An MLP machine learning (ML) algorithm [17] of one hidden layer learns the function:

where W_1 represents the weights of the input layer, W_2 represents the hidden layer, b_1 is the bias added to the hidden layer, b_2 is the output layer, and g is the activation function.

The GBM is a fixed-size DT that uses a boosting strategy [18]. This ML algorithm has a built-in feature selection and aims to provide the estimation or approximation  or the function $F^*(x)$ that maps x to y, minimizing the expected value using a loss function $L(y, F(x))$ over the joint distribution [19], given by:

$$F^* = arg\ min_F E_{y,x} L(y,F(x)) = arg\ min_F E_x [E_y (y,F(x)) \mid x] \quad \textbf{(2)}$$

A DT is an ML algorithm that usually uses a divide and conquer strategy to generate a directed acyclic graph by applying division rules based on information gain [20]. The algorithm has a built-in feature selection, and the information gain is guided by the concept of entropy H, which measures the randomness of a discrete random variable A (with domain $a_1, a_2, ..., a_n$), given by:

where p_i is the probability of observing each value $a_1, a_2, ..., a_n$. This algorithm enables a straightforward interpretation of results by following the decision rules of a unique tree.

The RF is an ML algorithm that relies on classification and regression trees, following specific tree growing rules, tree combination, self-testing, and postprocessing [21]. The algorithm has a built-in feature selection, assessed by the Gini impurity criterion index. The binary split of a node n is given by:

where p_j is the relative frequency of class j. This algorithm also enables a straightforward interpretation of results by following the decision rules of the trees.

As a variant of the GBM, the XGBoost is a regression tree with the same decision rules as a DT [22]. If the XGBoost ML algorithm consists of K DTs, the optimization objective function is given by:

where f_k is an independent tree with leaf scores, and F is the space of a regression tree. Both algorithms enable a straightforward interpretation of results.

The KNN is a distance-based ML algorithm that identifies a new instance based on a neighbor's distance [23]. An instance represents a point in the space, and the algorithm calculates the distance between two points using a metric such as the Euclidean distance, given by:

where x_i and x_j are vectors representing objects in the space, and  and  are the *l-th* elements of the vectors.

The SVM is a ML algorithm that handles binary data using a line to achieve the maximum distance between the data. The algorithm comprises four basic concepts: separation hyperplane, maximum margin hyperplane, soft margin, and kernel function [17]. For instance, the maximization of the margin hyperplane is given by:

where y_i are the output variables, x_i are input vectors, b is the bias, K is a dot products function (Kernel), and α_i is calculated by the maximization of:

where x_j are the named support vectors when α_i is greater than 0.

Finally, the LR is an extension of linear regression that estimates relations between variables using a sigmoid function during probabilistic classifications [24], given by:

where z is the weighted sum of the evidence of a class. Regularization can also be used to prevent overfitting. We applied the LR algorithm to compare a compact and linear model's performance with the previous ML approaches.

We used the Python programming language and the SciPy library [25] to implement and validate the classification models based on such algorithms. We applied the random search method to configure the algorithms' hyperparameters to improve performance carefully. The configurations can be verified in the GitHub repository [12].

Classification Metrics

We calculated the precision, accuracy score, recall, Brier Score, AUROC, and AUPRC for the classification models [26]. The precision represents the proportion of classifications that are true positives and is given by:

where TP is the true positives and FP is the false positives. The accuracy score presents fractions of correct classifications and is given by:

where A is the accuracy score,  is the classified value of a sample, y_i is the corresponding true value, n is the number of samples, and $I(x)$ is the indicator function.

The recall calculates the actual positives that are correctly positives and is given by:

where FN is the number of false negatives. It is relevant for evaluating classifications related to diagnosis due to the highly undesired impacts of false negatives.

The Brier Score provides the mean squared difference between predicted probabilities and expected results, given by:

where f_t is the predicted value, o_t is the expected value, and n is the number of samples.

Finally, the AUROC provides an overview of the diagnostic abilities of the models. However, the use of the AUPRC is usually recommended when handling problems using unbalanced data.

Results

The implementations of classification models using the MLP, GBM, DT, RF, XGBoost, KNN, SVM, and LR algorithms are available in the GitHub repository [12]. Using 10-fold cross-validation with five repetitions, the mean values of precision, accuracy score, recall, and Brier Score of the DT-based classification models were among the best results (Table 2). Such models presented similar results using the six data sets. For the *RT-PCR unbalanced/balanced* and *both unbalanced/balanced* data sets, the LR algorithm was outperformed by the other models. In the results, LR and LRR stand for models with weak and strong regularization, respectively.

Table 2. Results of 10-fold cross-validation for the classification models using the unbalanced and balanced data sets.

Data sets and models	Precision (%)	Accuracy score (%)	Recall (%)	Brier Score
RT-PCR[a] unbalanced and balanced				
MLP[b], unbalanced (balanced)	97.33 (95.86)	96.24 (95.81)	97.08 (95.80)	0.04 (0.04)
GBM[c], unbalanced (balanced)	97.32 (95.95)	96.30 (95.70)	97.17 (95.47)	0.04 (0.04)
RF[d], unbalanced (balanced)	97.42 (96.06)	96.55 (96.00)	97.46 (95.97)	0.04 (0.04)
DT[e], unbalanced (balanced)	97.49 (96.50)	96.33 (95.91)	97.04 (95.32)	0.04 (0.04)
XGBoost[f], unbalanced (balanced)	97.36 (95.94)	96.30 (95.52)	97.13 (95.10)	0.04 (0.04)
KNN[g], unbalanced (balanced)	97.38 (95.92)	96.55 (95.48)	97.50 (95.04)	0.03 (0.05)
SVM[h], unbalanced (balanced)	97.17 (95.84)	96.19 (95.58)	97.18 (95.34)	0.04 (0.04)
LRR[i], unbalanced (balanced)	86.97 (76.86)	86.72 (81.70)	94.37 (90.93)	0.13 (0.18)
LR[j], unbalanced (balanced)	87.00 (76.56)	86.72 (80.63)	94.33 (88.53)	0.13 (0.19)
Rapid unbalanced and balanced				
MLP, unbalanced (balanced)	99.33 (96.66)	98.70 (95.40)	99.32 (94.10)	0.01 (0.05)
GBM, unbalanced (balanced)	99.33 (96.18)	98.72 (95.33)	99.34 (94.50)	0.01 (0.05)
RF, unbalanced (balanced)	99.26 (96.42)	98.76 (95.21)	99.44 (93.98)	0.01 (0.05)
DT, unbalanced (balanced)	99.37 (95.51)	98.69 (94.59)	99.27 (93.67)	0.01 (0.05)
XGBoost, unbalanced (balanced)	99.33 (96.83)	98.72 (95.41)	99.34 (93.94)	0.01 (0.05)
KNN, unbalanced (balanced)	99.31 (97.43)	98.84 (94.58)	99.49 (91.63)	0.01 (0.05)
SVM, unbalanced (balanced)	99.30 (97.30)	98.73 (95.60)	99.37 (93.85)	0.01 (0.04)
LRR, unbalanced (balanced)	96.65 (82.00)	96.23 (84.22)	99.53 (87.93)	0.04 (0.16)
LR, unbalanced (balanced)	96.75 (84.75)	96.14 (85.33)	99.32 (86.32)	0.04 (0.15)
Both unbalanced and balanced				
MLP, unbalanced (balanced)	95.36 (93.53)	94.82 (89.18)	99.20 (84.23)	0.05 (0.11)
GBM, unbalanced (balanced)	95.23 (93.67)	94.73 (89.31)	99.25 (84.38)	0.05 (0.11)
RF, unbalanced (balanced)	95.31 (93.81)	94.87 (89.22)	99.32 (84.04)	0.05 (0.11)
DT, unbalanced (balanced)	95.43 (93.75)	94.79 (89.12)	99.10 (83.87)	0.05 (0.11)
XGBoost, unbalanced (balanced)	95.32 (93.60)	94.78 (89.22)	99.21 (84.24)	0.05 (0.11)
KNN, unbalanced (balanced)	95.50 (92.77)	91.09 (88.63)	94.81 (83.86)	0.09 (0.11)
SVM, unbalanced (balanced)	95.21 (93.36)	94.75 (89.33)	99.30 (84.73)	0.05 (0.11)
LRR, unbalanced (balanced)	92.45 (80.79)	92.04 (80.48)	99.48 (80.11)	0.08 (0.20)
LR, unbalanced (balanced)	92.49 (82.44)	91.98 (81.08)	99.36 (79.14)	0.08 (0.19)

[a]RT-PCR: reverse transcription polymerase chain reaction.

[b]MLP: multilayer perceptron.

[c]GBM: gradient boosting machine.

[d]RF: random forest.

[e]DT: decision tree.

[f]XGBoost: extreme gradient boosting.

[g]KNN: k-nearest neighbors.

[h]SVM: support vector machine.

[i]LRR: logistic regression (strong regularization).

[j]LR: logistic regression (weak regularization).

When removing features according to the chi-square results, there was a considerable decrease in the classification models' performances (Multimedia Appendix 1). However, in general, the classification models continued presenting good

performances. For example, the KNN classification model presented the lowest accuracy score (77.42%) using the *RT-PCR balanced* data set. The remaining classification models, considering all data sets, presented accuracy scores between 80.15% and 97.58%. Depending on the preprocessed data set, the LR (weak/strong regularization) continued to be outperformed by the other algorithms. Presenting such scenarios is relevant to analyze how the algorithms behave when models are implemented with reduced reported symptoms.

In addition, by computing the AUROC using the RT-PCR, rapid, and both test scenarios, the trade-offs between sensitivity (true-positive rate) and probability (false-positive rate) were identified, evidencing the diagnostic abilities of the classification models when the discrimination threshold is varied (Figure 4). The classification models presented high discriminatory power for all scenarios, with the curves closer to each graphic representation's upper left corner. However, for such scenarios, the KNN and SVM classification models presented the lowest discriminatory power.

Given the three unbalanced data sets, there were more negative than positive COVID-19 cases. We computed the AUPRC to verify the classification models when handling the minority class, analyzing the trade-off between precision and recall for different decision thresholds (Figure 5). The AUPRC was summarized using the average precision (AP), as a weighted mean of precision. The *RT-PCR unbalanced* data set was mildly unbalanced, with a baseline AUPRC of 0.33. The *rapid unbalanced* data set was highly unbalanced, with a baseline AUPRC of 0.04. This was also the case for the *both unbalanced* data set, with a baseline AUPRC of 0.08. The DT and XGBoost achieved the best AP value (65%) using the *RT-PCR unbalanced* data set. For the remaining scenarios, the classification models presented AP values between 80% and 96%.

We also applied the Friedman and Nemenyi tests to improve confidence in evaluating the classification models, observing that the experiments' results were statistically significant. The classification models were compared over the 6 data sets using the Friedman test [27]. This comparison focused on the recall results due to the highly undesired impacts of false negatives in the COVID-19 application scenario (Figure 6). The null hypothesis was that all classification models are equivalent and have equal mean ranks. The tests resulted in a $P<.001$ for the *RT-PCR unbalanced* (t=307.16), *RT-PCR balanced* (t=328.72), *rapid unbalanced* (t=247.43), *rapid balanced* (t=239.20), *both unbalanced* (t=226.98), and *both balanced* (t=343.10) data sets. The results showed that the difference between the mean recall values was probably real ($P\leq.1$). The Friedman test ranked the classification models for each data set, resulting in an average rank for each classification model.

Based on the Friedman test results, the Nemenyi test [27] was applied to compare the classification models using the mean ranks. The critical difference (CD) between the classification models was verified using the Nemenyi test, with α=0.1. The CD is relevant to highlight if the classification models are separated by an interval less than the CD, meaning that the classification models were statistically indistinguishable. Thus, for most of the data sets, the difference between LRR/LR (statistically indistinguishable) and the other classification models was highlighted by the CD using the mean recall results (Multimedia Appendix 2). Depending on the data set, MLP and GBM were also statistically indistinguishable, as was the case of DT, RF, XGBoost, KNN, and SVM.

From the classification metrics results and the Friedman and Nemenyi tests (Figure 6), the top five features of the classification models with the highest performances (ie, MLP, GBM, DT, RF, XGBoost, and SVM) were ranked using the permutation feature importance method. Each average importance and SD values were presented for the DT-based classification models and the RT-PCR, rapid, and both types scenarios (Table 3). The average importance and SD information relate to reducing the feature importance when a feature is not considered. For example, according to the frequency of symptoms and the number of confirmed cases (Figure 3), *fever* showed higher average importance values for almost all scenarios than other reported symptoms. We also applied the permutation feature importance method for the unbalanced data sets (Multimedia Appendix 3).

We also present the results achieved using the permutation feature importance method for detailing the feature ranking for classifications with MLP and SVM models (Table 4). For example, similar to the DT-based classification models, *fever* presented higher average importance values for almost all test scenarios than other symptoms reported by patients. For such algorithms, we also present the average importance and SD for the unbalanced data sets (Multimedia Appendix 3).

Therefore, the top five most significant features vary depending on the algorithm used to implement the classification model (Table 5). For the *RT-PCR balanced* data set, all algorithms prioritized the same top two features (ie, *fever* and *gender*), slightly differing in the top three and top five, while, for the *rapid balanced* data set, all algorithms prioritized the same top two features (ie, *dyspnea* and *olfactory disorders*), also slightly different in the top three, top four, and top five. For the *both balanced* data set, the algorithms prioritized the top two features similar to the classifications with the *RT-PCR balanced* data set. We also applied the permutation feature importance method to rank features using the unbalanced data sets (Multimedia Appendix 3).

In addition, to improve the experiments conducted to assist the COVID-19 test prioritization, we combined the classification models to define voting ensemble models using the majority voting strategy (Multimedia Appendix 4). Two combinations of classification models were considered for each data set: DT-based models (ie, GBM, DT, RF, and XGBoost) and non-DT models (ie, MLP, SVM, KNN, LRR, and LR). In general, for the voting ensemble models implemented with the six data sets, the mean results of classification metrics using 10-fold cross-validation were similar to those of the MLP, GBM, DT, RF, XGBoost, KNN, SVM, LRR, and LR models (Table 2).

Figure 4. The models' ROC curves with (A) RT-PCR unbalanced, (B) RT-PCR balanced, (C) rapid unbalanced, (D) rapid balanced, (E) both unbalanced, and (F) both balanced. AUC: area under the receiver operating characteristic curve; GBM: gradient boosting machine; KNN: k-nearest neighbors; LR: logistic regression (weak regularization); LRR: logistic regression (strong regularization); Mlp: multilayer perceptron; ROC: receiver operating characteristic; RT-PCR: reverse transcription polymerase chain reaction; SVM: support vector machine; XGBoost: extreme gradient boosting.

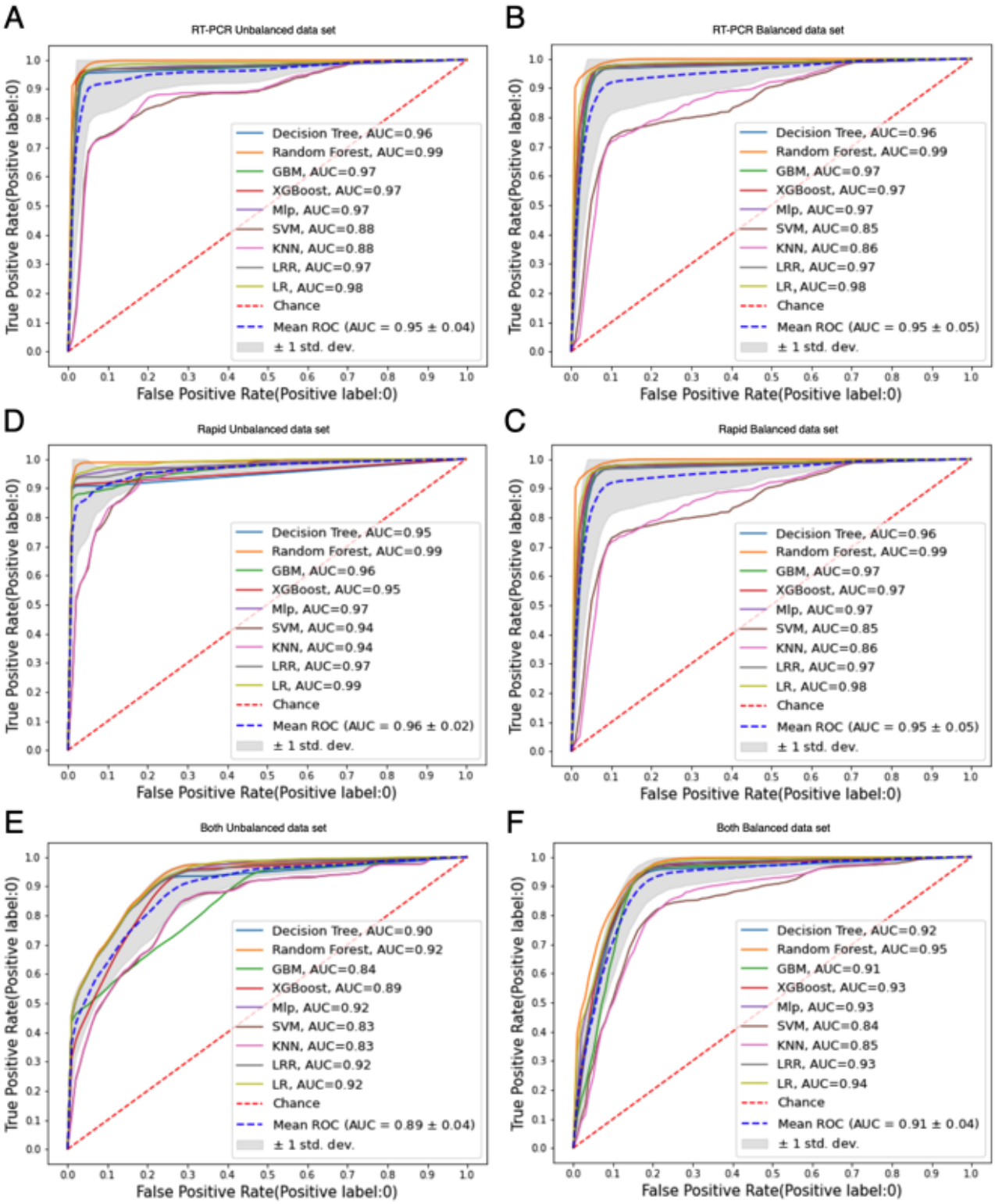


Figure 5. Models' precision-recall curve with (A) RT-PCR unbalanced data set, (B) rapid unbalanced data set, and (C) both unbalanced data set. AP: average precision; GBM: gradient boosting machine; KNN: k-nearest neighbors; LR: logistic regression (weak regularization); LRR: logistic regression (strong regularization); Mlp: multilayer perceptron; PR: precision-recall; RT-PCR: reverse transcription polymerase chain reaction; SVM: support vector machine; XGBoost: extreme gradient boosting.

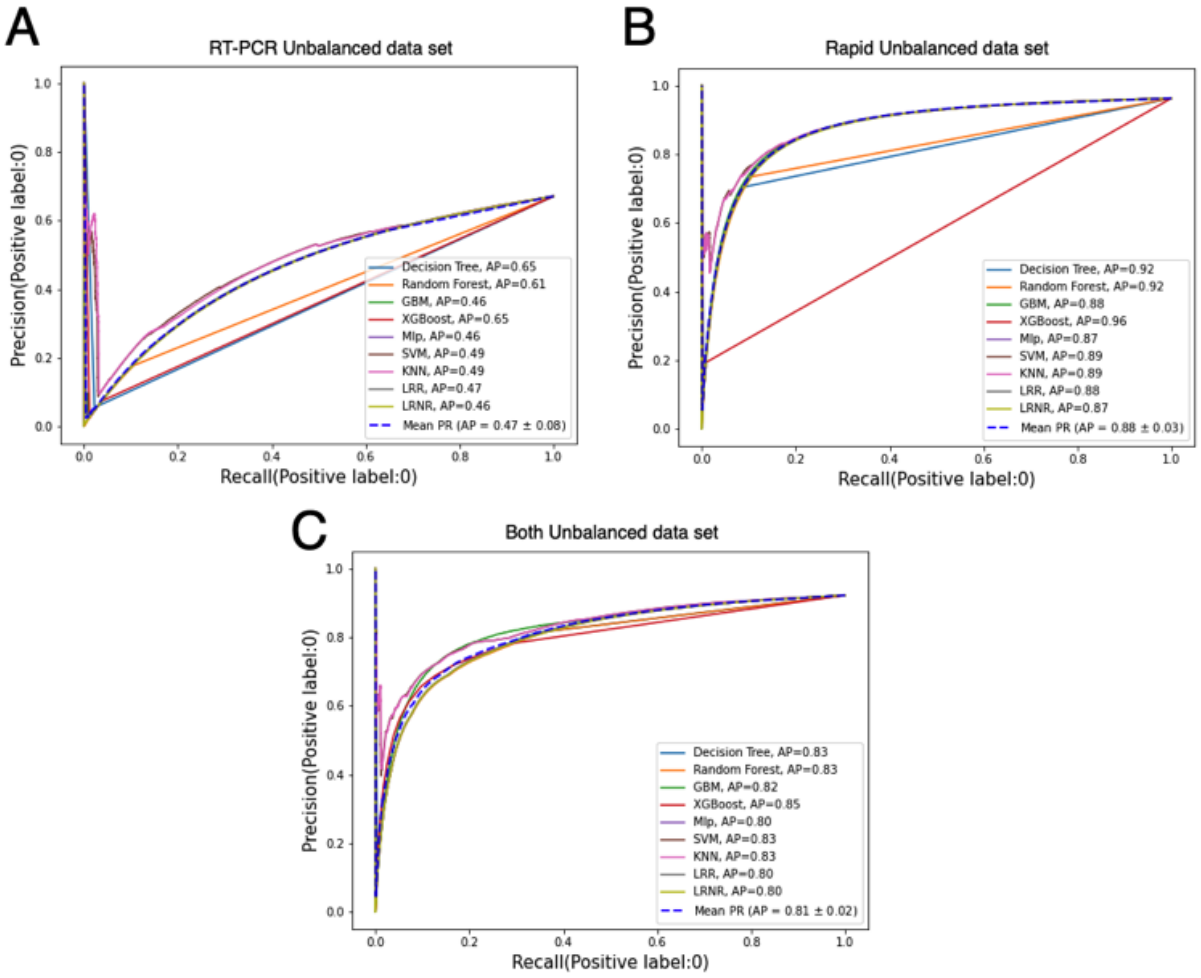


Figure 6. (A) The mean recall for the MLP, GBM, RF, DT, XGBoost, KNN, SVM, LRR, and LR classification models using the unbalanced data sets for RT-PCR, rapid, and both types. (B) The mean recall for the MLP, GBM, RF, DT, XGBoost, KNN, SVM, LRR, and LR classification models using the balanced data sets for RT-PCR, rapid, and both types. DT: decision tree; GBM: gradient boosting machine; KNN: k-nearest neighbors; LR: logistic regression (weak regularization); LRR: logistic regression (strong regularization); MLP: multilayer perceptron; RF: random forest; RT-PCR: reverse transcription polymerase chain reaction; SVM: support vector machine; XGBoost: extreme gradient boosting.

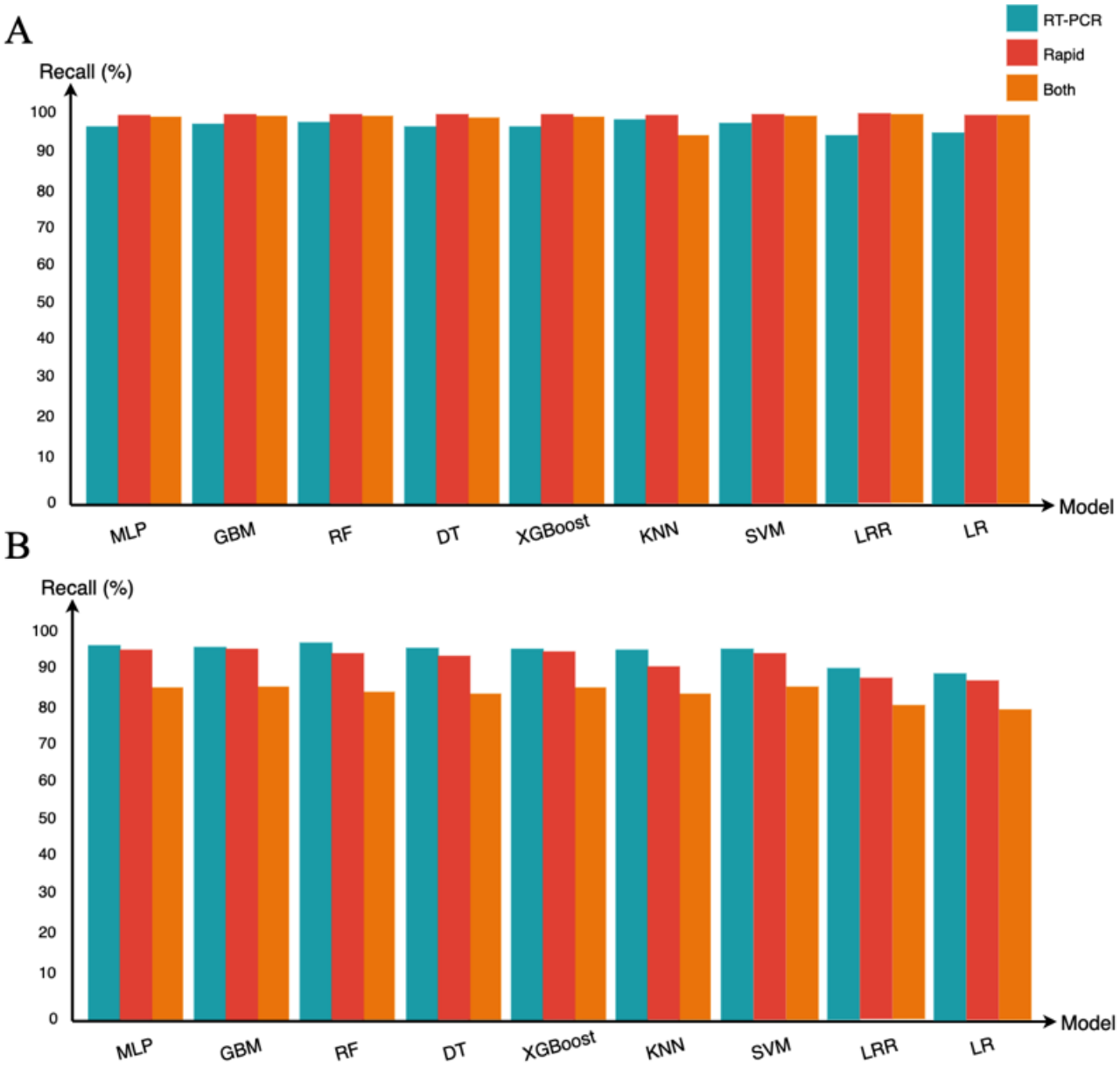


Table 3. The average importance and SD values for each feature for the decision tree–based classification models using the balanced data sets.

Data sets and features	GBM[a], mean (SD)	DT[b], mean (SD)	RF[c], mean (SD)	XGBoost[d], mean (SD)
RT-PCR[e] balanced				
Gender	0.233 (0.013)	0.245 (0.013)	0.238 (0.013)	0.233 (0.013)
Health professional	0.045 (0.005)	0.055 (0.005)	0.051 (0.005)	0.042 (0.005)
Fever	0.260 (0.012)	0.263 (0.012)	0.267 (0.013)	0.262 (0.013)
Sore throat	0.100 (0.007)	0.102 (0.007)	0.100 (0.007)	0.101 (0.006)
Dyspnea	0.096 (0.007)	0.100 (0.007)	0.098 (0.007)	0.092 (0.007)
Olfactory disorders	0.026 (0.004)	0.029 (0.003)	0.014 (0.004)	0.016 (0.003)
Cough	0.087 (0.007)	0.096 (0.007)	0.084 (0.007)	0.082 (0.008)
Coryza	0.026 (0.004)	0.027 (0.004)	0.022 (0.004)	0.014 (0.003)
Taste disorders	0.030 (0.004)	0.040 (0.004)	0.027 (0.004)	0.024 (0.003)
Headache	0.021 (0.004)	0.024 (0.005)	0.018 (0.004)	0.002 (0.003)
Rapid balanced				
Gender	0.135 (0.009)	0.109 (0.009)	0.122 (0.010)	0.123 (0.010)
Health professional	0.026 (0.005)	0.027 (0.005)	0.020 (0.004)	0.027 (0.005)
Fever	0.120 (0.012)	0.124 (0.012)	0.114 (0.011)	0.109 (0.010)
Sore throat	0.019 (0.005)	0.030 (0.005)	0.023 (0.004)	0.013 (0.004)
Dyspnea	0.184 (0.012)	0.179 (0.012)	0.187 (0.013)	0.179 (0.013)
Olfactory disorders	0.175 (0.012)	0.178 (0.013)	0.180 (0.014)	0.154 (0.012)
Cough	0.076 (0.009)	0.084 (0.011)	0.080 (0.008)	0.080 (0.008)
Coryza	0.090 (0.008)	0.092 (0.009)	0.087 (0.007)	0.052 (0.005)
Taste disorders	0.078 (0.008)	0.081 (0.009)	0.053 (0.007)	0.071 (0.007)
Headache	0.035 (0.007)	0.030 (0.007)	0.035 (0.006)	0.035 (0.007)
Both balanced				
Gender	0.159 (0.007)	0.160 (0.007)	0.153 (0.007)	0.156 (0.007)
Health professional	0.025 (0.004)	0.024 (0.004)	0.023 (0.004)	0.025 (0.004)
Fever	0.211 (0.010)	0.215 (0.011)	0.213 (0.010)	0.209 (0.010)
Sore throat	0.080 (0.005)	0.081 (0.005)	0.082 (0.005)	0.078 (0.005)
Dyspnea	0.077 (0.006)	0.075 (0.005)	0.073 (0.005)	0.076 (0.005)
Olfactory disorders	0.059 (0.006)	0.050 (0.005)	0.050 (0.005)	0.046 (0.005)
Cough	0.060 (0.005)	0.058 (0.005)	0.054 (0.005)	0.060 (0.005)
Coryza	0.047 (0.004)	0.042 (0.003)	0.040 (0.003)	0.044 (0.004)
Taste disorders	0.063 (0.006)	0.079 (0.006)	0.072 (0.005)	0.069 (0.005)
Headache	0.042 (0.005)	0.046 (0.005)	0.044 (0.005)	0.045 (0.005)

[a]GBM: gradient boosting machine.

[b]RF: random forest.

[c]DT: decision tree.

[d]XGBoost: extreme gradient boosting.

[e]RT-PCR: reverse transcription polymerase chain reaction.

Table 4. The average importance and SD for each feature for the MLP and SVM models and the balanced data sets.

Data sets and features	MLP[a], mean (SD)	SVM[b], mean (SD)
RT-PCR[c] balanced		
Gender	0.236 (0.013)	0.230 (0.013)
Health professional	0.048 (0.005)	0.042 (0.005)
Fever	0.262 (0.013)	0.257 (0.013)
Sore throat	0.103 (0.006)	0.098 (0.006)
Dyspnea	0.096 (0.007)	0.088 (0.007)
Olfactory disorders	0.027 (0.004)	0.015 (0.004)
Cough	0.084 (0.008)	0.078 (0.007)
Coryza	0.025 (0.004)	0.013 (0.004)
Taste disorders	0.031 (0.004)	0.020 (0.004)
Headache	0.023 (0.003)	0.002 (0.003)
Rapid balanced		
Gender	0.120 (0.009)	0.117 (0.010)
Health professional	0.033 (0.006)	0.029 (0.005)
Fever	0.115 (0.010)	0.105 (0.011)
Sore throat	0.012 (0.005)	0.023 (0.004)
Dyspnea	0.177 (0.013)	0.177 (0.014)
Olfactory disorders	0.157 (0.012)	0.149 (0.012)
Cough	0.082 (0.009)	0.076 (0.008)
Coryza	0.064 (0.006)	0.058 (0.006)
Taste disorders	0.072 (0.007)	0.055 (0.006)
Headache	0.036 (0.006)	0.028 (0.005)
Both balanced		
Gender	0.161 (0.007)	0.154 (0.007)
Health professional	0.025 (0.004)	0.024 (0.004)
Fever	0.207 (0.010)	0.193 (0.009)
Sore throat	0.084 (0.005)	0.075 (0.005)
Dyspnea	0.078 (0.006)	0.088 (0.006)
Olfactory disorders	0.055 (0.006)	0.049 (0.005)
Cough	0.062 (0.005)	0.071 (0.005)
Coryza	0.035 (0.003)	0.046 (0.003)
Taste disorders	0.071 (0.006)	0.068 (0.005)
Headache	0.051 (0.005)	0.045 (0.005)

[a]MLP: multilayer perceptron.

[b]SVM: support vector machine.

[c]RT-PCR: reverse transcription polymerase chain reaction.

Table 5. The five most significant factors for COVID-19 test prioritization using the classification models with the highest performances and the data sets.

Data sets and models	Top one	Top two	Top three	Top four	Top five
RT-PCR[a] balanced					
MLP[b]	Fever	Gender	Sore throat	Dyspnea	Cough
GBM[c]	Fever	Gender	Sore throat	Dyspnea	Cough
RF[d]	Fever	Gender	Sore throat	Dyspnea	Cough
DT[e]	Fever	Gender	Sore throat	Dyspnea	Cough
XGBoost[f]	Fever	Gender	Sore throat	Dyspnea	Cough
SVM[g]	Fever	Gender	Sore throat	Dyspnea	Cough
Rapid balanced					
MLP	Dyspnea	Olfactory disorders	Gender	Fever	Cough
GBM	Dyspnea	Olfactory disorders	Gender	Fever	Coryza
RF	Dyspnea	Olfactory disorders	Gender	Fever	Coryza
DT	Dyspnea	Olfactory disorders	Fever	Gender	Coryza
XGBoost	Dyspnea	Olfactory disorders	Gender	Fever	Cough
SVM	Dyspnea	Olfactory disorders	Gender	Fever	Cough
Both balanced					
MLP	Fever	Gender	Sore throat	Dyspnea	Taste disorders
GBM	Fever	Gender	Sore throat	Dyspnea	Taste disorders
RF	Fever	Gender	Sore throat	Dyspnea	Taste disorders
DT	Fever	Gender	Sore throat	Taste disorders	Dyspnea
XGBoost	Fever	Gender	Sore throat	Dyspnea	Taste disorders
SVM	Fever	Gender	Dyspnea	Sore throat	Cough

[a]RT-PCR: reverse transcription polymerase chain reaction.

[b]MLP: multilayer perceptron.

[c]GBM: gradient boosting machine.

[d]RF: random forest.

[e]DT: decision tree.

[f]XGBoost: extreme gradient boosting.

[g]SVM: support vector machine.

Discussion

Principal Findings

The raw data set's data preprocessing enabled the implementation, validation, and comparison of classification models with different characteristics such as using neural layers, tree combinations, and calculating the distance between data. The preprocessing also resulted in the public data availability of patients who were symptomatic tested using RT-PCR and rapid tests [10]. Thus, the data sets can be reused by other studies to improve the state of the art.

The algorithms were trained and tested using the unbalanced and balanced data sets, improving data representativity. The best classification metrics results were related to the RT-PCR and rapid tests scenarios using unbalanced and balanced data. Although the classification models' performance was similar for the RT-PCR and rapid tests scenarios, the RT-PCR test scenario is the most clinically relevant one due to the RT-PCR testing's high confidence. The RT-PCR test's precision increases confidence in the diagnosis, even if the patient was tested in the first days after symptoms onset. For both test scenarios with unbalanced data, although presenting a low Brier Score and high precision, accuracy score, and recall, the classification models presented a lower AUROC because of the higher negative than positive COVID-19 cases. For both test scenarios with balanced data, the Brier score continued to be low. The precision, accuracy, and AUROC were higher; however, the recall results were slightly decreased if compared to the unbalanced data results.

The recall metric is relevant due to the undesired impacts of false negatives in clinical practice. Thus, we improved the classification models' quality of comparisons by applying the Friedman and Nemenyi tests based on the six data sets' recall.

We used such statistical comparison results for defining the MLP, GBM, DT, RF, XGBoost, and SVM as the classification models with the highest performances for COVID-19 test prioritization in Brazil.

Given the classification models with the highest performances and the five most significant features for COVID-19 test prioritization, the fever's importance as one of the top two features is according to the aforementioned statistics (Figure 3). The statistics showed that fever was the second most frequent symptom reported by patients who were symptomatic, confirmed as COVID-19 cases. *Gender* and *dyspnea* were also among the highest-ranked features used by classification models. For example, for the *RT-PCR balanced* data set, observing the DT model's decision rules to get an overview of the role of gender in classifications, positive or negative decisions for males and females differed based on reported symptoms and the *health professional* feature. However, further investigation about the role of gender in classifications is recommended for future works.

Therefore, secondary RQ 1 was answered by showing that *gender* and *health professional* features are related to relevant demographic characteristics to support the COVID-19 test prioritization in Brazil (Tables 4 and 5). Secondary RQ 2 was also answered, showing that fever, sore throat, dyspnea, olfactory disorders, cough, coryza, taste disorders, and headache are relevant symptoms.

All DT-based classification models considered in this study are among the classification models with the highest performances, grouped based on the results of classification metrics and statistical tests. This fact is relevant due to the high levels of DTs' interpretability, positively impacting health care professionals' final decision making. In clinical practice, ML-based applications' acceptance increases when health care professionals can easily understand and interpret classification models' outputs to track decision-making logic [28]. Given the grouping of models with similar performances, we used the criterion of easy interpretability to answer secondary RQ 3. Thus, the DT classification model was considered the most suitable for COVID-19 test prioritization in Brazil. We configured the model with the Gini impurity criterion, best split strategy, no maximum depth, a minimum number of two samples split and one sample leaf, no minimum weighted fraction leaves and no impurity decrease and split, unlimited number of features and leaves, global random state instance, no class weight, and no pruning. As one of the classification models with the highest performances, DT provides a simple tree representation of the decision making, enabling a unique tree's straightforward interpretation by health care professionals.

To answer secondary RQ 4, we analyzed the DT model's classification results, observing that a considerable fraction of the incorrectly classified instances occurred when patients reported only one, two, or three symptoms. Furthermore, we conducted an experiment to verify the impacts of reducing features in the performance of the implemented classification models (Multimedia Appendix 1). For example, with the *both RT-PCR balanced* data set, when the symptoms of sore throat, dyspnea, headache, and coryza were not considered to implement the DT classification model, the performance results decreased considerably. This reduces the ability of the model to distinguish between positive and negative cases.

Although the DT is considered the most suitable model, all the other classification models that presented high performance were relevant to address COVID-19 test prioritization. In Brazil, due to other epidemics (eg, dengue fever [29]), many people report symptoms that may or may not be related to COVID-19. As a limited-income country, Brazil also has inefficient testing strategies such as shortages of COVID-19 tests. One of the available classification models can be applied for COVID-19 test prioritization during primary health care, with a mean accuracy score of at least 88.63%.

Comparison With Prior Work

The relevance of research addressing viral infection outbreaks is evidenced from the public administration (eg, surveillance systems) to the diagnosis viewpoint. For example, Son et al [30] used a South Korean time series of influenza incidence for early outbreak detection, aiming to assist the definition of control policies. Chatterjee et al [31] analyzed COVID-19 data sets to identify risks of spreading, identify correlated factors associated with the disease's spread, identify the impact of social isolation, and experiment with univariate long short-term memory models for forecasting of total cases and total deaths. In general, infectious disease research is guided by trends in data analytics [32].

Indeed, the COVID-19 pandemic is an example of a problematic scenario. Kumar [33] applied cluster analysis to study and improve the monitoring of SARS-CoV-2 infections in India, providing insights on clusters of affected Indian states and union territories. Besides aiming to improve the management of available resources, Khakharia et al [34] developed outbreak classification models for COVID-19 using data sets with information about patients who live in India, Bangladesh, the Democratic Republic of Congo, Pakistan, China, Philippines, Germany, Indonesia, Ethiopia, and Nigeria. Vaid et al [35] implemented and validated models (eg, XGBoost) to predict mortality and critical events using electronic health records of patients who tested positive for COVID-19 in New York City.

To assist COVID-19 detection, Brinati et al [36] validated models implemented using DT, extremely randomized trees, KNN, LR, naive Bayes, RF, and three-way RF algorithms. The authors considered COVID-19 detection using routine blood exams, gender, and age. The accuracy of the models ranged between 82% and 86%. However, the large number of required blood exams (ie, 13) was a limitation, which may compromise this approach's feasibility in low- and middle–income countries.

Ahamad et al [21] used a Chinese data set to assist the COVID-19 detection considering symptoms (ie, fever, cough, pneumonia, lung infection, coryza, muscle soreness, and diarrhea), gender, age, travel history, and isolation. The authors validated the XGBoost, SVM, DT, RF, and GBM models. XGBoost presented the highest accuracy with more than 85%, varying according to age. However, lung infection use, detected by chest images, increases costs and may limit the disease's rapid screening.

Aiming to improve confidence in screening COVID-19, Mei et al [37] used computerized tomography (CT) images along with symptoms (ie, fever, cough, and cough with sputum), exposure history, laboratory testing (ie, white blood cells, neutrophils, percentage neutrophils, lymphocytes, and percentage lymphocytes), age, gender, and temperature. They applied the deep convolutional neural network to analyze images, besides comparing the performance of SVM, RF, and MLP models, showing that MLP presented the highest accuracy score. Afterward, the authors combined images and clinical information. Similarly, requiring images increases costs and may limit the rapid screening of COVID-19 in low- and middle–income countries.

Finally, Zoabi et al [4] used gender, age, symptoms (eg, cough, fever, sore throat, shortness of breath, and headache), and contact with a confirmed case to classify positive and negative COVID-19 cases. The authors implemented a GBM model based on data reported by the Israeli Ministry of Health. The GBM model presented an AUROC of 86% and 90% using a reduced set of features and the complete set, respectively. Similar to our study, the authors reported the high importance of gender during the classifications. We also improved the state of the art by presenting a comparison of other implementations of classification models. Besides cough, fever, sore throat, shortness of breath, and headache, we used the symptoms of olfactory disorders, coryza, and taste disorders to improve the results.

In contrast to such prior works, we focused on raw data from 55,676 Brazilians and used features that do not require expensive exams such as CT images and blood tests. Symptoms included fever, sore throat, dyspnea, olfactory disorders, cough, coryza, taste disorders, and headache. The *gender* and *health professional* features were the additional information required to conduct the COVID-19 test prioritization using the classification models. *Gender* was also used as a feature by prior works [4,20,36,37]. The use of exams such as CT images and blood tests limits classification models' application scenarios because it is necessary to prioritize patients who are symptomatic for testing in the first days after symptoms onset.

Limitations

By preprocessing the 55,676 raw data, the *RT-PCR balanced* data set only included 1832 patients who were symptomatic, the *rapid balanced* data set included 1296 patients who were symptomatic, and the *both balanced* data set included 3128 patients who were symptomatic. However, to improve the strength of results and decrease size limitation, we also considered 3 unbalanced data sets. For example, the *both unbalanced* data set was composed of 20,021 patients who were symptomatic and tested for COVID-19.

Furthermore, in a real-world scenario, the number of patients who were asymptomatic with COVID-19 can also be considered a limitation to the classification models' applicability. In this case, this study continues to be relevant due to the remaining symptomatic cases that also require health care professionals and the government's attention. The evaluation of patients who are symptomatic is also relevant to prevent the unplanned use of COVID-19 testing resources due to other disease outbreaks in Brazil caused by other viral infections (eg, dengue, Zika, and chikungunya). Such viral infections present similar symptoms that may complicate health care professionals' decision on the adequate testing type needed.

The reduced number of symptoms reported by a patient who is symptomatic can also negatively impact the reuse classification models. Nevertheless, the feature ranking and other information (eg, contact with infected people) are relevant to complement the classification models during the decision making conducted by health care professionals and policy makers. We verified the impacts of reducing features in the performance of implemented classification models (Multimedia Appendix 1).

Finally, the number of classification models implemented, validated, and compared is another limitation of our study, given the wide variety of available algorithms and ensemble strategies. This limitation was reduced by selecting well-known algorithms based on trees, linear regression, statistical learning, distance, and the concept of neurons.

Clinical Practice Context

The availability of eHealth and mHealth systems is relevant to assist decision making in different scenarios. One such scenario is detecting COVID-19 in patients who reside in remote and hard-to-reach locations (eg, Amazonia or Latin America) [38]. Developers can integrate eHealth and mHealth systems with services that enable health care professionals to be alerted when the risk of disease is detected. The use of eHealth and mHealth systems should be encouraged, considering that the early detection of COVID-19 is essential in clinical practice to enable early medical attention, possibly reducing the negative impacts of late treatments. This type of eHealth and mHealth system can also benefit public health systems when factors related to the human condition (eg, fatigue and lack of experience) and the collapse of health services negatively influence health care professionals' decision making during patients' evaluation. Such scenarios are authentic in the context of the COVID-19 pandemic [39].

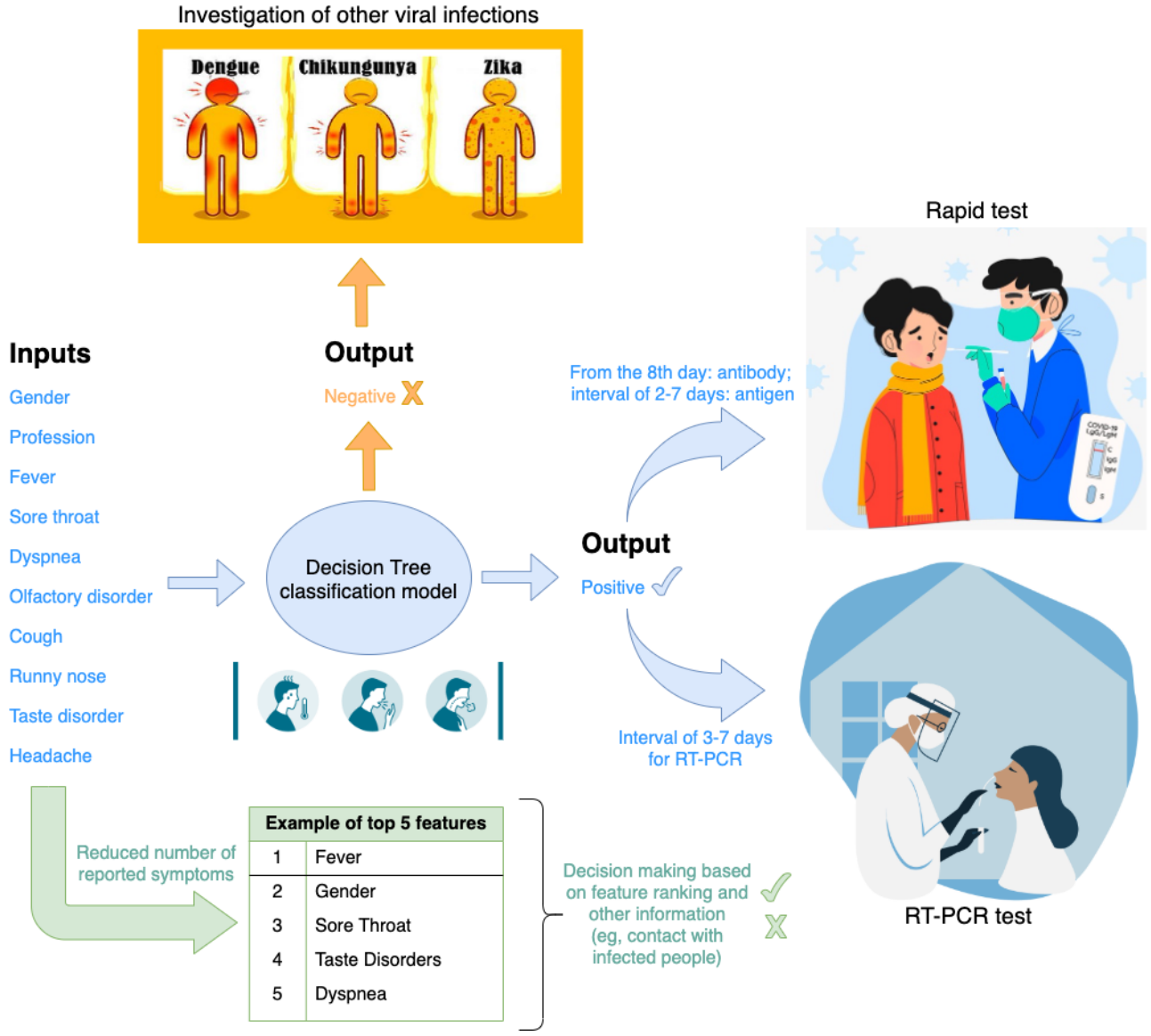
Therefore, the implemented classification models can be the basis for eHealth and mHealth systems to support health care professionals and policy makers during the COVID-19 test prioritization. To be applied in clinical practice and integrated with the current clinical workflow, the availability of the DT classification model and the use of feature ranking information through web services to be consumed by an eHealth or mHealth system is recommended. Such a system shall present classification results for health care professionals in a user-friendly manner. The straightforward interpretation of classification models is relevant to increase health care professionals' confidence in classification results. For example, the web services can be integrated with Brazilian public health facilities' systems to prioritize the reduced COVID-19 testing resources.

We present an application scenario integrating a clinical workflow and the DT classification model (Figure 7). The DT is used to prioritize patients who are symptomatic for COVID-19 testing. However, when the number of reported symptoms is too low, the classification models cannot distinguish between

positive and negative cases. In this case, health care professionals can reuse the feature ranking and other information (eg, contact with infected people) to make decisions about COVID-19 testing. Thus, the use of feature ranking information is guided by the answer of secondary RQ 4. If the result is not prioritized, the patient's clinical condition should be further investigated in regard to other viral diseases.

For the application scenario, there are five possible flows: (1) confirmed case with classification model and rapid test result, (2) confirmed case with classification model and RT-PCR test result, (3) confirmed case using feature ranking and rapid test result, (4) confirmed case using feature ranking and RT-PCR test result, and (5) negative case with the recommendation of investigation of other viral diseases. It is relevant to consider the days between the onset of symptoms and COVID-19 testing: closed interval of 3-7 days for RT-PCR test, from the eighth day for the rapid antibody test, and closed interval of 2-7 days for the rapid antigen test [40-42].

Figure 7. An application scenario to connect the decision tree classification model with a clinical workflow. The model guides the test prioritization of patients who were symptomatic suspected of COVID-19. RT-PCR: reverse transcription polymerase chain reaction.



Conclusions

The results showed the relevance of using classification models for COVID-19 test prioritization in Brazil, mainly based on the symptoms that do not require expensive exams. By comparing the classification models using raw data from 55,676 Brazilians, the 10-fold cross-validation method, classification metrics, and the Friedman and Nemenyi tests, the MLP, GBM, DT, RF, XGBoost, and SVM presented the highest performances with similar results.

DT-based classification models' high performances are relevant for our application scenario due to the high levels of DTs' interpretability, positively impacting health care professionals' final decision making. Therefore, applying the easy interpretability as an additional comparison criterion, DT was considered the most suitable classification model, effectively assisting in the decision making for prioritizing patients who are symptomatic for testing. Information about the features *gender*, *health professional*, *fever*, *sore throat*, *dyspnea*, *olfactory disorders*, *cough*, *coryza*, *taste disorders*, and

headache enable the COVID-19 test prioritization for patients who are symptomatic. The use of symptoms that do not require expensive exams contributes to assisting patients who live, for example, in needy and hard-to-reach communities. The results of feature ranking reported in this paper are also relevant to support a more detailed analysis in a scenario where a patient reports a reduced number of symptoms.

To improve testing prioritization, we plan to investigate the relationship between the symptoms reported by patients with COVID-19 and other widespread diseases in Brazil, such as dengue fever, Zika fever, and chikungunya. Thus, we aim to include implementing and validating classification models and developing and validating an eHealth system to support health care professionals and policy makers in decision making for testing strategies.

Acknowledgments

The authors thank the support of the Conselho Nacional de Desenvolvimento Científico e Tecnológico - Brazil; Federal University of the Agreste of Pernambuco, Coordenacão de Aperfeiçamento de Pessoal de Nível Superior - Brazil; and Programa de Pós-Graduação em Engenharia Elétrica, Federal University of Campina Grande. This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brazil: Project 88881.507204/2020-01.

The data sets generated or analyzed during this study are available in the data set of Brazilian patients who were symptomatic for screening the risk of COVID-19 [10]. Furthermore, the source codes are available in a GitHub COVID-19 repository [12].

Authors' Contributions

ÍVSS contributed to implementing the classification models, conducting the practical experimentation and validation of the models, and writing the paper. ÁS, LCS, LDS, EC, and AP contributed to the study's conception, research methodology, revision of implementation and validation results, and the paper's writing and revisions. ACMS and DFSS contributed to the revisions of the implementation, validation results, and paper.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Performance of classification models considering the chi-square results.
[DOCX File , 24 KB - jmir_v23i4e27293_app1.docx]

Multimedia Appendix 2
Results of statistical tests.
[DOCX File , 54 KB - jmir_v23i4e27293_app2.docx]

Multimedia Appendix 3
Feature ranking results for the unbalanced data set.
[DOCX File , 23 KB - jmir_v23i4e27293_app3.docx]

Multimedia Appendix 4
Mean values of classification metrics for the ensemble models.
[DOCX File , 16 KB - jmir_v23i4e27293_app4.docx]

References

1. Belard A, Buchman T, Forsberg J, Potter BK, Dente CJ, Kirk A, et al. Precision diagnosis: a view of the clinical decision support systems (CDSS) landscape through the lens of critical care. J Clin Monit Comput 2017 Apr;31(2):261-271. [doi: 10.1007/s10877-016-9849-1] [Medline: 26902081]
2. Elhoseny M, Abdelaziz A, Salama AS, Riad A, Muhammad K, Sangaiah AK. A hybrid model of Internet of Things and cloud computing to manage big data in health services applications. Future Generation Computer Syst 2018 Sep;86:1383-1394. [doi: 10.1016/j.future.2018.03.005]
3. Chatterjee A, Gerdes MW, Martinez S. eHealth initiatives for the promotion of healthy lifestyle and allied implementation difficulties. 2019 Presented at: International Conference on Wireless and Mobile Computing, Networking and Communications; October 21-23, 2019; Barcelona, Spain p. 1-8. [doi: 10.1109/wimob.2019.8923324]
4. Zoabi Y, Deri-Rozov S, Shomron N. Machine learning-based prediction of COVID-19 diagnosis based on symptoms. NPJ Digit Med 2021 Jan 04;4(1):3. [doi: 10.1038/s41746-020-00372-6] [Medline: 33398013]

5. Guimarães VHA, de Oliveira-Leandro M, Cassiano C, Marques ALP, Motta C, Freitas-Silva AL, et al. Knowledge about COVID-19 in Brazil: cross-sectional web-based study. JMIR Public Health Surveill 2021 Jan 21;7(1):e24756 [FREE Full text] [doi: 10.2196/24756] [Medline: 33400684]
6. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet 2020 Feb 15;395(10223):497-506 [FREE Full text] [doi: 10.1016/S0140-6736(20)30183-5] [Medline: 31986264]
7. Veiga E Silva L, de Andrade Abi Harb MDP, Teixeira Barbosa Dos Santos AM, de Mattos Teixeira CA, Macedo Gomes VH, Silva Cardoso EH, et al. COVID-19 mortality underreporting in Brazil: analysis of data from government internet portals. J Med Internet Res 2020 Aug 18;22(8):e21413 [FREE Full text] [doi: 10.2196/21413] [Medline: 32730219]
8. Ferrante L, Steinmetz WA, Almeida ACL, Leão J, Vassão RC, Tupinambás U, et al. Brazil's policies condemn Amazonia to a second wave of COVID-19. Nat Med 2020 Sep;26(9):1315. [doi: 10.1038/s41591-020-1026-x] [Medline: 32770168]
9. Monteiro de Oliveira M, Fuller TL, Brasil P, Gabaglia CR, Nielsen-Saines K. Controlling the COVID-19 pandemic in Brazil: a challenge of continental proportions. Nat Med 2020 Oct;26(10):1505-1506 [FREE Full text] [doi: 10.1038/s41591-020-1071-5] [Medline: 32887951]
10. dos Santos Santana Í, da Silveira ACM, Sobrinho A, Chaves e Silva L, da Silva LD, de Souza Santos DF, et al. A Brazilian dataset of symptomatic patients for screening the risk of COVID-19. Mendeley Data. 2021. URL: https://data.mendeley.com/datasets/b7zcgmmwx4/5 [accessed 2021-04-05]
11. Malta M, Murray L, da Silva CMFP, Strathdee SA. Coronavirus in Brazil: the heavy weight of inequality and unsound leadership. EClinicalMedicine 2020 Aug;25:100472 [FREE Full text] [doi: 10.1016/j.eclinm.2020.100472] [Medline: 32754684]
12. COVID-19-BRAZIL-ORGANIZATION / COVID-19. GitHub. URL: http://bit.ly/2Met3S4 [accessed 2021-04-05]
13. Chatterjee A, Gerdes MW, Martinez SG. Identification of risk factors associated with obesity and overweight-a machine learning overview. Sensors (Basel) 2020 May 11;20(9):2734 [FREE Full text] [doi: 10.3390/s20092734] [Medline: 32403349]
14. Gender and COVID-19: advocacy brief, 14 May 2020. World Health Organization. 2020. URL: https://apps.who.int/iris/handle/10665/332080 [accessed 2021-04-05]
15. Peckham H, de Gruijter NM, Raine C, Radziszewska A, Ciurtin C, Wedderburn LR, et al. Male sex identified by global COVID-19 meta-analysis as a risk factor for death and ITU admission. Nat Commun 2020 Dec 09;11(1):6317. [doi: 10.1038/s41467-020-19741-6] [Medline: 33298944]
16. Shilaskar S, Ghatol A. Diagnosis system for imbalanced multi-minority medical dataset. Soft Computing 2018 Apr 3;23(13):4789-4799. [doi: 10.1007/s00500-018-3133-x]
17. Almansour NA, Syed HF, Khayat NR, Altheeb RK, Juri RE, Alhiyafi J, et al. Neural network and support vector machine for the prediction of chronic kidney disease: a comparative study. Comput Biol Med 2019 Jun;109:101-111. [doi: 10.1016/j.compbiomed.2019.04.017] [Medline: 31054385]
18. Biau G, Cadre B, Rouvière L. Accelerated gradient boosting. Machine Learning 2019 Feb 4;108(6):971-992. [doi: 10.1007/s10994-019-05787-1]
19. Friedman JH. Greedy function approximation: a gradient boosting machine. Ann Stat 2001 Oct 1;29(5):1189-1232. [doi: 10.1214/aos/1013203451]
20. Xing W, Bei Y. Medical health big data classification based on KNN classification algorithm. IEEE Access 2020;8:28808-28819. [doi: 10.1109/access.2019.2955754]
21. Ahamad MM, Aktar S, Rashed-Al-Mahfuz M, Uddin S, Liò P, Xu H, et al. A machine learning model to identify early stage symptoms of SARS-Cov-2 infected patients. Expert Syst Appl 2020 Dec 01;160:113661 [FREE Full text] [doi: 10.1016/j.eswa.2020.113661] [Medline: 32834556]
22. Ma B, Meng F, Yan G, Yan H, Chai B, Song F. Diagnostic classification of cancers using extreme gradient boosting algorithm and multi-omics data. Comput Biol Med 2020 Jun;121:103761. [doi: 10.1016/j.compbiomed.2020.103761] [Medline: 32339094]
23. Sarica A, Cerasa A, Quattrone A. Random forest algorithm for the classification of neuroimaging data in Alzheimer's disease: a systematic review. Front Aging Neurosci 2017;9:329. [doi: 10.3389/fnagi.2017.00329] [Medline: 29056906]
24. Schober P, Vetter T. Logistic regression in medical research. Anesth Analg 2021 Feb 01;132(2):365-366 [FREE Full text] [doi: 10.1213/ANE.0000000000005247] [Medline: 33449558]
25. scikit-learn. URL: https://scikit-learn.org/stable/ [accessed 2021-02-21]
26. Ferri C, Hernández-Orallo J, Modroiu R. An experimental comparison of performance measures for classification. Pattern Recognition Lett 2009 Jan;30(1):27-38. [doi: 10.1016/j.patrec.2008.08.010]
27. Demsar J. Statistical comparisons of classifiers over multiple data sets. J Machine Learning Res 2006;7:1-30.
28. The Lancet Respiratory Medicine. Opening the black box of machine learning. Lancet Respir Med 2018 Nov;6(11):801. [doi: 10.1016/S2213-2600(18)30425-9] [Medline: 30343029]
29. Is it dengue or is it COVID-19? Centers for Disease Control and Prevention. 2020. URL: http://bit.ly/35u3t2p [accessed 2021-02-21]
30. Son W, Park J, Kwon O. Early detection of influenza outbreak using time derivative of incidence. EPJ Data Sci 2020;9(1):28 [FREE Full text] [doi: 10.1140/epjds/s13688-020-00246-7] [Medline: 32934899]

31. Chatterjee A, Gerdes MW, Martinez SG. Statistical explorations and univariate timeseries analysis on COVID-19 datasets to understand the trend of disease spreading and death. Sensors (Basel) 2020 May 29;20(11):3089 [FREE Full text] [doi: 10.3390/s20113089] [Medline: 32486055]
32. Kasson PM. Infectious disease research in the era of big data. Annu Rev Biomed Data Sci 2020 Jul 20;3(1):43-59. [doi: 10.1146/annurev-biodatasci-121219-025722]
33. Kumar S. Monitoring novel corona virus (COVID-19) infections in India by cluster analysis. Ann Data Sci 2020 May 19;7(3):417-425. [doi: 10.1007/s40745-020-00289-7]
34. Khakharia A, Shah V, Jain S, Shah J, Tiwari A, Daphal P, et al. Outbreak prediction of COVID-19 for dense and populated countries using machine learning. Ann Data Sci 2020 Oct 16;8(1):1-19. [doi: 10.1007/s40745-020-00314-9]
35. Vaid A, Somani S, Russak AJ, De Freitas JK, Chaudhry FF, Paranjpe I, et al. Machine learning to predict mortality and critical events in a cohort of patients with COVID-19 in New York City: model development and validation. J Med Internet Res 2020 Nov 06;22(11):e24018 [FREE Full text] [doi: 10.2196/24018] [Medline: 33027032]
36. Brinati D, Campagner A, Ferrari D, Locatelli M, Banfi G, Cabitza F. Detection of COVID-19 infection from routine blood exams with machine learning: a feasibility study. J Med Syst 2020 Jul 01;44(8):135 [FREE Full text] [doi: 10.1007/s10916-020-01597-4] [Medline: 32607737]
37. Mei X, Lee H, Diao K, Huang M, Lin B, Liu C, et al. Artificial intelligence-enabled rapid diagnosis of patients with COVID-19. Nat Med 2020 Aug;26(8):1224-1228 [FREE Full text] [doi: 10.1038/s41591-020-0931-3] [Medline: 32427924]
38. Sobrinho A, da Silva LD, Perkusich A, Pinheiro ME, Cunha P. Design and evaluation of a mobile application to assist the self-monitoring of the chronic kidney disease in developing countries. BMC Med Inform Decis Mak 2018 Jan 12;18(1):7 [FREE Full text] [doi: 10.1186/s12911-018-0587-9] [Medline: 29329530]
39. Ting DSW, Carin L, Dzau V, Wong TY. Digital technology and COVID-19. Nat Med 2020 Apr;26(4):459-461 [FREE Full text] [doi: 10.1038/s41591-020-0824-5] [Medline: 32284618]
40. Lima FET, Albuquerque NLSD, Florencio SDSG, Fontenele MGM, Queiroz APO, Lima GA, et al. Time interval between onset of symptoms and COVID-19 testing in Brazilian state capitals, August 2020. Epidemiol Serv Saude 2020;30(1):e2020788 [FREE Full text] [doi: 10.1590/S1679-4974202100010002] [Medline: 33331603]
41. Zhao J, Yuan Q, Wang H, Liu W, Liao X, Su Y, et al. Antibody responses to SARS-CoV-2 in patients with novel coronavirus disease 2019. Clin Infect Dis 2020 Nov 19;71(16):2027-2034 [FREE Full text] [doi: 10.1093/cid/ciaa344] [Medline: 32221519]
42. Burki TK. Testing for COVID-19. Lancet Respir Med 2020 Jul;8(7):e63-e64. [doi: 10.1016/s2213-2600(20)30247-2]

Abbreviations

AP: average precision
AUPRC: area under the precision-recall curve
AUROC: area under the receiver operating characteristic curve
CD: critical difference
CT: computerized tomography
DT: decision tree
GBM: gradient boosting machine
KNN: k-nearest neighbors
LR: logistic regression
LRR: logistic regression (strong regularization)
mHealth: mobile health
ML: machine learning
MLP: multilayer perceptron
RF: random forest
RQ: research question
RT-PCR: reverse transcriptase polymerase chain reaction
SVM: support vector machine
XGBoost: extreme gradient boosting

Edited by C Basch; submitted 25.01.21; peer-reviewed by X Cheng, A Chatterjee, D Huang; comments to author 16.02.21; revised version received 08.03.21; accepted 21.03.21; published 08.04.21.

Please cite as:
Viana dos Santos Santana Í, CM da Silveira A, Sobrinho Á, Chaves e Silva L, Dias da Silva L, Santos DFS, Gurjão EC, Perkusich A Classification Models for COVID-19 Test Prioritization in Brazil: Machine Learning Approach
J Med Internet Res 2021;23(4):e27293
URL: https://www.jmir.org/2021/4/e27293
doi:10.2196/27293
PMID:33750734

Original Paper

Evolving Epidemiological Characteristics of COVID-19 in Hong Kong From January to August 2020: Retrospective Study

Kin On Kwok[1,2,3], PhD; Wan In Wei[1], MPhil; Ying Huang[1], MSc; Kai Man Kam[1,2], MB; Emily Ying Yang Chan[1,4], MD; Steven Riley[5], PhD; Ho Hin Henry Chan[1], PhD; David Shu Cheong Hui[2,6], MD; Samuel Yeung Shan Wong[1], MD; Eng Kiong Yeoh[7], MBBS

[1]JC School of Public Health and Primary Care, The Chinese University of Hong Kong, Hong Kong, China (Hong Kong)

[2]Stanley Ho Centre for Emerging Infectious Diseases, The Chinese University of Hong Kong, Hong Kong, China (Hong Kong)

[3]Shenzhen Research Institute, The Chinese University of Hong Kong, Hong Kong, China (Hong Kong)

[4]GX Foundation, Hong Kong, China (Hong Kong)

[5]School of Public Health, Imperial College London, London, United Kingdom

[6]Department of Medicine and Therapeutics, The Chinese University of Hong Kong, Hong Kong, China (Hong Kong)

[7]Centre for Health Systems and Policy Research, The Chinese University of Hong Kong, Hong Kong, China (Hong Kong)

Corresponding Author:
Kin On Kwok, PhD
JC School of Public Health and Primary Care
The Chinese University of Hong Kong
Room 419, 4/F, School of Public Health Building
Prince of Wales Hospital, Shatin, NT
Hong Kong
China (Hong Kong)
Phone: 852 22528405
Email: kkokwok@cuhk.edu.hk

Abstract

Background: COVID-19 has plagued the globe, with multiple SARS-CoV-2 clusters hinting at its evolving epidemiology. Since the disease course is governed by important epidemiological parameters, including containment delays (time between symptom onset and mandatory isolation) and serial intervals (time between symptom onsets of infector-infectee pairs), understanding their temporal changes helps to guide interventions.

Objective: This study aims to characterize the epidemiology of the first two epidemic waves of COVID-19 in Hong Kong by doing the following: (1) estimating the containment delays, serial intervals, effective reproductive number (R_t), and proportion of asymptomatic cases; (2) identifying factors associated with the temporal changes of the containment delays and serial intervals; and (3) depicting COVID-19 transmission by age assortativity and types of social settings.

Methods: We retrieved the official case series and the Apple mobility data of Hong Kong from January-August 2020. The empirical containment delays and serial intervals were fitted to theoretical distributions, and factors associated with their temporal changes were quantified in terms of percentage contribution (the percentage change in the predicted outcome from multivariable regression models relative to a predefined comparator). R_t was estimated with the best fitted distribution for serial intervals.

Results: The two epidemic waves were characterized by imported cases and clusters of local cases, respectively. R_t peaked at 2.39 (wave 1) and 3.04 (wave 2). The proportion of asymptomatic cases decreased from 34.9% (0-9 years) to 12.9% (≥80 years). Log-normal distribution best fitted the 1574 containment delays (mean 5.18 [SD 3.04] days) and the 558 serial intervals (17 negative; mean 4.74 [SD 4.24] days). Containment delays decreased with involvement in a cluster (percentage contribution: 10.08%-20.73%) and case detection in the public health care sector (percentage contribution: 27.56%, 95% CI 22.52%-32.33%). Serial intervals decreased over time (6.70 days in wave 1 versus 4.35 days in wave 2) and with tertiary transmission or beyond (percentage contribution: –50.75% to –17.31%), but were lengthened by mobility (percentage contribution: 0.83%). Transmission within the same age band was high (18.1%). Households (69.9%) and social settings (20.3%) were where transmission commonly occurred.

Conclusions: First, the factors associated with reduced containment delays suggested government-enacted interventions were useful for achieving outbreak control and should be further encouraged. Second, the shorter serial intervals associated with the composite mobility index calls for empirical surveys to disentangle the role of different contact dimensions in disease transmission. Third, the presymptomatic transmission and asymptomatic cases underscore the importance of remaining vigilant about COVID-19. Fourth, the time-varying epidemiological parameters suggest the need to incorporate their temporal variations when depicting the epidemic trajectory. Fifth, the high proportion of transmission events occurring within the same age group supports the ban on gatherings outside of households, and underscores the need for residence-centered preventive measures.

(J Med Internet Res 2021;23(4):e26645) doi:10.2196/26645

KEYWORDS

SARS-CoV-2; COVID-19; evolving epidemiology; containment delay; serial interval; Hong Kong; epidemiology; public health; transmission; China; intervention; case study

Introduction

The novel coronavirus (SARS-CoV-2), which causes COVID-19, first appeared in Wuhan, China, in late December 2019 and quickly plagued the globe. The World Health Organization declared COVID-19 a pandemic on March 12, 2020. As of September 27, 2020, there have been 32.7 million cases and almost one million deaths worldwide [1]. Countries are experiencing the resurgence of COVID-19. For example, in the week of September 21-27, 2020, there were about 420,000 new cases in Europe [1], triggering another round of lockdown measures [2]. Researchers promptly summarized the case epidemiology during the early phase of the pandemic [3,4]. However, the enactment of nonpharmaceutical interventions and the presence of multiple genetic SARS-CoV-2 clusters [5] hint at important changes in the epidemiology of COVID-19.

Hong Kong is no exception with regard to COVID-19. The first epoch of the first wave of COVID-19 in Hong Kong took off after the first imported case reported on January 23, 2020 [6,7]. Initially, the epidemic was under control after prompt bundled public health interventions [8]. With the number of infections surging worldwide in mid-March 2020, Hong Kong faced the second epoch of the first wave of infection. Compulsory laboratory tests for all arriving passengers followed by 14-day compulsory quarantine spurred overseas residents to return, resulting in a small influx of imported cases. After peaking in March 2020, the number of cases remained low until a surge of local cases in July 2020, signifying the second wave of the epidemic in Hong Kong. This second wave represented the largest local outbreak in Hong Kong, which was likely attributable to initiation by imported cases coupled with the easing of social-distancing measures in July 2020.

The disease course of COVID-19 is governed by important epidemiological parameters, including containment delay and serial interval. The former has been shown to be associated with the infection source and number of doctor consultations [6], which in turn vary as the epidemic progresses, whereas the latter varies by virus type and subtype [9,10], the contact patterns between susceptible and infectious individuals [10], and the implementation of nonpharmaceutical interventions during epidemics [11].

Parameterization of mathematical models that account for the temporal variation of epidemiological characteristics would improve decisions regarding mitigation strategies. Moreover, containment delay increases opportunities for transmission and affects the effectiveness of control measures, whereas investigating ways to reduce containment delay could enhance control measures [12]. As such, we analyzed laboratory-confirmed COVID-19 case series in Hong Kong between January 2020 and August 2020 to quantify and identify the factors associated with the containment delay and serial interval.

Methods

Data Retrieval

We analyzed the case series provided by the Hong Kong Centre for Health Protection (HKCHP) from January 23, 2020, to August 2, 2020, from which we extracted the following: demographics, case classification, travel history, epidemiological links among cases, date of symptom onset, date of isolation, and the report date. Based on the order of settings embedded in a cluster and the case classification (cases, or close contact of cases), we compiled a line-list database of infector-infectee pairs (hereafter denoted as “paired data”).

Definitions

A laboratory-confirmed case (hereafter denoted as “a case”) and a cluster were defined previously [6]. In short, a case refers to an individual with SARS-CoV-2 detected in a clinical specimen, and a cluster refers to at least two cases that are epidemiologically linked. Further, a local cluster is defined as a cluster that consists of at least one local case. A cluster encompasses one or more orders of settings, referred to as primary, secondary, tertiary, and quaternary settings.

Containment delay and serial interval were defined previously [6]. In short, containment delay is the time elapsed between the first onset of symptoms and mandatory isolation of a case [6], and serial interval is the time interval between the symptom onset of an infector and an infectee [13]. Further, secondary transmission refers to the first generation of infections induced by a case, and infections caused by infectees of a secondary transmission are referred to as tertiary transmission. Accordingly, subsequent orders of transmission are named based on the aforementioned rationale. Effective reproductive number (R_t) is the average number of secondary cases generated by a primary case at any given time. It measures real-time

transmissibility in response to control measures. The epidemic will shrink if R_t is consistently smaller than 1, and vice versa.

Classifications of Cases

The HKCHP classified cases into six types according to their likely source of infection: imported cases, local cases, possibly local cases, and cases with epidemiological linkage with imported, local, or possibly local cases. Based on their travel history during the 14 days preceding the first symptom onset and their involvement in local clusters, the latter four types of cases were reclassified (Table 1) such that there were only three types of cases: (1) imported, (2) local, and (3) unclassified.

Table 1. Reclassification regime of cases.

Original classification	Reclassified classification	Number of cases (N=3512)
Imported cases	Imported cases	1045
Local cases	Local cases	901
Possibly local cases	Local cases[a]	2
Possibly local cases	Imported cases[b]	71
Possibly local cases	Unclassified cases[c]	30
Close contact of imported cases	Local	31
Close contact of local cases	Local	1370
Close contact of possibly local cases	Local	62

[a]Cases without travel history during the 14 days before their first symptom onset.

[b]Cases that had travel history during the 14 days before their first symptom onset and were not involved in any local cluster.

[c]Cases who either had travel history during the 14 days before their first symptom onset and were linked to a local cluster or did not have an onset/arrival date.

Symptom Profile

Symptoms manifested by cases, if any, were grouped based on the International Statistical Classification of Diseases and Related Health Problems (ICD-10) and the national ambulatory medical care survey [14] into 8 categories (Table S1 in Multimedia Appendix 1), including general (such as fever and headache) and respiratory (such as cough and sore throat) symptoms.

Mobility Index

The index is generated by Apple Maps, and represents the relative volume of routing requests (walking) for directions in Hong Kong compared to the baseline volume on January 13, 2020 [15]. The higher the index value is from the baseline, the higher the level of mobility. For ease of presentation, this index was further normalized with the value on January 18, 2020, set as 100.

Statistical Analysis

Characteristics of cases and epidemiological parameters were summarized with mean, standard deviation, percentage, frequency, and bootstrapped confidence interval. Missing isolation dates were replaced by rounding up the mean of available isolation days in a stratum, which was made up of cases that had the same likely source of infection, asymptomatic indicator, quarantine status, mode of detection, and report dates. If there was no available isolation date in a stratum, the mean day differences between report dates and isolation dates were calculated, and the missing isolation dates were imputed as follows: the corresponding report dates minus the mean day difference. For containment delay, only local cases who were neither quarantined nor under medical surveillance and had nonnegative containment delay were analyzed. For serial intervals, only settings that linked 2-4 cases, had identifiable infectors, and were related to local transmission were used to generate infector-infectee pairs for analysis. For R_t, only symptomatic local cases were considered.

A Markov chain Monte Carlo with doubly interval-censored likelihood [16] was adopted to fit the empirical containment delay and serial interval with four candidate distributions (gamma, lognormal, weibull, and normal) with credible intervals computed. The smallest value of the leave-one-out cross-validation information criterion (LOOIC) [17] indicates the best fitted distribution. Assuming the best fitted distribution for serial intervals as the weighted infectivity function, incidence data of local cases by dates of symptom onset with an 8-day window (to reflect the time between exposure to SARS-CoV-2 and symptom onset) were fitted by a novel and statistically robust tool for estimating R_t [18].

Assuming the respective best fitted distributions as the likelihood function, multivariate linear regression models were used to examine the associated factors for containment delays and serial intervals. The effect of each factor was measured in terms of percentage contribution as previously described [19], which is defined as the percentage change in the predicted value of the outcome (from the regression models) relative to a comparator set of covariate values (referred to as “the comparator”). The comparator refers to a transmission event fulfilling these conditions: the case (for containment delay) or the infector (for serial interval) was female, aged 0-30 years, free of chronic diseases, free of general and respiratory symptoms; the transmission happened in a household, during wave 1, and is of the secondary generation; and (for serial interval only) the mobility index being the baseline value on the onset date of the infector. Specifically, $\ln(\alpha_i + j + 1)$ and

$\ln(\beta_i + k + 1)$ were defined as the response variables for the multivariable regression analyses, where α_i and β_i are the containment delay and serial interval of the i^{th} case, respectively, and j and k are the maximum absolute values for the nonpositive containment delays and the nonpositive serial intervals, respectively.

A statistical significance of .05 was specified. All analyses were performed in R (version 3.6.3; R Foundation for Statistical Computing) [20] with the RStan package (version 2.19.3).

Ethical Statement

This study was approved by the Survey and Behavioral Research Ethics Committee of The Chinese University of Hong Kong (reference: SBRE-19-595).

Data Statement

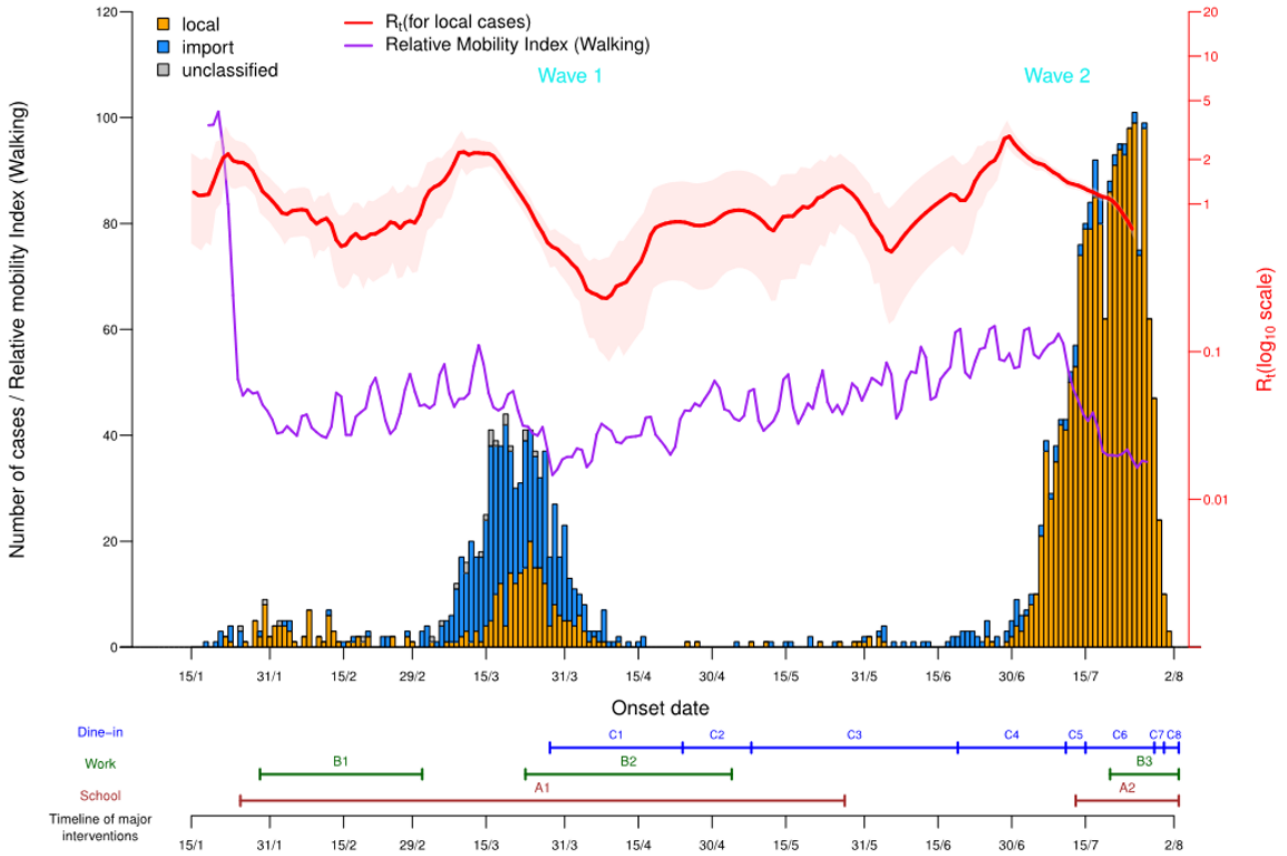
The availability of the data set is subject to approval from HKCHP and relevant government departments.

Results

Characteristics of the Two Epidemic Waves

The report date, June 14, 2020, defined two epidemic waves in Hong Kong: (1) January 23-June 14, 2020 (wave 1) and (2) June 15-August 2, 2020 (wave 2). This cut-off date separated the clusters in both waves without overlap. Two epochs were further defined within wave 1: (1) January 23-February 29, 2020 (epoch 1) and (2) March 1-June 14, 2020 (epoch 2). Wave 1 was initially dominated by imported cases from mainland China and local cases (epoch 1) and subsequently by imported cases from Europe and the Americas (epoch 2), whereas wave 2 was composed mainly of local cases (Figure 1).

Figure 1. Epidemic curve of COVID-19 and timeline for major interventions in Hong Kong. A1: School closure, January 25-May 26, 2020. A2: School closure, July 13, 2020 to ongoing as of August 2, 2020. B1: The government spearheaded work-from-home arrangements, January 29-March 1, 2020. B2: The government spearheaded work-from-home arrangements, March 23-May 3, 2020. B3: The government spearheaded work-from-home arrangements, July 20, 2020 to ongoing as of August 2, 2020. C1: Regulations imposed on dine-in services, March 28-April 23, 2020: (1) tables ≥1.5 meters apart, (2) ≤4 persons per table, and (3) number of customers ≤50% of capacity. C2: Regulations imposed on dine-in services, April 24-May 7, 2020: (1) tables ≥1.5 meters apart and (2) ≤4 persons per table. C3: Regulations imposed on dine-in services, May 8-June 18, 2020: (1) tables ≥1.5 meters apart and (2) ≤8 persons per table. C4: Regulations imposed on dine-in services, June 19-July 10, 2020: tables ≥1.5 meters apart. C5: Regulations imposed on dine-in services, July 11-July 14, 2020: (1) tables ≥1.5 meters apart, (2) ≤8 persons per table, and (3) number of customers ≤60% of capacity. C6: Regulations imposed on dine-in services, July 15-July 28, 2020: (1) tables ≥1.5 meters apart, (2) ≤4 persons per table, (3) number of customers ≤50% of capacity, and (4) no dine-in service from 6 PM to 4:59 AM every day. C7: No dine-in service at any time, July 29-July 30, 2020. C8: Regulations imposed on dine-in services, July 31, 2020 to ongoing as of August 2, 2020: (1) tables ≥1.5 meters apart, (2) ≤2 persons per table, (3) number of customers ≤50% of capacity, and (4) no dine-in service from 6 PM to 4:59 AM every day.



The government of Hong Kong has enacted multipronged interventions, and the major ones include school closures, work-from-home arrangements, and limiting customer flow and time for dine-in services (Figure 1). In addition, the public

voluntarily reduced their mobility substantially, such that the mobility index dropped from 100.2 initially to as low as 28.0 as the epidemic progressed (Figure 1).

R_t was generally below 1 throughout the epidemic, but it peaked at 2.39 in wave 1 and 3.04 in wave 2. This elevated R_t, coupled with the short doubling time of the epidemic size in wave 2 (Figure S1 in Multimedia Appendix 1), suggested that Hong Kong was on the verge of an uncontrolled outbreak during wave 2.

Characteristics of Cases

As of August 2, 2020, 3512 cases were reported. The mean age of cases was 43.91 (SD 20.27) years, with half of them being male (1748/3512). In addition, 6.2% (216/3512) had consulted with a doctor before case confirmation, and 9.7% (339/3512) had chronic conditions (Table S2 in Multimedia Appendix 1). Most cases (2649/3512, 75.4%) were symptomatic, with fever (929/2649, 35.1%) and cough (805/2649, 30.4%) being the most common symptoms (Table S3 in Multimedia Appendix 1). The proportion of asymptomatic cases decreased from the younger age groups to the older ones (0-9 years: 34.9%; 10-19 years: 34.9%; 20-29 years: 21.7%; 30-39 years: 22.4%; 40-49 years: 18.7%; 50-59 years: 16.1%; 60-69 years: 12.7%; 70-79 years: 21.7%; and ≥80 years: 12.9%). In terms of the probable source of infection, 67.4% (2366/3512) were locally acquired cases (Table 1), among whom 1575 cases were symptomatic and neither quarantined nor under medical surveillance (Table 2).

Table 2. Classification of 3512 cases in Hong Kong, as of August 2, 2020.

Case classification	Wave 1 (n=1110), n (%)	Wave 2 (n=2402), n (%)	Total (n=3512), n (%)
Imported case			
Symptomatic	554 (49.9)	94 (3.9)	648 (18.5)
Asymptomatic	189 (17)	231 (9.6)	420 (12)
Missing	0 (0)	48 (2)	48 (1.4)
Local case			
Symptomatic			
Quarantine[a]	31 (2.8)	158 (6.6)	189 (5.4)
Nonquarantine			
Medical surveillance	91 (8.2)	125 (5.2)	216 (6.2)
Others[b]	176 (15.9)	1399 (58.2)	1575 (44.8)
Asymptomatic	39 (3.5)	246 (10.2)	285 (8.1)
Missing	0 (0)	101 (4.2)	101 (2.9)
Unclassified	30 (2.7)	0 (0)	30 (0.9)

[a]This included home/hotel confinee cases, camp/center quarantine cases, and cases released from home/hotel quarantine before getting infected.

[b]This included cases with these modes of case detection: enhanced lab surveillance, enhanced surveillance in private, enhanced surveillance at general outpatient clinics and accident and emergency departments, meeting reporting criteria, diagnosed in private clinic, test at private clinic, and being under Tier 7 classification.

Containment Delay

After excluding one case with negative containment delay, 1574 cases were included in the estimation. Log-normal distribution (mean 5.18 [SD 3.04] days) fitted the empirical containment delay best (LOOIC: 7525.6; Table S4 in Multimedia Appendix 1), which ranged from 4.38 days (95% empirical CI [eCI] 3.80-4.95) for cases belonging to the secondary settings or beyond in a cluster to 6.48 days (95% eCI 6.15-6.82) for cases identified through a private mode of detection (Table 3).

Table 3. Estimates of and factors associated with containment delay based on 1574 cases.

Factors	Cases, n (%)	Subgroup-specific estimates (95% empirical CI)	Percentage contribution (95% CI)
Age group (years)			
0-30	301 (19.1)	4.87 (4.55, 5.18)	Reference[c]
31-45	358 (22.7)	4.86 (4.58, 5.14)	0.04 (–7.82, 8.48)
46-60	448 (28.5)	5.27 (4.99, 5.55)	6.86 (–1.14, 15.41)
≥61	467 (29.7)	5.52 (5.21, 5.82)	11.29 (3.05, 20.09)
Sex			
Female	834 (53)	5.15 (4.95, 5.35)	Reference[c]
Male	740 (47)	5.20 (4.98, 5.42)	2.11 (–3.17, 7.63)
Number of general symptoms			
0	985 (62.6)	5.38 (5.20, 5.56)	Reference[c]
≥1	589 (37.4)	4.83 (4.58, 5.08)	–12.45 (–17.39, –7.27)
Number of respiratory symptoms			
0	1002 (63.7)	5.17 (4.99, 5.35)	Reference[c]
1	459 (29.2)	5.20 (4.91, 5.49)	–1.25 (–7.09, 4.91)
≥2	113 (7.2)	5.08 (4.54, 5.62)	1.86 (–8.77, 13.57)
Order of settings			
None	423 (26.9)	5.71 (5.43, 6.00)	Reference[c]
Primary	1045 (66.4)	5.03 (4.85, 5.22)	–10.08 (–15.44, –4.43)
Secondary or beyond	106 (6.7)	4.38 (3.80, 4.95)	–20.73 (–29.61, –10.93)
Mode of case detection			
Private[a]	347 (22)	6.48 (6.15, 6.82)	Reference[c]
Public[b]	1227 (78)	4.80 (4.64, 4.96)	–27.56 (–32.33, –22.52)
Wave			
1	176 (11.2)	5.29 (4.78, 5.80)	Reference[c]
2	1398 (88.8)	5.16 (5.00, 5.31)	–7.23 (–15.58, 1.82)

[a]Private mode includes diagnosis in private, enhanced surveillance in private, and private test.

[b]Public mode includes enhanced lab surveillance, enhanced surveillance at general outpatient clinics and accident and emergency departments, meeting reporting criteria, Tier 6, and Tier 7.

[c]The reference is a common comparator across all variables.

Containment delay varied by several factors (Table 3; Figure S2A in Multimedia Appendix 1). Notably, the percentage contribution (relative to the comparator) significantly increases with older age (≥61 years) by 11.29% (95% CI 3.05%-20.09%). On the contrary, the percentage contribution decreases with the following: (1) the presence of general symptoms (12.45%, 95% CI 7.27%-17.39%); (2) the transmission event originating from a primary setting of a cluster (10.08%, 95% CI 4.43%-15.44%); (3) the transmission event originating from a secondary setting or beyond of a cluster (20.73%, 95% CI 10.93%-29.61%); and (4) private mode of case detection (27.56%, 95% CI 22.52%-32.33%). The containment delay was shorter in wave 2 than in wave 1, though not statistically significant; the percentage contribution by wave 2 was –7.23% (95% CI –15.58% to 1.82%).

Serial Interval

Initially, there were 847 settings linking at least two cases. After removing 144 settings consisting of solely imported cases, 72 settings with ≥5 cases (as the simultaneous presence of a large number of cases in a setting obscures the transmission link between cases), and 104 settings in which the infectors could not be identified (for example, there was no, or more than one, index case), 528 settings remained. From these 528 settings, 757 infector-infectee pairs were generated. After further removing 4 pairs with duplicated infectees and 195 pairs with missing onset dates of infectors/infectees, 558 paired data were included in the analysis. The mean number of infectees per infector was 1.46.

There were 17 negative serial intervals (range: –5 to –1 days). Log-normal distribution (mean 4.74 [SD 4.24] days) fitted the

overall empirical serial intervals best (LOOIC: 3095.5; Table S4 in Multimedia Appendix 1). The serial intervals were 6.70 days (95% eCI 5.45-7.95) and 4.35 days (95% eCI 4.00-4.70) in waves 1 and 2, respectively. Further, the subgroup estimates of serial intervals ranged from 2.18 days (95% eCI –0.52 to 4.88) for quaternary transmission or beyond to 5.85 days (95% eCI 4.57-7.13) among the infector-infectee pairs with infectors with chronic conditions (Table 4).

Table 4. Estimates of and factors associated with serial interval based on 558 infector-infectee pairs.

Factors	Pairs, n (%)	Subgroup-specific estimates (95% empirical CI)	Percentage contribution (95% CI)
Age group of infector (years)			
0-30	77 (13.8)	4.12 (3.41, 4.83)	Reference[c]
31-45	121 (21.7)	4.43 (3.53, 5.33)	–0.24 (–16.50, 17.88)
46-60	179 (32.1)	4.84 (4.26, 5.43)	10.47 (–6.17, 28.92)
≥61	181 (32.4)	5.12 (4.42, 5.82)	7.57 (–9.28, 26.33)
Sex of infector			
Female	295 (52.9)	4.83 (4.34, 5.33)	Reference[c]
Male	263 (47.1)	4.64 (4.10, 5.18)	–5.52 (–14.81, 4.37)
Presence of chronic conditions among infectors			
No	485 (86.9)	4.58 (4.20, 4.95)	Reference[c]
Yes	73 (13.1)	5.85 (4.57, 7.13)	0.41 (–15.17, 17.67)
Number of general symptoms presented by infectors			
0	338 (60.6)	4.78 (4.32, 5.25)	Reference[c]
≥1	220 (39.4)	4.69 (4.09, 5.28)	–6.34 (–16.48, 4.53)
Number of respiratory symptoms presented by infectors			
0	336 (60.2)	4.22 (3.83, 4.61)	Reference[c]
≥1	222 (39.8)	5.54 (4.84, 6.23)	8.29 (–3.07, 20.48)
Type of setting			
Household	390 (69.9)	4.83 (4.41, 5.25)	Reference[c]
Institution	10 (1.8)	5.00 (2.59, 7.41)	4.05 (–30.27, 47.56)
Social activity	113 (20.3)	4.27 (3.54, 4.99)	–9.17 (–20.78, 3.41)
Work	45 (8.1)	5.16 (3.12, 7.19)	–4.00 (–21.02, 15.12)
Wave			
1	94 (16.8)	6.70 (5.45, 7.95)	Reference[c]
2	464 (83.2)	4.35 (4.00, 4.70)	–33.90 (–45.08, –21.63)
Order of transmission			
Secondary	445 (79.7)	4.89 (4.49, 5.30)	Reference[c]
Tertiary	102 (18.3)	4.36 (3.46, 5.27)	–17.31 (–28.84, –4.75)
Quaternary or beyond	11 (2)	2.18 (–0.52, 4.88)	–50.75 (–72.53, –23.45)
Relative mobility index[a,b,c]	N/A[d]	N/A	0.83 (0.32, 1.33)

[a]An 8-day lag was assumed to account for the time between exposure to SARS-CoV-2 and the first symptom onset, based on the estimated 7.76-day incubation period [21].

[b]The original mobility index was further adjusted relative to January 18, 2020, which has a value of 100.

[c]The reference is a common comparator across all variables.

[d]N/A: not applicable.

Serial intervals varied by several factors (Table 4; Figure S2B in Multimedia Appendix 1). They were significantly shorter in wave 2 than in wave 1: the percentage contribution by wave 2 was –33.9% (95% CI –45.08% to –21.63%). Serial intervals

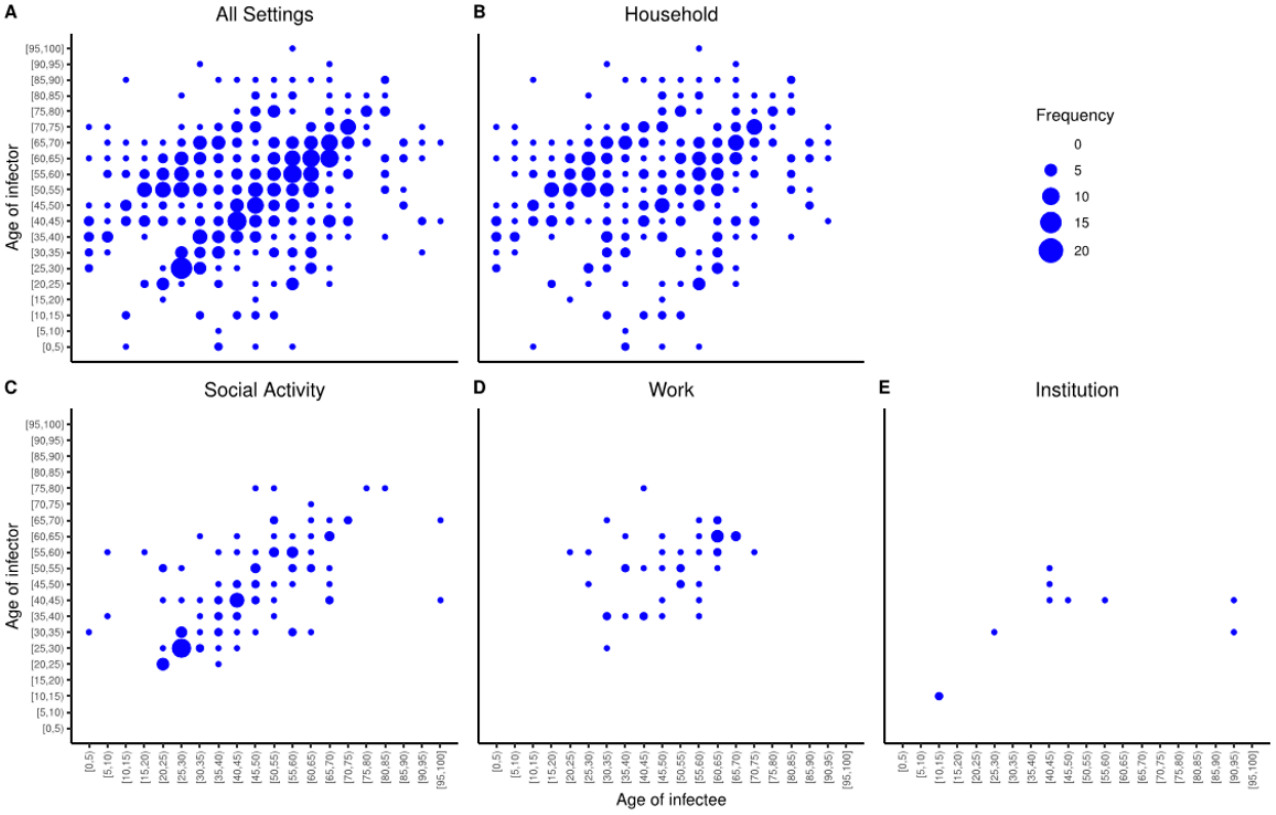
also decreased with a tertiary transmission (percentage contribution: –17.31%, 95% CI –28.84% to –4.75%), and a quaternary transmission or beyond (percentage contribution: –50.75%, 95% CI –72.53% to –23.45%). On the other hand, the 8-day lagged relative mobility index (assuming an incubation period of 7.76 days [21], which reflected the possible exposure date for each infector-infectee pair) lengthened the serial interval (percentage contribution: 0.83%, 95% CI 0.32%-1.33%).

Transmission Events

Transmission events that occurred within the same age band were high (101/558, 18.1%) when using 5-year age bands (Figure 2A). Similar age transmission patterns were also observed in households (Figure 2B) and in social settings (Figure 2C), but were less obvious in work (Figure 2D) and institutional (Figure 2E) settings. In the household setting, transmission from infectors in older age groups (50-70 years) to infectees in younger age groups (15-40 years) was commonly observed (Figure 2B). The age transmission matrix asymmetry in all settings well reflected this phenomenon as well as the lack of transmission from younger age groups to older ones (Figure 2A).

Importantly, households (390/558, 69.9%) and social settings (113/558, 20.3%) were the two most common settings for transmission in Hong Kong (Table 4) in both epidemic waves (Table S5 in Multimedia Appendix 1).

Figure 2. Age-specific transmission events in (A) all settings, (B) households, (C) social settings, (D) work settings, and (E) institutions.



Discussion

Summary of Study Findings

Understanding the evolving epidemiology of an infectious disease is vital for guiding infection control policies. From the case series between January 23, 2020, and August 2, 2020, we identified two waves of epidemics. R_t was in general below the outbreak threshold, suggesting that interventions, whether adopted voluntarily by the community [22] or institutionalized by the government [8], successfully interrupted the transmission of COVID-19. As the epidemic grew in Hong Kong, containment delays and serial intervals were shortened over time. The shortening of the former was associated with the manifestation of general symptoms, involvement in clusters, and public mode of case detection, whereas that of the latter was associated with later transmission generations and lower mobility. The occurrence of intra–age group transmission is more common than inter–age group transmission, and households and social settings together account for 90.14% of all identified transmission events.

Result Implications

Our results have five implications. First, the factors associated with a reduction in containment delay suggested that government-enacted interventions were useful in achieving COVID-19 outbreak control in Hong Kong and should be further encouraged. Containment delay played a major role in determining whether an outbreak was controllable; assuming 80% of contacts can be traced, the chance of controlling an outbreak fell from 89% to 31% if the containment delay increased from 3.43 days to 8.09 days [23]. The difference in containment delay experienced by cases detected by different modes (private: 6.48 days; public: 4.80 days) pinpointed the worthiness of continual investment in public modes of case detection, including setting up community testing centers and

mobile specimen collection stations. Further, as reflected by the percentage contribution by cases involved in a cluster (–20.73% to –10.08%), contact tracing and the follow-up quarantine of individuals with epidemiological links with cases were useful in reducing containment delay; contact tracing should involve a high tracing ratio (preferably ≥80%) and should be done when there are only a few initial cases [23].

Second, the association between decreasing serial intervals over time and lower mobility aligned with the contention that serial intervals are shortened by nonpharmaceutical interventions [11]. Reduced serial intervals indicated faster case generation replacement, which may be attributable to the institutionalized social-distancing policies (Figure 1) that diminished the geographical reach of citizens. This is a tradeoff between time spent in a place versus number of places visited in a limited time, such that citizens may stay in confined locations (eg, home) longer. Although the underlying mechanisms remain undetermined, we hypothesize that confined geographical movement would, in reality, intensify the proximity of contacts between successive case generations. This hypothesis is in line with earlier findings that more time spent in close proximity to the index case shortened serial intervals of influenza [10]. Although the adopted mobility index remained composite, empirical contact surveys—such as the ones by Mossong and colleagues [24] and Kwok and colleagues [19]—that disentangle the interplay of contact dimensions will advance research in the area of evolving epidemiology. On one hand, it is of interest to dissect the role of different contact dimensions on disease transmission; on the other, the potential of other composite social mobility measures—such as the Twitter Social Mobility Index [25]—in estimating epidemiological parameters should be explored.

Third, the presence of negative serial intervals, suggestive of presymptomatic transmission and asymptomatic cases, is a reminder for the community to stay on guard against the resurgence of COVID-19. Infections happening before symptom onset would impede the effectiveness of control measures. This presymptomatic fraction appeared to be low in this study (17/558; 3.05%), but it can be as high as 12.6% [26]. Meanwhile, the proportion of asymptomatic local cases was 12.0% (285/2366) and the number of accumulated infections so far (ie, 3512 cases among more than 7 million people in Hong Kong) is not high enough to confer herd immunity when compared with the conservative 5.66% previously suggested [27]. Together, coupled with the fact that vaccinations will take some time, these results suggest that the community should remain vigilant against any resurgence.

Fourth, temporal variations of key epidemiological parameters should be considered. It is common to assume that the empirical data (of the epidemiological parameters) resembles theoretical distributions. However, the correlation of lower mobility (as proxy to voluntary or compulsory social distancing en masse), involvement in local clusters (as proxy to early government actions on case tracing), and government-level case identification with either containment delay or serial interval observed in this study suggests that epidemiological parameters are dynamic throughout the epidemic. Furthermore, with SARS-CoV-2 transcending international borders, it has mutated into different clusters or subtypes [28-30]. These subtypes differ by their intrinsic properties, exhibiting variations in COVID-19 epidemiology [30]. Therefore, caution must be taken to interpret findings from infectious disease models that assume static epidemiological parameters.

Fifth, the high proportion of transmission events within the same age group supports the ban on gatherings outside of households. The observed intra–age group transmission of COVID-19 echoed the contact assortativity by age of the Hong Kong population [19], and is in line with earlier research [31]: social contact should be considered together with age when it comes to determining the driving force of the incidence of respiratory infections. Further, the asymmetric age transmission matrix revealed that children rarely infected others in the first two epidemic waves. This phenomenon may be attributable to the continual school closure, resulting in fewer social interactions than usual among children in Hong Kong. With the reopening of schools, and hence more social mixing among children, transmission chains branching from children are possible; therefore, older adults who are frequently in contact with children should be prioritized to receive the COVID-19 vaccine, as older adults have a higher risk of COVID-19–related mortality [32]. In addition, the abundance of transmission events in households underscores the need for residence-centered preventive measures, as per the lesson learned from aerosol transmission (through defective plumbing) of the 2003 severe acute respiratory syndrome virus in the Amoy Gardens housing complex in Hong Kong [33].

Study Limitations

There are two study limitations that bear mentioning. First, the data in this study, including self-reported symptoms and contact history, were subject to recall bias. However, the medical surveillance in place to monitor the contacts of cases may lessen the data uncertainties. Second, some cases might not have been captured by the present surveillance system in Hong Kong due to the underdiagnosis of mild cases and asymptomatic individuals, who might play a role in the transmission chain involving unlinked local cases.

Acknowledgments

This study is funded by Direct Grant for Research (number: 2019.020) from The Chinese University of Hong Kong, and by the Centre for Health Systems and Policy Research which is supported by the Tung Foundation. The authors would like to thank Hong Kong Centre for Health Protection and Food and Health Bureau for providing the data, and Margaret Tsoi for technical support. KOK also acknowledges support from Health and Medical Research Fund (reference numbers: INF-CUHK-1, 17160302,

18170312), General Research Fund (reference numbers: 14112818, 24104920), and Wellcome Trust Fund (United Kingdom, 200861/Z/16/Z).

Authors' Contributions

KOK, SYSW, and EKY conceptualized the study and collected the data; KOK, WIW, YH, and HHHC analyzed the data; KOK and WIW wrote the first draft of the manuscript; and all authors interpreted the data, edited the manuscript, and provided critical comments. KOK and WIW contributed equally as the joint corresponding authors. SYSM and EKY contributed equally as the joint last authors.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary data.
[DOCX File , 195 KB - jmir_v23i4e26645_app1.docx]

References

1. World Health Organization. Weekly Epidemiological Update - Coronavirus disease 2019 (COVID-19). 2020 Sep 27. URL: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200928-weekly-epi-update.pdf?sfvrsn=9e354665_6 [accessed 2020-09-30]
2. Covid: How are European countries tackling the pandemic? BBC News. URL: https://www.bbc.com/news/explainers-53640249
3. Park M, Cook AR, Lim JT, Sun Y, Dickens BL. A Systematic Review of COVID-19 Epidemiology Based on Current Evidence. J Clin Med 2020 Mar 31;9(4):967 [FREE Full text] [doi: 10.3390/jcm9040967] [Medline: 32244365]
4. Khalili M, Karamouzian M, Nasiri N, Javadi S, Mirzazadeh A, Sharifi H. Epidemiological characteristics of COVID-19: a systematic review and meta-analysis. Epidemiol Infect 2020 Jun 29;148:e130. [doi: 10.1017/s0950268820001430]
5. Yang X, Dong N, Chan EW, Chen S. Genetic cluster analysis of SARS-CoV-2 and the identification of those responsible for the major outbreaks in various countries. Emerg Microbes Infect 2020 Dec 11;9(1):1287-1299 [FREE Full text] [doi: 10.1080/22221751.2020.1773745] [Medline: 32525765]
6. Kwok K, Wong V, Wei W, Wong S, Tang JWT. Epidemiological characteristics of the first 53 laboratory-confirmed cases of COVID-19 epidemic in Hong Kong, 13 February 2020. Euro Surveill Internet 2020;25(16):1. [doi: 10.2807/1560-7917.es.2020.25.16.2000155]
7. Kwok KO, Wong V, Wei VWI, Wong SYS, Tang JW. Novel coronavirus (2019-nCoV) cases in Hong Kong and implications for further spread. J Infect 2020 Jun;80(6):671-693 [FREE Full text] [doi: 10.1016/j.jinf.2020.02.002] [Medline: 32092391]
8. Wong SY, Kwok KO, Chan FK. What can countries learn from Hong Kong's response to the COVID-19 pandemic? CMAJ 2020 May 11;192(19):E511-E515 [FREE Full text] [doi: 10.1503/cmaj.200563] [Medline: 32332040]
9. Vink M, Bootsma M, Wallinga J. Serial intervals of respiratory infectious diseases: a systematic review and analysis. Am J Epidemiol 2014 Nov 01;180(9):865-875. [doi: 10.1093/aje/kwu209] [Medline: 25294601]
10. Levy JW, Cowling BJ, Simmerman JM, Olsen SJ, Fang VJ, Suntarattiwong P, et al. The serial intervals of seasonal and pandemic influenza viruses in households in Bangkok, Thailand. Am J Epidemiol 2013 Jun 15;177(12):1443-1451 [FREE Full text] [doi: 10.1093/aje/kws402] [Medline: 23629874]
11. Ali ST, Wang L, Lau EHY, Xu X, Du Z, Wu Y, et al. Serial interval of SARS-CoV-2 was shortened over time by nonpharmaceutical interventions. Science 2020 Aug 28;369(6507):1106-1109 [FREE Full text] [doi: 10.1126/science.abc9004] [Medline: 32694200]
12. Marinović AB, Swaan C, van Steenbergen J, Kretzschmar M. Quantifying Reporting Timeliness to Improve Outbreak Control. Emerg Infect Dis 2015 Feb;21(2):209-216. [doi: 10.3201/eid2102.130504]
13. Fine PEM. The interval between successive cases of an infectious disease. Am J Epidemiol 2003 Dec 01;158(11):1039-1047. [doi: 10.1093/aje/kwg251] [Medline: 14630599]
14. Meads S, McLemore T. The national ambulatory medical care survey: symptom classification. Vital Health Stat 2 1974 May(63):1-41 [FREE Full text] [doi: 10.1037/e444962004-001] [Medline: 25093394]
15. Mobility trends reports. Apple. URL: https://www.apple.com/covid19/mobility [accessed 2020-08-20]
16. Nishiura H, Linton NM, Akhmetzhanov AR. Serial interval of novel coronavirus (COVID-19) infections. Int J Infect Dis 2020 Apr;93:284-286 [FREE Full text] [doi: 10.1016/j.ijid.2020.02.060] [Medline: 32145466]
17. Vehtari A, Gelman A, Gabry J. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. Stat Comput 2016 Aug 30;27(5):1413-1432. [doi: 10.1007/s11222-016-9696-4]
18. Cori A, Ferguson N, Fraser C, Cauchemez S. A new framework and software to estimate time-varying reproduction numbers during epidemics. Am J Epidemiol 2013 Nov 01;178(9):1505-1512 [FREE Full text] [doi: 10.1093/aje/kwt133] [Medline: 24043437]

19. Kwok KO, Cowling B, Wei V, Riley S, Read JM. Temporal variation of human encounters and the number of locations in which they occur: a longitudinal study of Hong Kong residents. J R Soc Interface 2018 Jan 24;15(138):20170838 [FREE Full text] [doi: 10.1098/rsif.2017.0838] [Medline: 29367241]
20. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing URL: https://www.r-project.org/ [accessed 2021-03-25]
21. Qin J, You C, Lin Q, Hu T, Yu S, Zhou X. Estimation of incubation period distribution of COVID-19 using disease onset forward time: A novel cross-sectional and forward follow-up study. Sci Adv 2020 Aug 07;6(33):eabc1202 [FREE Full text] [doi: 10.1126/sciadv.abc1202] [Medline: 32851189]
22. Kwok KO, Li KK, Chan HHH, Yi YY, Tang A, Wei WI, et al. Community Responses during Early Phase of COVID-19 Epidemic, Hong Kong. Emerg Infect Dis 2020 Jul;26(7):1575-1579 [FREE Full text] [doi: 10.3201/eid2607.200500] [Medline: 32298227]
23. Hellewell J, Abbott S, Gimma A, Bosse NI, Jarvis CI, Russell TW, et al. Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts. The Lancet Global Health 2020 Apr;8(4):e488-e496. [doi: 10.1016/s2214-109x(20)30074-7]
24. Mossong J, Hens N, Jit M, Beutels P, Auranen K, Mikolajczyk R, et al. Social contacts and mixing patterns relevant to the spread of infectious diseases. PLoS Med 2008 Mar 25;5(3):e74 [FREE Full text] [doi: 10.1371/journal.pmed.0050074] [Medline: 18366252]
25. Xu P, Dredze M, Broniatowski DA. The Twitter Social Mobility Index: Measuring Social Distancing Practices With Geolocated Tweets. J Med Internet Res 2020 Dec 03;22(12):e21499 [FREE Full text] [doi: 10.2196/21499] [Medline: 33048823]
26. Du Z, Xu X, Wu Y, Wang L, Cowling BJ, Meyers LA. Serial Interval of COVID-19 among Publicly Reported Confirmed Cases. Emerg Infect Dis 2020 Jun;26(6):1341-1343 [FREE Full text] [doi: 10.3201/eid2606.200357] [Medline: 32191173]
27. Kwok KO, Lai F, Wei WI, Wong SYS, Tang JW. Herd immunity - estimating the level required to halt the COVID-19 epidemics in affected countries. J Infect 2020 Jun;80(6):e32-e33 [FREE Full text] [doi: 10.1016/j.jinf.2020.03.027] [Medline: 32209383]
28. Forster P, Forster L, Renfrew C, Forster M. Phylogenetic network analysis of SARS-CoV-2 genomes. Proc Natl Acad Sci U S A 2020 Apr 28;117(17):9241-9243 [FREE Full text] [doi: 10.1073/pnas.2004999117] [Medline: 32269081]
29. To K, Chan W, Ip J, Chu A, Tam A, Liu R, et al. Unique SARS-CoV-2 clusters causing a large COVID-19 outbreak in Hong Kong. Clin Infect Dis 2020 Aug 05:ciaa1119 [FREE Full text] [doi: 10.1093/cid/ciaa1119] [Medline: 32756996]
30. Tang X, Wu C, Li X, Song Y, Yao X, Wu X, et al. On the origin and continuing evolution of SARS-CoV-2. Natl Sci Rev 2020;7(6):1012-1023. [doi: 10.1093/nsr/nwaa036]
31. Kwok KO, Cowling BJ, Wei VWI, Wu KM, Read JM, Lessler J, et al. Social contacts and the locations in which they occur as risk factors for influenza infection. Proc Biol Sci 2014 Aug 22;281(1789):20140709 [FREE Full text] [doi: 10.1098/rspb.2014.0709] [Medline: 25009062]
32. Ho FK, Petermann-Rocha F, Gray SR, Jani BD, Katikireddi SV, Niedzwiedz CL, et al. Is older age associated with COVID-19 mortality in the absence of other risk factors? General population cohort study of 470,034 participants. PLoS One 2020 Nov 5;15(11):e0241824 [FREE Full text] [doi: 10.1371/journal.pone.0241824] [Medline: 33152008]
33. Yu IT, Li Y, Wong TW, Tam W, Chan AT, Lee JH, et al. Evidence of Airborne Transmission of the Severe Acute Respiratory Syndrome Virus. N Engl J Med 2004 Apr 22;350(17):1731-1739. [doi: 10.1056/nejmoa032867]

Abbreviations

eCI: empirical confidence interval
HKCHP: Hong Kong Centre for Health Protection
LOOIC: leave-one-out cross-validation information criterion

Edited by C Basch; submitted 19.12.20; peer-reviewed by S Lawpoolsri; comments to author 16.01.21; revised version received 22.02.21; accepted 19.03.21; published 16.04.21.

Please cite as:
Kwok KO, Wei WI, Huang Y, Kam KM, Chan EYY, Riley S, Chan HHH, Hui DSC, Wong SYS, Yeoh EK
Evolving Epidemiological Characteristics of COVID-19 in Hong Kong From January to August 2020: Retrospective Study
J Med Internet Res 2021;23(4):e26645
URL: https://www.jmir.org/2021/4/e26645
doi:10.2196/26645
PMID:33750740

Original Paper

Association of Perceived Threat, Negative Emotions, and Self-Efficacy With Mental Health and Personal Protective Behavior Among Chinese Pregnant Women During the COVID-19 Pandemic: Cross-sectional Survey Study

Phoenix Kit Han Mo[1], PhD; Vivian Wai In Fong[1], MSc; Bo Song[2], MMed; Jiangli Di[2], PhD; Qian Wang[2], PhD; Linhong Wang[3], MMed

[1]Center for Health Behaviours Research, School of Public Health and Primary Care, Chinese University of Hong Kong, Hong Kong, Hong Kong

[2]National Center for Women and Children's Health, Chinese Center for Disease Control and Prevention, Beijing, China

[3]National Center for Chronic and Noncommunicable Disease Control and Prevention, Chinese Center for Disease Control and Prevention, Beijing, China

Corresponding Author:
Qian Wang, PhD
National Center for Women and Children's Health
Chinese Center for Disease Control and Prevention
No 12 Dahuisi Road
Haidian District
Beijing, 100081
China
Phone: 86 010 62170970
Email: qianawang@chinawch.org.cn

Abstract

Background: COVID-19 is an emerging infectious disease that has created health care challenges worldwide. Pregnant women are particularly affected by this disease.

Objective: The aims of this study are to assess the levels of perceived threat (susceptibility, severity, impact), negative emotions (fear, worry), and self-efficacy of pregnant women in China related to COVID-19 and to examine their associations with mental health (depression and anxiety) and personal protective behavior (wearing a face mask).

Methods: A total of 4087 pregnant women from China completed a cross-sectional web-based survey between March 3 and 10, 2020.

Results: The prevalence of probable depression and anxiety was 48.7% (1989/4087) and 33.0% (1347/4087), respectively; 23.8% participants (974/4087) reported always wearing a face mask when going out. Of the 4087 participants, 32.1% (1313) and 36.4% (1490) perceived themselves or their family members to be susceptible to COVID-19 infection, respectively; 3216-3518 (78.7%-86.1%) agreed the disease would have various severe consequences. Additionally, 2275 of the 4087 participants (55.7%) showed self-efficacy in protecting themselves from contracting COVID-19, and 2232 (54.6%) showed efficacy in protecting their family members; 1303 (31.9%) reported a high level of fear of the disease, and 2780-3056 (68.0%-74.8%) expressed worry about various aspects of COVID-19. The results of the multivariate multinominal logistic regression analyses showed that perceived severity, perceived impact, fear, and worry were risk factors for probable depression and anxiety, while self-efficacy was a protective factor. The results of the multivariate logistic regression analysis showed that perceived susceptibility was associated with always wearing a face mask.

Conclusions: Chinese pregnant women showed high levels of mental distress but low levels of personal protective behavior during the COVID-19 pandemic. Interventions are needed to promote the mental health and health behavior of pregnant women during the pandemic.

(J Med Internet Res 2021;23(4):e24053) doi:10.2196/24053

KEYWORDS

COVID-19; pregnant women; depression; anxiety; self-efficacy; mental health; survey; threat; emotion

Introduction

COVID-19 as an Emerging Infectious Disease

In December 2019, a cluster of viral pneumonia cases of unknown cause, later named COVID-19, were detected in Wuhan, Hubei Province. After that, more cases of new infections and deaths were reported across cities in China. Due to the highly contagious nature of the disease, the COVID-19 outbreak spread worldwide in less than three months. As of February 1, 2021, a total of 102,339,513 confirmed cases of COVID-19 and 2,217,005 deaths were reported worldwide [1]. The outbreak of COVID-19 represents a public health emergency of international concern.

Measures to Prevent the Spread of COVID-19 in China

Following the initial outbreak of COVID-19 in Wuhan, numerous measures were enacted to prevent further spread of the disease to other parts of China. On January 23, 2020, the Chinese government shut down Wuhan and two additional cities (Wenzhou and Shenzhen) by suspending all modes of transportation to and from these cities. Other measures included cancelling or postponing large public events (ie, Chinese New Year celebrations), prohibition of attendance at school and work, and closure of public amenities. These measures were part of the social distancing policies, which aimed to limit social contacts and human-to-human transmission [2]. Moreover, public information and education campaigns that encouraged personal protective behaviors were promoted through social media marketing and traditional media. A number of protective measures were also implemented to prevent the spread of COVID-19 among members of the workforce who wished to return to work [3]. These measures included enhancing the practice of hand hygiene and wearing a face mask, as well as organizational measures such as improvement of workplace hygiene; tracking the physical health status of employees; and dissemination of facts about COVID-19 prevention [3]. On February 2, 2020, the National Health Commission of China issued a new notice proposing a list of guidelines and recommendations for pregnant women and health care professionals in response to COVID-19. This notice urged strengthening of health counselling, screening, and follow-ups for pregnant women. Recommendations on personal protective measures, including social distancing, hand hygiene, and the use of a face mask when going out, were also stated in the notice [4].

Possible Impact of COVID-19 on Pregnant Women

Due to the immunological changes that occur during pregnancy, pregnant women are generally susceptible to respiratory pathogens and the development of severe pneumonia. Therefore, it is believed that pregnant women may be more susceptible to COVID-19 [5]. Recent reviews found that the most commonly reported symptoms of COVID-19 in pregnant women were fever and dry cough, followed by fatigue, diarrhea, dyspnea, lymphocytopenia with elevated C-reactive protein, sore throat, and myalgia [6-8]. These clinical characteristics of COVID-19 are similar to those of nonpregnant adult patients [6-8]. More complications were observed in symptomatic pregnant women, including intensive care unit admissions, mechanical ventilation, and death [9,10]. Moreover, studies have shown that infected pregnant women are at significantly higher risk for caesarean delivery and preterm birth than the general pregnant population [6,7,11], which is associated with increased risk and complications in both mothers and neonates. A number of fetal and neonatal complications, such as stillbirth, neonatal death, low birth weight, fetal distress, thrombocytopenia accompanied by abnormal liver function, neonatal asphyxia, and death, were also reported [7,12]. Although some reviews reported a low risk of vertical transmission [8,13], it is important to note that most studies to date have found no conclusive evidence of vertical transmission of COVID-19 [6,7,10,11,14]. According to the World Health Organization, mothers with suspected or confirmed COVID-19 are recommended to initiate and continue breastfeeding and mother-infant contact because the benefits of breastfeeding and mother-infant contact substantially outweigh the potential risks of transmission of COVID-19 to the child [15]. As the situation of COVID-19 is constantly changing worldwide, understanding the effect of COVID-19 on pregnancy would be important and beneficial for the development of treatment and management to reduce the morbidity and complications of COVID-19 in both mothers and neonates. Very few studies to date have examined the perceptions, emotions, and mental and behavioral responses to COVID-19 of pregnant women in China.

Mental Health and Personal Preventive Behavior During the COVID-19 Period

The uncertainty of COVID-19 has caused substantial psychological distress to the public. In studies among Chinese people, it was found that during the COVID-19 epidemic, 53.8% of respondents rated the psychological impact of the COVID-19 outbreak as moderate or severe [16], between 16.1% and 20.1% Chinese people scored higher than the cutoff for depression [12,17], and between 20.4% and 35.1% scored higher than the cutoff for anxiety [12,17,18]. Recent studies have revealed that females experience much greater levels of anxiety, depression, and stress than their male counterparts during the COVID-19 pandemic [19-21]. Studies have also shown that pregnant women experienced greater psychological distress than the general population during the COVID-19 epidemic [22,23], as they may face difficulties in accessing health care services due to suspension of transportation and other nonemergency services. Health services for pregnant women may also become limited as the health system is strained by the increasing number of infections. Furthermore, visiting clinics and hospitals for medical checkups may place pregnant women at increased risk of infection. Given the potential risk of COVID-19 infection in pregnancy, pregnant women may feel anxious visiting these facilities and worry about being infected during antenatal checkup. Furthermore, other studies that examined the effect of lockdown and quarantine on the mental health of pregnant population demonstrated a higher prevalence of symptoms of

psychological distress (ie, anxiety, depression, posttraumatic stress symptoms, and insomnia [24,25]). As antenatal stress and mental health problems are associated with multiple adverse outcomes for the child [26,27], it would be pertinent to examine the prevalence and associated factors of mental distress among pregnant women during the COVID-19 period so that early interventions could be designed.

The virus of COVID-19 can be transmitted through respiratory droplets and contacts. From the experience of past epidemic outbreaks, personal protective behaviors such as wearing a mask are recommended to offer protections against virus infection. A recent study reported that the frequency of always wearing a face mask in the past week was 76.1%-85.5% among participants in Wuhan and Shanghai during the COVID-19 period [18]. Research has shown that countries with higher proportions of citizens that used face masks had fewer COVID-19 cases and controlled the epidemic much earlier than countries that discouraged the use of face masks [28]. The frequent use of face masks was found to be associated with a lower number of COVID-19 cases; the perceived benefits of wearing a face mask (eg, effectiveness in preventing virus spread from asymptomatic patients) also boosted confidence and reduced the risk of adverse mental health among Chinese people [28]. Studies among pregnant women during the severe acute respiratory syndrome (SARS) outbreak in Hong Kong found that approximately 70% wore a mask all or most of the time [29]. It was expected that the frequency of wearing a face mask among pregnant women would be high during the COVID-19 pandemic.

Factors Related to Mental Health and Personal Protective Behavior

Cognitions are important determinants of health behaviors and mental health. Protection motivation theory is a prominent theory that highlights the important role of cognition in health [30]. Protection motivation theory posits that two parallel cognitive processes function to predict an individual's motivation to perform a health behavior: threat appraisal and coping appraisal. Threat appraisal focuses on perceived susceptibility (ie, estimation of the possibility of harm from a threat) and perceived severity (ie, estimation of the degree of harm resulting from the threat), while coping appraisal focuses on self-efficacy (ie, level of confidence in coping with the threats) and other factors that may increase or decrease the adaptive response. The perception that the pandemic has caused a significant impact on one's life may decrease the probability of the adaptive response. According to protection motivation theory, increased levels of threat and coping appraisal increase individuals' protection motivation, leading to the performance of a health behavior. Protection motivation theory has been applied to COVID-19 preventive behavior among health care workers and the general public [31,32]. A meta-analysis also confirmed that increases in perceived susceptibility, perceived severity, and self‐efficacy facilitated adaptive intentions or behaviors [33].

Consistent with protection motivation theory, extensive studies have also demonstrated the important role of perceived threats on behavioral and psychological responses to a disease. A review by Bish and Michie [34] identified perceived susceptibility to the disease and perceived severity of the disease as important predictors of protective behaviors during a pandemic. A recent study revealed that individuals' perceived severity of the COVID-19 outbreak is related to an increased level of mental health problems in the Chinese public [35]. Studies from other countries, such as the Philippines, Vietnam, and Turkey, also supported the findings that the perceived impact, perceived susceptibility, and perceived severity of COVID-19 are related to increased levels of mental health problems and preventive behaviors [20,21,36].

The literature has also suggested self-efficacy as an important determinant of mental health and behavior. Studies on epidemic outbreaks, such as SARS and H1N1, have found that self-efficacy was associated with better mental health [37] and higher levels of protective behaviors [38,39], while a lower level of self-efficacy was associated with higher fear related to the disease [40]. Studies during the COVID-19 epidemic also demonstrated that stronger self-confidence, stronger confidence in one's health care providers, and having a good perception of one's health status were significantly associated with lower risks of anxiety, depression, and stress as well as with a lower psychological impact of the pandemic [18,20].

The literature has also highlighted the importance of emotional factors in mental health and behavior. According to the appraisal tendency theory, emotions guide specific cognitive response or appraisal, leading to an effect on mental and behavioral outcomes [41]. Negative emotions, such as fear and worry, are associated with a tendency to perceive a situation as uncertain and less controllable, which prompts individuals to engage in precautionary behaviors. Furthermore, negative emotions may impair individuals' cognitive processes in coping with stressful encounters [42], which drives them to perform a range of precautionary behaviors regardless of their scientific value. Supporting this notion, studies conducted during pandemics, such as H1N1 [43-45] and COVID-19 [36], have shown that negative emotions, such as fear of the pandemic, are associated with the practice of protective behaviors. Negative emotions related to epidemic outbreaks have also been consistently found to be detrimental to mental health [40,46]. It is therefore conjectured that during the COVID-19 pandemic period, pregnant women who showed negative emotions toward COVID-19 may have poorer mental health but be more likely to adopt protective behaviors.

This Study

This study examined the prevalence and identified factors for depression, anxiety, and frequency of face mask wearing during the COVID-19 period among pregnant women in China. Based on the protection motivation theory and the appraisal tendency theory, the roles of perceived susceptibility, perceived severity, perceived impact of COVID-19, self-efficacy, and negative emotions (ie, fear and worry about COVID-19) were examined. It was hypothesized that perceived threat and negative emotions would be risk factors for depression and anxiety, while self-efficacy would be a protective factor. Furthermore, perceived threat, negative emotions, and self-efficacy were

hypothesized to be protective factors of the frequency of wearing a face mask.

Methods

Participants

The target participants were pregnant women who were currently using health care services from maternal health care institutions in Mainland China. Pregnant women who intended to continue the pregnancy were eligible for the study, and those who planned to terminate their pregnancy were excluded.

Procedure

A web-based cross-sectional survey was conducted from March 3-10, 2020. Participants were recruited from maternal health care centers in various provinces of China (ie, Beijing, Chongqing, Guangdong, Guangxi, Hainan, Shandong, Tianjin, and Xinjiang). These maternal health care centers provided antenatal health services to all pregnant women within the region, they contained a record and contact details of those women who were using their health services. Eligible women were first identified from the record and were invited to take part in the survey through WeChat. A quick response (QR) code and link to the web-based survey was provided; interested participants could directly access the web-based survey through scanning the QR code or clicking the link. Information about the purpose and procedure of the survey was provided on the first page of the web-based survey. Participants were assured that no direct identifiers (eg, name, email address, ID number) were collected, all data collected would remain confidential, and only the research team would have access to the data. Refusal to take part in the survey would not affect the services they would obtain at the health care center. Women who agreed to take part in the survey were asked to provide informed consent by clicking the "I agree" button before starting the web-based survey. The survey took 15-20 minutes to complete. Participants received no incentive for their participation. Ethical approval was obtained from the Survey and Behavioral Research Ethics Committee of The Chinese University of Hong Kong (SBRE-19-395). Approximately 5540 invitation messages were sent out, and a total of 4087 completed responses were collected (response rate: 73.8%).

Measures

Sociodemographic Characteristics

Participants were asked to report their age, education level, and employment status.

Pregnancy-Related Characteristics

Participants were asked to report their parity, gestational age, and whether they had any pregnancy-related complications.

Perceived Susceptibility to COVID-19

The participants' perceived susceptibility to COVID-19 was measured by 2 items on the likelihood of oneself and one's family members contracting COVID-19. Items were rated on a 4-point Likert scale from 1, very little, to 4, very much, with a higher score indicating a higher level of perceived susceptibility. The internal reliability of the items was satisfactory (Cronbach α=.94).

Perceived Severity of COVID-19

The participants' perception of the severity of COVID-19 was measured by 3 items (eg, "Would maternal infection with COVID-19 affect the health of the newborn?"). These items were rated on a 4-point Likert Scale from 1, very little, to 4, very much, with a higher score indicating a higher level of perceived severity. The internal reliability of the items was satisfactory (Cronbach α =.92).

Perceived Impact of COVID-19

Participants were given a checklist and were asked to rate whether COVID-19 had affected any part of their daily lives (ie, work, financial income, family relationship, social interactions, others). The number of items endorsed reflected the level of impact. The possible score ranged from 0 to 5, with a higher score indicating a higher level of perceived impact of COVID-19.

Self-efficacy

Self-efficacy was measured by 2 items. Participants were asked to rate their level of confidence of protecting themselves and their family members from contracting COVID-19. Items were rated on a 4-point Likert scale from 1, very little, to 4, very much, with a higher score indicating a higher level of self-efficacy. The internal reliability of the items was satisfactory (Cronbach α=.94).

Fear

Fear was measured by a single item. Participants were asked to rate their level of fear of COVID-19 on a 4-point Likert scale from 1, very little, to 4, very much, with a higher score indicating a higher level of fear.

Worry

Worry was measured by 4 items. Participants were asked to rate their level of worry regarding different aspects related to COVID-19 (eg, "You will be infected with COVID-19 when you attend the prenatal check-up"). Items were rated on a 4-point Likert scale from 1, very little, to 4, very much, with a higher score indicating a higher level of worry. The internal reliability of the items was satisfactory (Cronbach α=.92).

Frequency of Face Mask Wearing

Participants were asked to report their frequency of wearing a face mask when going out on a 4-point Likert scale from 1, never, to 4, always. A cutoff of always wearing a face mask was set in this study; this cutoff has been used in previous studies [44,47].

Depression

Depression was measured by the 9-item Patient Health Questionnaire (PHQ-9) [48]. The Chinese version has been validated and used in the Chinese population [49,50]. Participants were asked to rate how often they have been bothered by symptoms in the past 2 weeks on a 4-point Likert scale from 0, not at all, to 3, almost every day. The total score ranges from 0 to 27, with a higher score indicating a higher

level of depression. Scores of 0-4, 5-9, 10-14, 15-19, and 20-27 represent minimal, mild, moderate, moderately severe, and severe depression, respectively.

Anxiety

Anxiety was assessed by the 7-item General Anxiety Disorder scale (GAD-7) [51]. It is a brief, self-reported scale for identifying probable cases of generalized anxiety disorder (GAD). The Chinese version has been validated and used in the Chinese population [52,53]. Participants were asked how often they experienced each symptom in the past 2 weeks on a 4-point Likert scale from 0, not at all, to 3, nearly every day. The GAD-7 total score ranges from 0 to 21, with a higher score indicating a higher level of anxiety. Scores of 0-4, 5-9, 10-14, and 15-21 represent minimal, mild, moderate, and severe anxiety, respectively.

Analysis

Descriptive statistics on the participants' sociodemographic and pregnancy-related characteristics, perceived threat, self-efficacy, negative emotions, and prevalence of depression, anxiety, and frequency of face mask wearing are presented. To identify significant factors of depression and anxiety, univariate multinominal logistic regressions were first conducted to examine the association between all factors with depression and anxiety, and respective odds ratios derived from univariate logistic regression (ORus) and 95% confidence intervals are presented. To control for the potential effect of sociodemographic and pregnancy-related characteristics, all sociodemographic characteristics, pregnancy-related characteristics, and independent variables with $P<.05$ in the univariate multinominal regression models were then subjected to multivariate multinominal logistic regression analysis; the resulting multivariate odds ratios (ORms) are reported.

To identify significant factors of always wearing a face mask when going out, univariate logistic regressions were first conducted to examine the association between all factors and the outcome, and the respective ORus and 95% confidence intervals are presented. To control for the potential effect of sociodemographic and pregnancy-related characteristics, all sociodemographic characteristics, pregnancy-related characteristics, and independent variables with $P<.05$ in the univariate logistic regressions were then subjected to a multivariate logistic regression analysis, and the resulting ORms were reported. Data analyses were performed using SPSS version 21.0 (IBM Corporation), with a P value of <.05 being considered statistically significant.

Results

Descriptive Statistics of the Participants

Slightly more than two-thirds (2743/4087, 67.1%) of the participants were aged 30 years or less. Nearly half of the participants (1989/4087, 48.7%) had received a postsecondary level of education. Approximately half of the participants (2022/4087, 49.5%) were nulliparous, and a similar number (1860/4087, 45.5%) were in their third trimester. A small number of participants (6.6%) reported having some pregnancy-related complications (Table 1).

Table 1. Characteristics of the participants (N=4087).

Characteristic	Value, n (%)
Sociodemographic characteristics	
Age (years)	
≥19	57 (1.4)
20-25	872 (21.3)
26-30	1814 (44.4)
31-35	1055 (25.8)
36-40	239 (5.8)
≥41	50 (1.3)
Education level	
Primary or below	111 (2.7)
Junior secondary	1072 (26.2)
Senior secondary	915 (22.4)
Matriculation	1051 (25.7)
Undergraduate	831 (20.3)
Postgraduate or above	107 (2.6)
Pregnancy-related characteristics	
Parity	
Nulliparous	2022 (49.5)
Primiparous	1820 (44.5)
Multiparous	245 (6.0)
Gestational age	
First trimester (12 weeks or below)	855 (21.2)
Second trimester (13-26 weeks)	1362 (33.3)
Third trimester (27 weeks or above)	1860 (45.5)
Pregnancy-related complications	
No	3816 (93.4)
Yes	271 (6.6)
Personal protective behavior	
Frequency of wearing a face mask when going out	
Never	1262 (30.9)
Seldom	1398 (34.2)
Sometimes	453 (11.1)
Always	974 (23.8)
Mental health	
Depression (measured by PHQ-9[a])	
Minimal (0-4)	2098 (51.3)
Mild (5-9)	1163 (28.5)
Moderate (10-14)	504 (12.3)
Moderately severe (15-19)	227 (5.6)
Severe (20-27)	95 (2.3)
Anxiety (measured by GAD-7[b])	

Characteristic	Value, n (%)
Minimal (0-4)	2740 (67.0)
Mild (5-9)	919 (22.5)
Moderate (10-14)	318 (7.8)
Severe (15-21)	110 (2.7)

[a]PHQ-9: 9-item Patient Health Questionnaire.

[b]GAD-7: 7-item General Anxiety Disorder Scale.

Prevalence of Depression, Anxiety and Frequency of Wearing a Face Mask

The prevalence of mild to severe depression (PHQ score >5) and mild to severe anxiety (GAD-7 score >5) was 48.7% (1989/4087) and 33.0% (1347/4087), respectively. Less than a quarter of the participants (974/4087, 23.8%) reported always wearing a face mask when going out (Table 1).

Perceived Threats, Negative Emotions, and Self-Efficacy

One-third of the 4087 participants perceived that they (1313, 32.1%) or their family members (1490, 36.4%) were likely to be infected with COVID-19 (ie, perceived susceptibility), and a sizable number of participants (3216, 78.7%, to 3518, 86.1%) agreed that the disease would have various severe consequences, such as mother-to-child transmission (ie, perceived severity; Table 2). Approximately one-fifth of respondents (894/4087, 21.9%) reported a score of 3 or above on perceived impact (Table 3). Slightly more than half were confident that they could protect themselves (2275/4087, 55.7%) or their family members (2232/4087, 54.6%) from contracting COVID-19. Approximately one-third (1303/4087, 31.9%) reported a high level of fear of the disease, and more than two thirds (2780/4087, 68.0%, to 3056/4087, 74.8%) showed worry about various aspects of COVID-19 (Table 2).

Table 2. Perceived threats, negative emotions, and self-efficacy of pregnant women during the COVID-19 period (N=4087).

Variable	Value, n (%)			
	Very little	Little	Much	Very much
Perceived susceptibility				
Likelihood of oneself contracting COVID-19	788 (19.3)	1986 (48.6)	944 (23.1)	369 (9.0)
Likelihood of one's family members contracting COVID-19	713 (17.4)	1884 (46.1)	1052 (25.7)	438 (10.7)
Perceived severity				
COVID-19 will be transmitted from mother to child	261 (6.4)	610 (14.9)	2114 (51.7)	1102 (27.0)
Maternal infection of COVID-19 will be more difficult to cure than in the general population	212 (5.2)	479 (11.7)	2074 (50.7)	1322 (32.3)
Maternal infection of COVID-19 will affect the health of the child	203 (5.0)	366 (9.0)	1944 (47.6)	1574 (38.5)
Self-efficacy				
Confidence of protecting oneself from contracting COVID-19	478 (11.7)	1334 (32.6)	1588 (38.9)	687 (16.8)
Confidence of protecting family members from contracting COVID-19	443 (10.8)	1412 (34.5)	1601 (39.2)	631 (15.4)
Fear				
Level of fear of COVID-19	666 (16.3)	2118 (51.8)	991 (24.2)	312 (7.6)
Worry				
Worry about being infected with COVID-19 when attending a prenatal checkup	258 (6.3)	773 (18.9)	1983 (48.5)	1073 (26.3)
Worry about one's hospital delivery arrangement being affected due to COVID-19	371 (9.1)	936 (22.9)	1841 (45.0)	939 (23.0)
Worry that accompanied delivery will not be available due to COVID-19	398 (9.7)	865 (21.2)	1844 (45.1)	980 (24.0)
Worry that child health services will be affected after delivery due to COVID-19	339 (8.3)	746 (18.3)	1957 (47.9)	1045 (25.6)

Table 3. Scores of perceived impact of COVID-19 (N=4087).

Total score of perceived impact	Value, n (%)
1	2138 (52.3)
2	1055 (25.8)
3	592 (14.5)
4-5	302 (7.4)

Multinominal Logistic Regression Models for Depression

The results of the univariate multinominal logistic regression analyses showed that among all the background characteristics, being multiparous was a protective factor for mild depression (ORu 0.63, 95% CI 0.45-0.88) and being a farmer was a protective factor for moderately severe depression (ORu 0.72, 95% CI 0.53-0.96) (Table 4). Among the cognitive and psychological variables, perceived susceptibility (ORu 1.17-1.26), perceived severity (ORu 1.14-1.52) and fear (ORu 1.46-2.12) were risk factors for all levels of depression, and perceived impact was a risk factor for mild, moderate, and moderately severe depression (ORu 1.14-1.32). On the other hand, self-efficacy (ORu 0.89, 95% CI 0.82-0.96) was a significant protective factor for moderately severe depression (Table 4). The results of the multivariate multinominal regression analyses showed that after adjusting for significant background variables, perceived severity (ORm 1.08 and 1.09 for mild and moderate depression, respectively), perceived impact (ORm 1.08 and 1.12 for mild and moderate depression, respectively), fear (ORm 1.26-1.70 for all levels of depression), and worry (ORm 1.05 and 1.27 for moderate and severe depression, respectively) were significant risk factors, while self-efficacy was a protective factor (ORm 0.79-0.87 for moderate to severe depression) to different levels of depression (Table 5).

Table 4. Univariate multinominal logistic regression of depression among pregnant women (N=4087).

Variable	ORu[a] (95% CI)			
	Mild depression	Moderate depression	Moderately severe depression	Severe depression
Sociodemographic characteristics				
Age (years)				
≤19[b]	1	1	1	1
20-25	0.94 (0.50-1.77)	1.21 (0.49-3.00)	1.31 (0.39-4.44)	0.63 (0.18-2.21)
26-30	1.08 (0.58-2.00)	1.21 (0.50-2.96)	0.88 (0.26-2.96)	0.36 (0.10-1.23)
31-35	0.97 (0.52-1.81)	1.06 (0.43-2.61)	1.19 (0.35-4.02)	0.37 (0.10-1.31)
36-40	1.05 (0.53-2.06)	1.14 (0.43-3.00)	0.79 (0.20-3.04)	0.47 (0.11-2.00)
≥41	0.38 (0.14-1.07)	0.88 (0.26-3.03)	0.58 (0.09-3.75)	0.59 (0.09-3.76)
Education level				
Primary or below[b]	1	1	1	1
Junior secondary	1.07 (0.65-1.76)	0.70 (0.40-1.23)	0.60 (0.17-2.11)	3.87 (0.52-28.70)
Senior secondary	1.40 (0.85-2.30)	0.85 (0.48-1.50)	0.52 (0.23-1.19)	2.90 (0.38-21.91)
Matriculation	1.54 (0.94-2.52)	0.83 (0.47-1.45)	0.73 (0.33-1.61)	2.17 (0.29-16.48)
Undergraduate	1.61 (0.98-2.65)	0.81 (0.46-1.44)	0.52 (0.23-1.19)	1.87 (0.24-14.52)
Postgraduate or above	1.90 (1.01-3.58)	0.87 (0.39-1.94)	0.60 (0.17-2.11)	2.40 (0.21-27.25)
Employment				
Unemployed/housewife/student[b]	1	1	1	1
Farmer	1.01 (0.86-1.18)	0.96 (0.78-1.18)	0.72 (0.53-0.96)*	0.84 (0.54-1.30)
Employed	0.87 (0.65-1.18)	0.98 (0.67-1.45)	1.16 (0.72-1.88)	0.99 (0.45-2.18)
Pregnancy-related characteristics				
Parity				
Nulliparous[b]	1	1	1	1
Primiparous	0.90 (0.77-1.04)	0.87 (0.71-1.07)	1.30 (0.97-1.73)	1.30 (0.84-1.99)
Multiparous	0.63 (0.45-0.88)**	0.89 (0.59-1.34)	1.59 (0.95-2.67)	1.31 (0.57-2.99)
Gestational age				
First trimester (12 weeks or below)[b]	1	1	1	1
Second trimester (13-26 weeks)	1.00 (0.82-1.22)	0.95 (0.72-1.25)	0.94 (0.65-1.36)	0.75 (0.41-1.38)
Third trimester (27 weeks or above)	0.95 (0.79-1.15)	1.15 (0.89-1.48)	0.82 (0.58-1.17)	1.18 (0.70-2.01)
Pregnancy-related complications				
No[b]	1	1	1	1
Yes	1.12 (0.84-1.48)	1.02 (0.69-1.51)	1.03 (0.59-1.79)	0.81 (0.32-2.02)
Cognitive and psychological variables				
Perceived susceptibility	1.17 (1.12-1.22)***	1.25 (1.18-1.32)***	1.26 (1.17-1.37)***	1.22 (1.09-1.38)**
Perceived severity	1.14 (1.10-1.18)***	1.20 (1.15-1.26)***	1.22 (1.14-1.31)***	1.52 (1.34-1.72)***
Perceived impact	1.14 (1.06-1.22)**	1.20 (1.09-1.33)***	1.32 (1.16-1.51)***	1.13 (0.92-1.39)
Self-efficacy	1.01 (0.97-1.05)	0.96 (0.91-1.02)	0.89 (0.82-0.96)*	0.90 (0.80-1.01)
Fear	1.46 (1.33-1.60)***	1.86 (1.65-2.10)***	2.10 (1.78-2.49)***	2.12 (1.66-2.71)***
Worry	1.09 (1.07-1.12)	1.15 (1.12-1.19)	1.20 (1.15-1.26)	1.41 (1.29-1.53)

[a]ORu: odds ratio derived from univariate multinominal logistic regression.

[b]Reference category.
*P<.05.
**P<.01.
***P<.001.

Table 5. Multivariate multinominal logistic regression of depression among pregnant women (N=4087).

Variable	ORm[a] (95% CI)			
	Mild depression	Moderate depression	Moderately severe depression	Severe depression
Perceived susceptibility	1.05 (0.99-1.12)	1.04 (0.96-1.12)	1.02 (0.92-1.13)	0.93 (0.81-1.07)
Perceived severity	1.09 (1.04-1.14)*	1.08 (1.01-1.15)**	1.02 (0.93-1.12)	1.14 (0.97-1.35)
Perceived impact	1.08 (1.00-1.17)**	1.12 (1.02-1.24)**	1.22 (1.07-1.40)	0.99 (0.81-1.23)
Self-efficacy	0.96 (0.92-1.00)	0.87 (0.82-0.93)*	0.79 (0.72-0.86)*	0.82 (0.72-0.93)**
Fear	1.26 (1.11-1.42)*	1.65 (1.40-1.94)*	1.91 (1.53, 2.40)*	1.70 (1.24-2.32)**
Worry	1.01 (0.98-1.05)	1.05 (1.00-1.10)**	1.11 (1.03-1.19)**	1.27 (1.13-1.43)*

[a]ORm: odds ratio derived from multivariate multinominal logistic regression that included all sociodemographic variables, pregnancy-related variables, and cognitive and psychological variables that were significant at the P<.05 level in the univariate multinominal logistic regression analysis.
*P<.001.
**P<.05.

Multinominal Logistic Regression Models for Anxiety

Results from univariate multinominal logistic regressions showed that among all the background characteristics, being multiparous was a risk factor (ORu 2.51, 95% CI 1.32-4.75), while being a farmer was a protective factor (ORu 60, 95% CI 0.40-0.90) for severe anxiety. Education level (ORu 0.32 to 0.46 for matriculation and undergraduate level) was a protective factor for moderate anxiety. Among the cognitive and psychological variables, perceived susceptibility (ORu 1.20-1.24), perceived severity (ORu 1.19-1.52), fear (ORu 1.62-3.03), and worry (ORu 1.16-1.41) were risk factors for all levels of anxiety, and perceived impact was a risk factor for mild and moderate anxiety (ORu 1.16-1.29). On the other hand, self-efficacy (ORu 0.90-0.93) was a significant protective factor for moderate and severe anxiety (Table 6). The results from the multivariate multinominal regression analysis showed that after adjusting for significant background variables, fear (ORm 1.36-2.88) and worry (ORm 1.09-1.22) were risk factors, while self-efficacy was a protective factor (ORm 0.77-0.90) for all levels of anxiety. Furthermore, perceived severity (ORm 1.07, 95% CI 1.02-1.13) was a risk factor for mild anxiety, and perceived impact (ORm 1.22, 95% CI 1.09-1.37) was a risk factor for moderate anxiety (Table 7).

Table 6. Univariate multinominal logistic regression of anxiety among pregnant women (N=4087).

Variable	ORu[a] (95% CI)		
	Mild anxiety	Moderate anxiety	Severe anxiety
Sociodemographic characteristics			
Age (years)			
≤19[b]	1	1	1
20-25	0.57 (0.22-1.50)	0.99 (0.38-2.61)	0.59 (0.17-2.03)
26-30	0.92 (0.47-1.80)	0.74 (0.29-1.93)	0.32 (0.10-1.10)
31-35	0.75 (0.40-1.41)	0.76 (0.29-2.00)	0.53 (0.15-1.81)
36-40	0.77 (0.41-1.42)	0.57 (0.19-1.70)	0.51 (0.13-2.07)
≥41	0.72 (0.38-1.35)	0.57 (0.13-2.56)	0.63 (0.10-4.00)
Education level			
Primary or below[b]	1	1	1
Junior secondary	0.99 (0.59-1.64)	0.58 (0.32-1.04)	5.25 (0.72-38.54)
Senior secondary	1.21 (0.73-2.02)	0.69 (0.38-1.23)	2.45 (0.32-18.49)
Matriculation	1.19 (0.72-1.98)	0.46 (0.25-0.83)*	2.05 (0.27-15.46)
Undergraduate	1.31 (0.79-2.19)	0.32 (0.17-0.59)**	1.94 (0.25-14.91)
Postgraduate or above	1.36 (0.70-2.63)	0.60 (0.25-1.44)	2.11 (0.19-23.87)
Employment			
Unemployed/housewife/student[b]	1	1	1
Farmer	1.08 (0.92-1.27)	0.78 (0.61-1.00)	0.60 (0.40-0.90)*
Employed	0.66 (0.93-0.68)	1.09 (0.71-1.68)	1.00 (0.51-1.97)
Pregnancy-related characteristics			
Parity			
Nulliparous[b]	1	1	1
Primiparous	0.88 (0.76-1.03)	1.06 (0.84-1.35)	1.32 (0.88-1.98)
Multiparous	0.86 (0.62-1.21)	1.37 (0.86-2.17)	2.51 (1.32-4.75)*
Gestational age			
First trimester (12 weeks or below)[b]	1	1	1
Second trimester (13-26 weeks)	0.97 (0.78-1.20)	1.22 (0.88-1.70)	0.84 (0.48-1.48)
Third trimester (27 weeks or above)	1.31 (1.07-1.60)*	1.21 (0.88-1.66)	1.36 (0.82-2.24)
Pregnancy-related complications			
No[b]	1	1	1
Yes	1.39 (1.04-1.84)*	1.33 (0.86-2.06)	.90 (0.39-2.08)
Cognitive and psychological variables			
Perceived susceptibility	1.20 (1.15-1.26)***	1.28 (1.20-1.37)***	1.24 (1.11-1.38)***
Perceived severity	1.21 (1.16-1.25)***	1.19 (1.12-1.26)***	1.52 (1.35-1.71)***
Perceived impact	1.16 (1.08-1.25)***	1.29 (1.15-1.44)***	1.15 (0.95-1.38)
Self-efficacy	0.98 (0.94-1.02)	0.93 (0.87-0.99)*	0.90 (0.80-0.99)*
Fear	1.62 (1.47-1.78)***	2.13 (1.85-2.45)***	3.03 (2.41-3.81)***
Worry	1.16 (1.13-1.19)***	1.20 (1.15-1.25)***	1.41 (1.30-1.53)***

[a]ORu: Odds ratio derived from univariate multinominal logistic regression.

[b]Reference category.

*P<.05.
**P<.01.
***P<.001.

Table 7. Multivariate multinominal logistic regression of anxiety among pregnant women (N=4087).

Variable	ORm[a] (95% CI)		
	Mild anxiety	Moderate anxiety	Severe anxiety
Perceived susceptibility	1.04 (0.98-1.10)	1.03 (0.94-1.13)	0.88 (0.74-1.02)
Perceived severity	1.07 (1.02-1.13)*	0.99 (0.91-1.07)	1.12 (0.96-1.31)
Perceived impact	1.07 (0.99-1.16)	1.22 (1.09-1.37)**	1.03 (0.85-1.24)
Self-efficacy	0.90 (0.86-0.95)***	0.83 (0.77-0.89)***	0.77 (0.68-0.87)***
Fear	1.36 (1.19-1.55)***	1.88 (1.55-2.29)***	2.88 (2.15-3.88)***
Worry	1.09 (1.05-1.13)***	1.12 (1.05-1.19)***	1.22 (1.09-1.37)***

[a]ORm: odds ratio derived from multivariate multinominal logistic regression that included all sociodemographic variables, pregnancy-related variables, and cognitive and psychological variables that were significant at the P<.05 level in the univariate multinominal logistic regression analysis.
*P<.05.
**P<.01.
***P<.001.

Logistic Regression Models for Always Wearing a Face Mask When Going Out

Results from the univariate logistic regression analyses showed that among all the background characteristics, being in the third trimester (ORu 0.64, 95% CI 0.53-0.77) and pregnancy-related complications (ORu 0.69, 95% CI 0.50-0.95) were risk factors for always wearing a face mask when going out. Having a postgraduate level of education or above was a protective factor for always wearing a face mask (ORu 2.78, 95% CI 1.53-5.05) (Table 8). Among the cognitive and psychological variables, perceived susceptibility (ORu 1.07, 95% CI 1.02-1.11), perceived severity (ORu 1.05, 95% CI 1.02-1.09), and worry (ORu 1.03, 95% CI 1.00-1.05) were associated with always wearing a face mask. The results of the multivariate logistic regression analysis showed that after adjusting for significant background variables, perceived susceptibility was the only significant factor (ORm 1.05, 95% CI 1.00-1.10) for always wearing a face mask (Table 8).

Table 8. Logistic regression of always wearing a face mask when going out among pregnant women (N=4087).

Variable	Always wearing a face mask when going out	
	ORu[a] (95% CI)	ORm[b] (95% CI)
Sociodemographic characteristics		
Age (years)		
≤19[c]	1	N/A[d]
20-25	1.36 (0.68-2.75)	N/A
26-30	1.61 (0.81-3.20)	N/A
31-35	1.32 (0.66-2.65)	N/A
36-40	1.51 (0.72-3.17)	N/A
≥41	2.42 (0.99-5.95)	N/A
Education level		
Primary or below[c]	1	1
Junior secondary	1.98 (0.60-1.58)	0.92 (0.56-1.49)
Senior secondary	0.92 (0.57-1.50)	0.84 (0.51-1.39)
Matriculation	1.31 (0.81-2.12)	1.20 (0.73-1.98)
Undergraduate	1.57 (0.97-2.55)	1.45 (0.87-2.41)
Postgraduate or above	2.78 (1.53-5.05)***	2.43 (1.31-4.54)**
Employment		
Unemployed/housewife/student[c]	1	1
Farmer	1.25 (1.06-1.46)*	0.99 (0.83-1.19)
Employed	0.88 (0.65-1.20)	0.91 (0.66-1.26)
Pregnancy-related characteristics		
Parity		
Nulliparous[c]	1	N/A
Primiparous	0.86 (0.74-1.00)	N/A
Multiparous	0.99 (0.73-1.35)	N/A
Gestational age		
First trimester (12 weeks or below)[c]	1	1
Second trimester (13-26 weeks)	0.86 (0.70-1.04)	0.90 (0.74-1.09)
Third trimester (27 weeks or above)	0.64 (0.53-0.77)***	0.68 (0.56-0.83)***
Pregnancy-related complications		
No[c]	1	1
Yes	0.69 (0.50-0.95)*	0.70 (0.51-0.97)*
Cognitive and psychological variables		
Perceived susceptibility	1.07 (1.02-1.11)**	1.05 (1.00-1.10)*
Perceived severity	1.05 (1.02-1.09)**	1.04 (.99, 1.09)
Perceived impact	1.07 (1.00-1.15)	N/A
Self-efficacy	1.00 (0.97-1.05)	N/A
Fear	1.07 (0.98-1.17)	N/A
Worry	1.03 (1.00-1.05)*	1.02 (0.98-1.05)

[a]ORu: odds ratio derived from univariate logistic regression.

[b]ORm: odds ratio derived from multivariate logistic regression that included all socio-demographic variables, pregnancy-related variables, and cognitive

and psychological variables that were significant at the *P*<.05 level in the univariate logistic regression analysis.

[c]Reference variable.

[d]N/A: not applicable.

*P<.05.

**P<.01.

***P<.001.

Discussion

Principal Findings

The COVID-19 pandemic is an emerging and rapidly evolving situation. People receive information from various sources may form different perceptions about COVID-19. In this study, between 32.1% (1313/4087) and 36.4% (1490/4087) of the participants perceived that they or their family members were susceptible to the disease, and between 78.7% (3216/4087) and 86.1% (3518/4087) believed that the disease would have various serious consequences. Participants also reported that the disease has affected various parts of their lives. The perceived susceptibility and severity of COVID-19 were considerably higher compared to those observed in a study conducted among residents of Wuhan and Shanghai, which showed that 12.5%-18.6% of participants perceived that they were likely to be infected with COVID-19 and 12%-19.9% rated the disease as serious [18]. The figures were also higher compared with those in studies conducted during other epidemics, such as SARS (eg, approximately 20% of pregnant women considered themselves likely to contract SARS during the SARS epidemic period) [29]. Our findings suggest that a substantial number of pregnant women overestimated their risk of contracting COVID-19 and the severity of infection. Furthermore, slightly more than half of the participants reported high levels of self-efficacy. However, this figure was significantly lower than that in the study conducted in Wuhan and Shanghai, which showed that 86.8% to 87.8% of participants perceived a high level of confidence in taking measures to protect themselves from COVID-19 [18].

Negative emotions towards COVID-19 were commonly documented in this study. Approximately one-third of the 4087 participants (1303, 31.9%) showed high levels of fear towards the disease, and more than two thirds (2780, 68.0%, to 3056, 74.8%) reported different worries related to the COVID-19 pandemic, including worries of contracting the infection when attending antenatal checkups and worries that their delivery and child health services would be affected due to the epidemic. The level of negative emotions is comparable to that documented among pregnant women and the general public during the SARS period in Hong Kong, for which studies found that more than half of the participants worried about themselves or their family members contracting SARS and two-thirds of the women were scared of going to the hospital for antenatal visits [29,54]. Due to the high level of transmissibility of COVID-19 and the disruptions that the pandemic has caused in daily life, it is not surprising that high levels of fear and worry related to the disease were reported in this study. The psychological impact of an epidemic outbreak on pregnant women warrants immediate attention.

It is important to note that nearly half of this sample (1989/4087, 48.7%) scored above the cutoff for probable depression, and one-third (1347/4087, 33.0%) scored above the cutoff for probable anxiety. The prevalence was significantly higher than that reported in the general population of pregnant women [55] as well as that reported among the general public during the COVID-19 period [12,17,18]. These findings were also in line with those conducted in countries such as Canada and Japan, which also showed elevated depression and anxiety among pregnant women during the COVID-19 pandemic [56,57]. Our findings suggest that mental health problems were heightened among pregnant women during the COVID-19 pandemic. Furthermore, it is surprising that although a high level of perceived susceptibility, perceived severity, and perceived impact were reported, less than a quarter of the sample (974/4087, 23.8%) reported always wearing a mask when going out. The prevalence of wearing a face mask was substantially lower compared to that reported in studies among pregnant women and the general public during the SARS epidemic (ie, approximately 70%) [29,54], and the one reported in Wuhan and Shanghai during the COVID-19 pandemic [18]. Our findings show that pregnant women reported poor mental health and personal protective behavior during the COVID-19 epidemic. Efforts should be made to provide education to this specific at-risk group.

This study also identified some factors associated with mental health and protective behavior. The results of the multivariate analyses showed that among the cognitive variables, perceived severity and perceived impact of COVID-19 were associated with higher levels of depression and anxiety. These findings are in line with those of previous studies of epidemic outbreaks such as SARS, in which individuals who had a negative appraisal of the disease reported mental health problems [40,46]. Individuals who endorsed higher levels of negative impact of COVID-19 on their daily lives and perceived that they would experience serious consequences of the disease may have been more likely to exhibit negative emotional responses. Furthermore, perceived susceptibility was associated with higher likelihood of wearing a face mask when going out. These findings are consistent with the extant theories, such as protection motivation theory, that higher perceptions of risk and threat are associated with higher levels of engagement in health behaviors [33].

Consistent with previous studies that documented a positive association between self-efficacy and mental health outcomes [37,58], this study found that self-efficacy was associated with lower levels of depression and anxiety. Individuals with higher level of self-efficacy may have better ability to regulate the distress and negative mental health impact of the COVID-19 pandemic. They may be more likely to use more positive coping strategies to manage their emotions when encountering adversities associated with the disease. Surprisingly, despite

the extensive evidence on the positive role of self-efficacy in explaining a range of health behaviors, self-efficacy had no significant association with wearing a face mask when going out in this study. In this study, self-efficacy was conceptualized as the perceived capability of protecting oneself or one's family members from being infected with COVID-19; it was not behavior-specific in nature and may therefore have failed to predict face mask–wearing behavior.

Consistent with the literature that documented the negative impact of negative emotions and mental health [59,60], fear and worry were associated with higher levels of depression and anxiety in this study. However, no significant association between fear and worry with always wearing a mask when going out was found in this study. The findings suggest that the current sample of pregnant women may be more likely to regulate their negative emotions through mental health rather than behavioral response. More research is needed to elucidate the association between negative emotions and mental and behavioral outcomes under different contexts.

Implications

The findings of this study provide implications for health care professionals and policy makers to address the threats posed to pregnant women by COVID-19 and by any possible future incidences of epidemic outbreaks. As poor health of pregnant women will lead to undesirable outcomes of both the mother and child, interventions to mitigate the negative effects of COVID-19 in this population is highly essential. First, this study found that cognitive perceptions on perceived threats of COVID-19 were associated with wearing a mask. To promote realistic risk perceptions and effective precautions of COVID-19, communication through various channels is essential. There is a need for governments and health care professionals to provide scientific information on the transmissibility and consequences of the disease and the efficacy of protective measures. However, because perceived threats are also associated with worse mental health outcomes, it is important to aim these health messages at increasing individuals' threat-related beliefs while at the same time reducing misconceptions and adverse emotional and mental health outcomes.

Increasing self-efficacy should also be considered, as it has been found to be associated with better mental health. This increase can be achieved through cognitive restructuring of perceived capability in self-care and reappraisal of stressors related to the COVID-19 epidemic, as well as by providing encouragement and role modelling of positive behaviors. Digital psychotherapy, namely internet cognitive behavioral therapy (iCBT), has been demonstrated to be effective in combating the growing prevalence of mental health problems related to COVID-19 [61]. Previous studies have found that iCBT is efficacious in treating depression, anxiety, and insomnia [61,62]. In the context of COVID-19, iCBT could be a useful means for pregnant women to obtain mental health support without concern about contracting the disease from face-to-face contact [63]. As health care professionals are well-placed to recognize potential mental health problems at an early stage, routine screening for depression and other mental health conditions in pregnant women should be considered in obstetrical settings during the COVID-19 pandemic. Moreover, support from family, friends, and health care professionals can also promote the mental health of pregnant women by enhancing their perceived efficacy in coping with the disease.

Finally, the negative associations between negative emotions and mental health also suggest that health education should seek ways to disseminate a realistic level of risk that will not induce excessive worry and fear. It has also been shown that uncertainty can cause fear during pregnancy [64]. It is important to equip women with adequate knowledge about COVID-19 to reduce their fear and worry regarding the disease. Emotion regulation training can also be offered to improve their skills in regulating negative emotions.

Limitations

The study was cross-sectional in nature, so causality cannot be assumed. Data were collected in a few cities; therefore, the findings may not be representative of the entire population of pregnant women in China. This study mainly used self-reported questionnaires to measure psychiatric symptoms and did not make clinical diagnoses. The standard for establishing a psychiatric diagnosis involves a structured clinical interview and functional neuroimaging [65,66]; therefore, the prevalence of mental health problems might have been overestimated. In addition, as participation was voluntary, the provinces that agreed to take part might have provided better services and support to pregnant women during the pandemic, leading to bias in the mental health and protective behaviors of the sample. Caution is needed when generalizing the findings to the population of pregnant women in China. In addition, the income levels of the pregnant women were not recorded. Lastly, as no validated scale on measuring perceptions of COVID-19 is available, variables were assessed by self-developed items with references to previous studies on other epidemics (eg, SARS), and some variables (ie, level of fear) were measured by a single item. The reliability of the items should be interpreted with caution.

Conclusion

This study provided important insights on pregnant women's perceptions of and emotional reactions to COVID-19 as well as their potential influence on behavioral and mental health. This study demonstrated that pregnant women reported a high level of perceived threat towards COVID-19; however, the frequency at which they wore a mask when going out was suboptimal. Most of the women showed negative emotions and mental responses. Our findings provide an important guide for health care professionals and policy makers to develop strategies to alleviate the negative effects of the COVID-19 epidemic.

Conflicts of Interest

None declared.

References

1. Coronavirus disease (COVID-19) weekly epidemiological update and weekly operational update. World Health Organization. URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports [accessed 2021-03-26]
2. Chen S, Yang J, Yang W, Wang C, Bärnighausen T. COVID-19 control in China during mass population movements at New Year. Lancet 2020 Mar 07;395(10226):764-766 [FREE Full text] [doi: 10.1016/S0140-6736(20)30421-9] [Medline: 32105609]
3. Tan W, Hao F, McIntyre R, Jiang L, Jiang X, Zhang L, et al. Is returning to work during the COVID-19 pandemic stressful? A study on immediate mental health status and psychoneuroimmunity prevention measures of Chinese workforce. Brain Behav Immun 2020 Jul;87:84-92 [FREE Full text] [doi: 10.1016/j.bbi.2020.04.055] [Medline: 32335200]
4. Doing a good job in the prevention and control of pneumonia outbreaks in children and pregnant women with new coronavirus infection. Webpage in Chinese. National Health Commission of the People's Republic of China. URL: http://www.gov.cn/xinwen/2020-02/02/content_5473939.htm [accessed 2021-03-26]
5. Liu H, Wang L, Zhao S, Kwak-Kim J, Mor G, Liao A. Why are pregnant women susceptible to COVID-19? An immunological viewpoint. J Reprod Immunol 2020 Jun;139:103122 [FREE Full text] [doi: 10.1016/j.jri.2020.103122] [Medline: 32244166]
6. Matar R, Alrahmani L, Monzer N, Debiane LG, Berbari E, Fares J, et al. Clinical Presentation and Outcomes of Pregnant Women With Coronavirus Disease 2019: A Systematic Review and Meta-analysis. Clin Infect Dis 2021 Feb 01;72(3):521-533 [FREE Full text] [doi: 10.1093/cid/ciaa828] [Medline: 32575114]
7. Yang Z, Wang M, Zhu Z, Liu Y. Coronavirus disease 2019 (COVID-19) and pregnancy: a systematic review. J Matern Fetal Neonatal Med 2020 Apr 30:1-4. [doi: 10.1080/14767058.2020.1759541] [Medline: 32354293]
8. Ashraf MA, Keshavarz P, Hosseinpour P, Erfani A, Roshanshad A, Pourdast A, et al. Coronavirus disease 2019 (COVID-19): a systematic review of pregnancy and the possibility of vertical transmission. J Reprod Infertil 2020;21(3):157-168 [FREE Full text] [Medline: 32685412]
9. Delahoy MJ, Whitaker M, O'Halloran A, Chai SJ, Kirley PD, Alden N, COVID-NET Surveillance Team. Characteristics and maternal and birth outcomes of hospitalized pregnant women with laboratory-confirmed COVID-19 - COVID-NET, 13 States, March 1-August 22, 2020. MMWR Morb Mortal Wkly Rep 2020 Sep 25;69(38):1347-1354 [FREE Full text] [doi: 10.15585/mmwr.mm6938e1] [Medline: 32970655]
10. Karimi-Zarchi M, Neamatzadeh H, Dastgheib SA, Abbasi H, Mirjalili SR, Behforouz A, et al. Vertical transmission of coronavirus disease 19 (COVID-19) from infected pregnant mothers to neonates: a review. Fetal Pediatr Pathol 2020 Jun;39(3):246-250 [FREE Full text] [doi: 10.1080/15513815.2020.1747120] [Medline: 32238084]
11. Gao Y, Ye L, Zhang J, Yin Y, Liu M, Yu H, et al. Clinical features and outcomes of pregnant women with COVID-19: a systematic review and meta-analysis. BMC Infect Dis 2020 Aug 03;20(1):564 [FREE Full text] [doi: 10.1186/s12879-020-05274-2] [Medline: 32746801]
12. Zhu H, Wang L, Fang C, Peng S, Zhang L, Chang G, et al. Clinical analysis of 10 neonates born to mothers with 2019-nCoV pneumonia. Transl Pediatr 2020 Feb;9(1):51-60 [FREE Full text] [doi: 10.21037/tp.2020.02.06] [Medline: 32154135]
13. Jafari M, Pormohammad A, Sheikh Neshin SA, Ghorbani S, Bose D, Alimohammadi S, et al. Clinical characteristics and outcomes of pregnant women with COVID-19 and comparison with control patients: A systematic review and meta-analysis. Rev Med Virol 2021 Jan 02:e2208 [FREE Full text] [doi: 10.1002/rmv.2208] [Medline: 33387448]
14. Wu X, Sun R, Chen J, Xie Y, Zhang S, Wang X. Radiological findings and clinical characteristics of pregnant women with COVID-19 pneumonia. Int J Gynaecol Obstet 2020 Jul;150(1):58-63. [doi: 10.1002/ijgo.13165] [Medline: 32270479]
15. Clinical management of COVID-19: interim guidance, 27 May 2020. World Health Organization. URL: https://apps.who.int/iris/handle/10665/332196 [accessed 2021-03-26]
16. Wang C, Pan R, Wan X, Tan Y, Xu L, Ho CS, et al. Immediate psychological responses and associated factors during the initial stage of the 2019 coronavirus disease (COVID-19) epidemic among the general population in China. Int J Environ Res Public Health 2020 Mar 06;17(5):1729 [FREE Full text] [doi: 10.3390/ijerph17051729] [Medline: 32155789]
17. Huang Y, Zhao N. Generalized anxiety disorder, depressive symptoms and sleep quality during COVID-19 epidemic in China: a web-based cross-sectional survey. medRxiv. Preprint posted online on March 09, 2020. [doi: 10.1101/2020.02.19.20025395]
18. Qian M, Wu Q, Wu P, Hou Z, Liang Y, Cowling B. Psychological responses, behavioral changes and public perceptions during the early phase of the COVID-19 outbreak in China: a population based cross-sectional survey. medRxiv. Preprint posted online on February 22, 2020. [doi: 10.1101/2020.02.18.20024448]
19. Le XTT, Dang AK, Toweh J, Nguyen QN, Le HT, Do TTT, et al. Evaluating the Psychological Impacts Related to COVID-19 of Vietnamese People Under the First Nationwide Partial Lockdown in Vietnam. Front Psychiatry 2020;11:824 [FREE Full text] [doi: 10.3389/fpsyt.2020.00824] [Medline: 32982807]

20. Tee ML, Tee CA, Anlacan JP, Aligam KJG, Reyes PWC, Kuruchittham V, et al. Psychological impact of COVID-19 pandemic in the Philippines. J Affect Disord 2020 Dec 01;277:379-391 [FREE Full text] [doi: 10.1016/j.jad.2020.08.043] [Medline: 32861839]
21. Tran BX, Nguyen HT, Le HT, Latkin CA, Pham HQ, Vu LG, et al. Impact of COVID-19 on economic well-being and quality of life of the Vietnamese during the national social distancing. Front Psychol 2020;11:565153. [doi: 10.3389/fpsyg.2020.565153] [Medline: 33041928]
22. Hessami K, Romanelli C, Chiurazzi M, Cozzolino M. COVID-19 pandemic and maternal mental health: a systematic review and meta-analysis. J Matern Fetal Neonatal Med 2020 Nov 01:1-8. [doi: 10.1080/14767058.2020.1843155] [Medline: 33135523]
23. Durankuş F, Aksu E. Effects of the COVID-19 pandemic on anxiety and depressive symptoms in pregnant women: a preliminary study. J Matern Fetal Neonatal Med 2020 May 18:1-7. [doi: 10.1080/14767058.2020.1763946] [Medline: 32419558]
24. Ceulemans M, Hompes T, Foulon V. Mental health status of pregnant and breastfeeding women during the COVID-19 pandemic: A call for action. Int J Gynaecol Obstet 2020 Oct;151(1):146-147. [doi: 10.1002/ijgo.13295] [Medline: 32620037]
25. Ravaldi C, Ricca V, Wilson A, Homer C, Vannacci A. Previous psychopathology predicted severe COVID-19 concern, anxiety, and PTSD symptoms in pregnant women during "lockdown" in Italy. Arch Womens Ment Health 2020 Dec;23(6):783-786 [FREE Full text] [doi: 10.1007/s00737-020-01086-0] [Medline: 33215247]
26. Glover V. Maternal depression, anxiety and stress during pregnancy and child outcome; what needs to be done. Best Pract Res Clin Obstet Gynaecol 2014 Jan;28(1):25-35. [doi: 10.1016/j.bpobgyn.2013.08.017] [Medline: 24090740]
27. Waters CS, Hay DF, Simmonds JR, van Goozen SHM. Antenatal depression and children's developmental outcomes: potential mechanisms and treatment options. Eur Child Adolesc Psychiatry 2014 Oct 19;23(10):957-971. [doi: 10.1007/s00787-014-0582-3] [Medline: 25037152]
28. Wang C, Chudzicka-Czupała A, Grabowski D, Pan R, Adamus K, Wan X, et al. The association between physical and mental health and face mask use during the COVID-19 Pandemic: a comparison of two countries with different views and practices. Front Psychiatry 2020;11:569981 [FREE Full text] [doi: 10.3389/fpsyt.2020.569981] [Medline: 33033485]
29. Lee DTS, Sahota D, Leung TN, Yip ASK, Lee FFY, Chung TKH. Psychological responses of pregnant women to an infectious outbreak: a case-control study of the 2003 SARS outbreak in Hong Kong. J Psychosom Res 2006 Nov;61(5):707-713 [FREE Full text] [doi: 10.1016/j.jpsychores.2006.08.005] [Medline: 17084150]
30. Prentice-Dunn S, Rogers RW. Protection motivation theory and preventive health: beyond the health belief model. Health Educ Res 1986;1(3):153-161. [doi: 10.1093/her/1.3.153]
31. Bashirian S, Jenabi E, Khazaei S, Barati M, Karimi-Shahanjarini A, Zareian S, et al. Factors associated with preventive behaviours of COVID-19 among hospital staff in Iran in 2020: an application of the Protection Motivation Theory. J Hosp Infect 2020 Jul;105(3):430-433 [FREE Full text] [doi: 10.1016/j.jhin.2020.04.035] [Medline: 32360337]
32. Kim JK, Crimmins EM. Age differences in the relationship between threatening and coping mechanisms and preventive behaviors in the time of COVID-19 in the United States: Protection Motivation Theory. Res Psychother 2020 Dec 31;23(3):485 [FREE Full text] [doi: 10.4081/ripppo.2020.485] [Medline: 33585294]
33. Floyd D, Prentice-dunn S, Rogers R. A Meta-Analysis of Research on Protection Motivation Theory. J Appl Social Pyschol 2000 Feb;30(2):407-429. [doi: 10.1111/j.1559-1816.2000.tb02323.x]
34. Bish A, Michie S. Demographic and attitudinal determinants of protective behaviours during a pandemic: a review. Br J Health Psychol 2010 Dec 24;15:797-820. [doi: 10.1348/135910710x485826]
35. Li J, Yang A, Dou K, Wang L, Zhang M, Lin X. Chinese public's knowledge, perceived severity, and perceived controllability of the COVID-19 and their associations with emotional and behavioural reactions, social participation, and precautionary behaviour: a national survey. PsyArXiv. Preprint posted online on February 27, 2020. [doi: 10.31234/osf.io/5tmsh]
36. Yıldırım M, Geçer E, Akgül Ö. The impacts of vulnerability, perceived risk, and fear on preventive behaviours against COVID-19. Psychol Health Med 2021 Jan;26(1):35-43. [doi: 10.1080/13548506.2020.1776891] [Medline: 32490689]
37. Mak WW, Law RW, Woo J, Cheung FM, Lee D. Social support and psychological adjustment to SARS: the mediating role of self-care self-efficacy. Psychol Health 2009 Mar;24(2):161-174. [doi: 10.1080/08870440701447649] [Medline: 20186649]
38. Liao Q, Cowling B, Lam WT, Ng MW, Fielding R. Situational awareness and health protective responses to pandemic influenza A (H1N1) in Hong Kong: a cross-sectional study. PLoS One 2010 Oct 12;5(10):e13350 [FREE Full text] [doi: 10.1371/journal.pone.0013350] [Medline: 20967280]
39. Liao Q, Cowling BJ, Lam WWT, Fielding R. Factors affecting intention to receive and self-reported receipt of 2009 pandemic (H1N1) vaccine in Hong Kong: a longitudinal study. PLoS One 2011 Mar 11;6(3):e17713 [FREE Full text] [doi: 10.1371/journal.pone.0017713] [Medline: 21412418]
40. Ho SMY, Kwong-Lo RSY, Mak CWY, Wong JS. Fear of severe acute respiratory syndrome (SARS) among health care workers. J Consult Clin Psychol 2005 Apr;73(2):344-349. [doi: 10.1037/0022-006x.73.2.344]
41. Smith CA, Ellsworth PC. Patterns of cognitive appraisal in emotion. J Pers Soc Psychol 1985 Apr;48(4):813-838. [Medline: 3886875]

42. Shields GS, Moons WG, Tewell CA, Yonelinas AP. The effect of negative affect on cognition: anxiety, not anger, impairs executive function. Emotion 2016 Sep;16(6):792-797 [FREE Full text] [doi: 10.1037/emo0000151] [Medline: 27100367]
43. Lau JT, Griffiths S, Choi KC, Tsui HY. Avoidance behaviors and negative psychological responses in the general population in the initial stage of the H1N1 pandemic in Hong Kong. BMC Infect Dis 2010 May 28;10:139 [FREE Full text] [doi: 10.1186/1471-2334-10-139] [Medline: 20509887]
44. Lau JTF, Griffiths S, Choi K, Lin C. Prevalence of preventive behaviors and associated factors during early phase of the H1N1 influenza epidemic. Am J Infect Control 2010 Jun;38(5):374-380 [FREE Full text] [doi: 10.1016/j.ajic.2010.03.002] [Medline: 20569849]
45. Wong LP, Sam I. Behavioral responses to the influenza A(H1N1) outbreak in Malaysia. J Behav Med 2011 Mar;34(1):23-31 [FREE Full text] [doi: 10.1007/s10865-010-9283-7] [Medline: 20680674]
46. Bonanno GA, Ho SMY, Chan JCK, Kwong RSY, Cheung CKY, Wong CPY, et al. Psychological resilience and dysfunction among hospitalized survivors of the SARS epidemic in Hong Kong: a latent class approach. Health Psychol 2008 Sep;27(5):659-667. [doi: 10.1037/0278-6133.27.5.659] [Medline: 18823193]
47. Mo PK, Lau JT. Illness representation on H1N1 influenza and preventive behaviors in the Hong Kong general population. J Health Psychol 2015 Dec;20(12):1523-1533. [doi: 10.1177/1359105313516031] [Medline: 24423574]
48. Bian C, Li C, Duan Q, Wu H. Reliability and validity of patient health questionnaire: depressive syndrome module for outpatients. Sci Res Essays 2011;6(2):278-282 [FREE Full text]
49. Zhang Y, Ting R, Lam M, Lam J, Nan H, Yeung R, et al. Measuring depressive symptoms using the Patient Health Questionnaire-9 in Hong Kong Chinese subjects with type 2 diabetes. J Affect Disord 2013 Nov;151(2):660-666. [doi: 10.1016/j.jad.2013.07.014] [Medline: 23938133]
50. Li Z, Dai J, Wu N, Gao J, Fu H. The mental health and depression of rural-to-urban migrant workers compared to non-migrant workers in Shanghai: a cross-sectional study. Int Health 2019 Oct 31;11(S1):S55-S63. [doi: 10.1093/inthealth/ihz081] [Medline: 31670817]
51. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
52. Zeng QZ, He YL, Liu H, Miao JM, Chen JX, Xu HN, et al. Reliability and validity of Chinese version of the Generalized Anxiety Disorder 7-item (GAD-7) scale in screening anxiety disorders in outpatients from traditional Chinese internal department. Chinese Mental Health Journal 2013;27(3):163-168.
53. Ying DG, Jiang S, Yang H, Zhu S. Frequency of generalized anxiety disorder in Chinese primary care. Postgrad Med 2010 Jul;122(4):32-38. [doi: 10.3810/pgm.2010.07.2173] [Medline: 20675969]
54. Lau JTF, Yang X, Pang E, Tsui HY, Wong E, Wing YK. SARS-related perceptions in Hong Kong. Emerg Infect Dis 2005 Mar;11(3):417-424 [FREE Full text] [doi: 10.3201/eid1103.040675] [Medline: 15757557]
55. Parcells D. Women's mental health nursing: depression, anxiety and stress during pregnancy. J Psychiatr Ment Health Nurs 2010 Nov;17(9):813-820. [doi: 10.1111/j.1365-2850.2010.01588.x] [Medline: 21040227]
56. Lebel C, MacKinnon A, Bagshawe M, Tomfohr-Madsen L, Giesbrecht G. Elevated depression and anxiety symptoms among pregnant individuals during the COVID-19 pandemic. J Affect Disord 2020 Dec 01;277:5-13 [FREE Full text] [doi: 10.1016/j.jad.2020.07.126] [Medline: 32777604]
57. Matsushima M, Horiguchi H. The COVID-19 pandemic and mental well-being of pregnant women in Japan: need for economic and social policy interventions. Disaster Med Public Health Prep 2020 Sep 10:1-6 [FREE Full text] [doi: 10.1017/dmp.2020.334] [Medline: 32907687]
58. Lowe NK. Self-efficacy for labor and childbirth fears in nulliparous pregnant women. J Psychosom Obstet Gynaecol 2000 Dec;21(4):219-224. [doi: 10.3109/01674820009085591] [Medline: 11191169]
59. Johnson M, Folkman S. Side effect and disease related symptom representations among HIV+ adults on antiretroviral therapy. Psychol Health Med 2004 May;9(2):139-148. [doi: 10.1080/13548500410001670672]
60. Frostholm L, Oernboel E, Christensen KS, Toft T, Olesen F, Weinman J, et al. Do illness perceptions predict health outcomes in primary care patients? A 2-year follow-up study. J Psychosom Res 2007 Mar;62(2):129-138. [doi: 10.1016/j.jpsychores.2006.09.003] [Medline: 17270570]
61. Ho CS, Chee CY, Ho RC. Mental health strategies to combat the psychological impact of COVID-19 beyond paranoia and panic. Ann Acad Med Singap 2020 Mar 16;49(3):155-160 [FREE Full text] [Medline: 32200399]
62. Soh HL, Ho RC, Ho CS, Tam WW. Efficacy of digital cognitive behavioural therapy for insomnia: a meta-analysis of randomised controlled trials. Sleep Med 2020 Nov;75:315-325. [doi: 10.1016/j.sleep.2020.08.020] [Medline: 32950013]
63. Zhang MWB, Ho RCM. Moodle: the cost effective solution for internet cognitive behavioral therapy (I-CBT) interventions. Technol Health Care 2017;25(1):163-165. [doi: 10.3233/THC-161261] [Medline: 27689560]
64. Melender H. Fears and coping strategies associated with pregnancy and childbirth in Finland. J Midwifery Womens Health 2002 Jul 08;47(4):256-263. [doi: 10.1016/s1526-9523(02)00263-5]
65. Ho CSH, Lim LJH, Lim AQ, Chan NHC, Tan RS, Lee SH, et al. Diagnostic and predictive applications of functional near-infrared spectroscopy for major depressive disorder: a systematic review. Front Psychiatry 2020;11:378 [FREE Full text] [doi: 10.3389/fpsyt.2020.00378] [Medline: 32477179]

66. Husain SF, Yu R, Tang T, Tam WW, Tran B, Quek TT, et al. Validating a functional near-infrared spectroscopy diagnostic paradigm for Major Depressive Disorder. Sci Rep 2020 Jun 16;10(1):9740 [FREE Full text] [doi: 10.1038/s41598-020-66784-2] [Medline: 32546704]

Abbreviations

GAD: generalized anxiety disorder
GAD-7: 7-item General Anxiety Disorder scale
iCBT: internet cognitive behavioral therapy
ORm: multivariate odds ratio
ORu: odds ratio derived from univariate logistic regression
PHQ-9: 9-item Patient Health Questionnaire
QR: quick response
SARS: severe acute respiratory syndrome

Edited by G Eysenbach; submitted 02.09.20; peer-reviewed by R Ho, Y Huang, K Linden; comments to author 14.01.21; revised version received 03.02.21; accepted 08.02.21; published 12.04.21.

Please cite as:
Mo PKH, Fong VWI, Song B, Di J, Wang Q, Wang L
Association of Perceived Threat, Negative Emotions, and Self-Efficacy With Mental Health and Personal Protective Behavior Among Chinese Pregnant Women During the COVID-19 Pandemic: Cross-sectional Survey Study
J Med Internet Res 2021;23(4):e24053
URL: https://www.jmir.org/2021/4/e24053
doi:10.2196/24053
PMID:33729983

Original Paper

Assessing Public Interest Based on Wikipedia's Most Visited Medical Articles During the SARS-CoV-2 Outbreak: Search Trends Analysis

Jędrzej Chrzanowski[1*]; Julia Sołek[1,2*], MD; Wojciech Fendler[1], MD, PhD; Dariusz Jemielniak[3], PhD

[1]Department of Biostatistics and Translational Medicine, Medical University of Łódź, Łódź, Poland

[2]Department of Pathology, Medical University of Łódź, Łódź, Poland

[3]Management in Networked and Digital Societies, Kozminski University, Warszawa, Poland

[*]these authors contributed equally

Corresponding Author:
Wojciech Fendler, MD, PhD
Department of Biostatistics and Translational Medicine
Medical University of Łódź
Mazowiecka 15
Łódź, 92-215
Poland
Phone: 48 422722585
Email: wojciech.fendler@umed.lodz.pl

Related Article:

This is a corrected version. See correction statement: https://www.jmir.org/2021/4/e29598

Abstract

Background: In the current era of widespread access to the internet, we can monitor public interest in a topic via information-targeted web browsing. We sought to provide direct proof of the global population's altered use of Wikipedia medical knowledge resulting from the new COVID-19 pandemic and related global restrictions.

Objective: We aimed to identify temporal search trends and quantify changes in access to Wikipedia Medicine Project articles that were related to the COVID-19 pandemic.

Methods: We performed a retrospective analysis of medical articles across nine language versions of Wikipedia and country-specific statistics for registered COVID-19 deaths. The observed patterns were compared to a forecast model of Wikipedia use, which was trained on data from 2015 to 2019. The model comprehensively analyzed specific articles and similarities between access count data from before (ie, several years prior) and during the COVID-19 pandemic. Wikipedia articles that were linked to those directly associated with the pandemic were evaluated in terms of degrees of separation and analyzed to identify similarities in access counts. We assessed the correlation between article access counts and the number of diagnosed COVID-19 cases and deaths to identify factors that drove interest in these articles and shifts in public interest during the subsequent phases of the pandemic.

Results: We observed a significant (P<.001) increase in the number of entries on Wikipedia medical articles during the pandemic period. The increased interest in COVID-19–related articles temporally correlated with the number of global COVID-19 deaths and consistently correlated with the number of region-specific COVID-19 deaths. Articles with low degrees of separation were significantly similar (P<.001) in terms of access patterns that were indicative of information-seeking patterns.

Conclusions: The analysis of Wikipedia medical article popularity could be a viable method for epidemiologic surveillance, as it provides important information about the reasons behind public attention and factors that sustain public interest in the long term. Moreover, Wikipedia users can potentially be directed to credible and valuable information sources that are linked with the most prominent articles.

(J Med Internet Res 2021;23(4):e26331) doi:10.2196/26331

KEYWORDS

COVID-19; pandemic; media; Wikipedia; internet; online health information; information seeking; interest; retrospective; surveillance; infodemiology; infoveillance

Introduction

After the COVID-19 pandemic outbreak began, a new concern for public health emerged—predicting and preventing the spread of the disease. The increased media coverage on the COVID-19 pandemic focused the public's attention and likely affected the most popular internet search terms, thus altering people's behavior worldwide [1,2]. The public consumption of COVID-19 information in digital media is directly associated with preventive behaviors, including regularly washing hands with soap and water, staying away from crowded places, and wearing face masks in public [3]. Bragazzi et al [4] have suggested that internet search trend data can be used to build predictive models of disease spread and help with containing the pandemic. Bento et al [5] has shown that the internet search term "coronavirus" increased in popularity immediately after the day of the first COVID-19 case announcement. However, the term's popularity returned to the baseline level in less than 1-2 weeks. After this period, other terms that pertained to community-level policies (ie, quarantine, school closures, and COVID-19 tests) or personal health strategies (ie, masks, grocery delivery, and over-the-counter medications) emerged as the most-searched terms [5]. Other studies have shown that public interest in specific search terms is more associated with reported deaths and media coverage than with real epidemiologic situations [6]. Moreover, it has been observed that people quickly experience information overload, which results in information avoidance [7]. Therefore, it seems logical to identify the most credible and reliable sources and use them to inform the public during the "information-hungry" period of any subsequent pandemics that may occur.

Wikipedia is considered a key web-based source of health information, and people are more willing to seek information that is published in Wikipedia than information from any other health websites [8]. Wikipedia's quality has been a highly debated topic, and many researchers are highly skeptical of the platform [9]. Nevertheless, in 2005, a comparative study that was published in *Nature* reported that Wikipedia was competing head-to-head with Britannica [10]. Since then, many different studies on Wikipedia's accuracy have shown that it is quite on par with published professional sources, as it provides reliable information [11,12]. Studies have reported that the quality of medical information that is available in Wikipedia is consistently high [13,14]. It has also been reported that Wikipedia is more reliable than several published sources, despite its low readability [15]. Medical articles in Wikipedia are based on reliable sources and mainly skew toward the most prominent academic journals [16]. More and more scholars have embraced the use of Wikipedia in the classroom [17,18]. The wide acceptance of Wikipedia is also related to institutional recognition. The American Psychological Association and the Association for Psychological Science have encouraged their members to edit Wikipedia articles [19,20].

With regard to medical information, Wikipedia's popularity exceeds that of the National Health Service, WebMD, Mayo Clinic, and World Health Organization (WHO) websites combined. This is mostly due to the highly accurate information that is provided by the editors of Wikiproject Medicine [21]. Wikiprojects are edited by self-organized groups of volunteers who are involved in curating information on a specific topic [13]. Since their reliability in providing medical information is so high, the WHO has partnered with the Wikimedia Foundation to expand access to trusted COVID-19–related information on Wikipedia [22]. Several studies have indicated that medical students who use Wikipedia to prepare for their exams receive better grades than students who only rely on textbooks [23]. Wikipedia has become a vital tool for global public health promotion [24].

The quality and popularity of Wikipedia's medical content make analyzing the most popular medical articles extremely interesting from a medical professional's point of view. Additionally, changes in viewership and peaks in interest are of high relevance to public health specialists. Given that Wikipedia has over 300 language variations that differ significantly in terms of rules, culture, and the presentation of knowledge [25], it is also interesting to compare changes in the popularity of Wikipedia medical articles across languages. Hence, to better understand public interest in high-quality health information during the COVID-19 pandemic, we conducted an analysis of daily visits to Wikipedia articles that were published from 2015 to 2020 and sought to identify possible factors for attracting public attention.

Methods

Detecting and Quantifying the Surge in COVID-19–Related Searches

We collected a list of 37,880 articles that were curated by the English Wikipedia Medicine Project. We derived the daily access counts (ie, from July 1, 2015, to September 13, 2020) of these articles by using ToolForge (ie, a pageview analytics tool). Daily access was defined as each visit to a given article on a given date. These data did not contain user-specific information.

We limited our selection to the 100 most accessed English Wikipedia Medicine Project articles that were published from July 1, 2015, to September 12, 2020. No other filters were applied to included Wikipedia articles. By using an interwiki mechanism (ie, cross-references for different language versions of Wikipedia), we identified articles that reported on the same subject in nine other language versions of Wikipedia (ie, French, German, Swedish, Dutch, Russian, Italian, Spanish, Polish, and Vietnamese). The articles were matched, and their daily access counts from July 1, 2015, to September 13, 2020 were obtained. These data were stored in an Excel sheet that used English Wikipedia article names. Extracted daily access data are provided in Multimedia Appendix 1.

To provide additional context with regard to the ongoing COVID-19 pandemic, we also obtained pandemic-related data. We decided not to use reports on the total number of daily confirmed SARS-CoV-2 infection cases due to the possible influencing effect of introducing different public testing schemes. Other factors, such as the total number of daily deaths, are less likely to be altered by regional politics and health care discrepancies. Thus, we decided to measure the effect of the COVID-19 pandemic by analyzing the total number of global deaths and region-specific deaths, which were defined as the cumulative number of deaths resulting from SARS-CoV-2 infection for all reported countries and language-specific countries, respectively. These data were obtained from standardized public information in the Our World in Data University of Oxford initiative website [26].

The preprocessing of Wikipedia article access data was performed in Python 3.8, Excel version 2011 (Microsoft Corporation). Due to the influences of day-to-day differences in Wikipedia access and absolute differences in access to the different language versions of Wikipedia, we standardized access data by calculating daily article access as percentages of language-specific access to Wikipedia on a given day.

To identify deviations in the stability of article visits, we chose articles that exhibited the least altered access patterns throughout the investigated period. Articles were selected based on their relative stability (ie, the SD of percent article access in a 30-day moving window divided by the 30-day moving average [ie, mean] of percent article access), which was calculated for the full duration of investigated period. Relative stability was assessed to identify articles that exhibited relatively unchanged access patterns throughout the studied period. Reference articles were defined as the 20 most stable articles across all languages. These articles provided the highest mean daily access percentages for all evaluated periods. As such, a reference article could be used for direct comparisons with other highly accessed articles. They also provided metrics that were the least affected by changes in Wikipedia use.

Statistical Analysis

A statistical analysis was conducted to identify access patterns in Wikipedia and Wikipedia Medicine Project articles that were published in 2020 and to determine these patterns' association with the COVID-19 pandemic. First, we investigated the association between Wikipedia access before and during the COVID-19 pandemic. Second, we determined whether there were specific articles of interest before and during the COVID-19 pandemic (ie, excluding the years that were associated with other epidemics that were covered in media). We also investigated whether the total number of global and regional deaths resulting from SARS-CoV-2 were associated with increased access to articles of interest during the COVID-19 pandemic.

We also investigated whether there was a difference between navigation to Wikipedia articles of interest before and during the COVID-19 pandemic. To this end, we determined the minimum number of links required to navigate between two articles within Wikipedia (ie, the degree of separation [DOS]). The statistical analysis was conducted in Python 3.8 and Statistica 13.3 (Statsoft, TIBCO Software Inc, Dell Inc). The DOS was assessed by using the PHP (hypertext preprocessor) language and jQuery library, which were implemented in the Degrees of Wikipedia tool [27]. To facilitate the straightforward representation of DOS results, we provided the DOSs of articles of interest that were found across the highest number of Wikipedia language versions. Due to the observed lack of a normal distribution in article access, we used nonparametric methods. We considered a P value of <.05 to be statistically significant.

Abnormalities in Wikipedia access patterns that occurred from July 1, 2015, to September 13, 2020, were investigated by using the Chi-square goodness-of-fit test. To identify the qualitative changes in the annual top 10 most accessed Wikipedia Medicine Project articles, we used word clouds as graphical representations of access patterns.

Identifying Shifts in Search Patterns Before and During the COVID-19 Pandemic

We determined whether the COVID-19 pandemic altered temporal article access patterns. This was achieved by using a k-nearest neighbor (kNN) unsupervised clustering method for analyzing selected Wikipedia language versions. As we were interested in the patterns of access to articles that were of increased user interest in a given period, we performed separate unsupervised clustering analyses for the 2017-2019 and 2020 periods. We excluded the 2015-2016 period due to the possible influence of the Zika virus epidemic, which could have promoted interests in the general population that were similar to interests during the COVID-19 pandemic. As such, we were able to identify articles of user interest during the 2017-2019 and 2020 periods.

We provided a tabular representation of kNN-recognized articles that were associated with increased access before and during the COVID-19 pandemic. The weights that were detected by the kNN algorithm were used to analyze monthly changes in the public's interest in overrepresented articles during the COVID-19 pandemic and these articles' relation to the total number of global and regional deaths resulting from SARS-CoV-2 infection. This correlation was tested by performing a Spearman rank correlation analysis. We further divided the data by month (ie, March to September 2020) to evaluate whether the correlation changed between months.

We compared navigation patterns between articles of interest that were published before and during the COVID-19 pandemic by using the DOSs of reference articles. These patterns were tested with the Kruskall-Wallis test, Dunn posthoc test, and U Mann-Whitney test.

Results

Characteristics of Wikipedia Medicine Project Articles

We analyzed visits to Wikipedia Medicine Project articles that were published in 10 language versions of Wikipedia from 2015 to 2020. These languages included English, French, German, Swedish, Dutch, Russian, Italian, Spanish, Polish, and

Vietnamese. We analyzed 100 articles from English Wikipedia after mapping the other language versions of Wikipedia (ie, 100 articles from French Wikipedia, 98 from Dutch Wikipedia, 98 from Spanish Wikipedia, 97 from Russian Wikipedia, 96 from Italian Wikipedia, 95 from German Wikipedia, 94 from Polish Wikipedia, and 93 from Vietnamese Wikipedia). Due to its low coverage (ie, <90 matched articles, as determined via the interwiki mechanism), Swedish Wikipedia was removed from the analysis. Wikipedia articles without coverage across all languages included the following: “Bed bug” (25% coverage), “Coronavirus” (50% coverage), “Adderall” (62.5% coverage), “Bubonic plague” (62.5% coverage), “Chronic traumatic encephalopathy” (62.5% coverage); “Plantar fasciitis” (75% coverage), “Alprazolam” (87.5% coverage), “Cancer” (87.5% coverage), “Elizabeth Holmes” (87.5% coverage), “Escitalopram” (87.5% coverage), “Ketogenic diet” (87.5% coverage), “Lorazepam” (87.5% coverage), “Project MKUltra” (87.5% coverage), “Psychopathy” (87.5% coverage), and “Trypophobia” (87.5% coverage). Full lists of articles that are available for each language are provided in Multimedia Appendix 2, Supplementary Table S1.

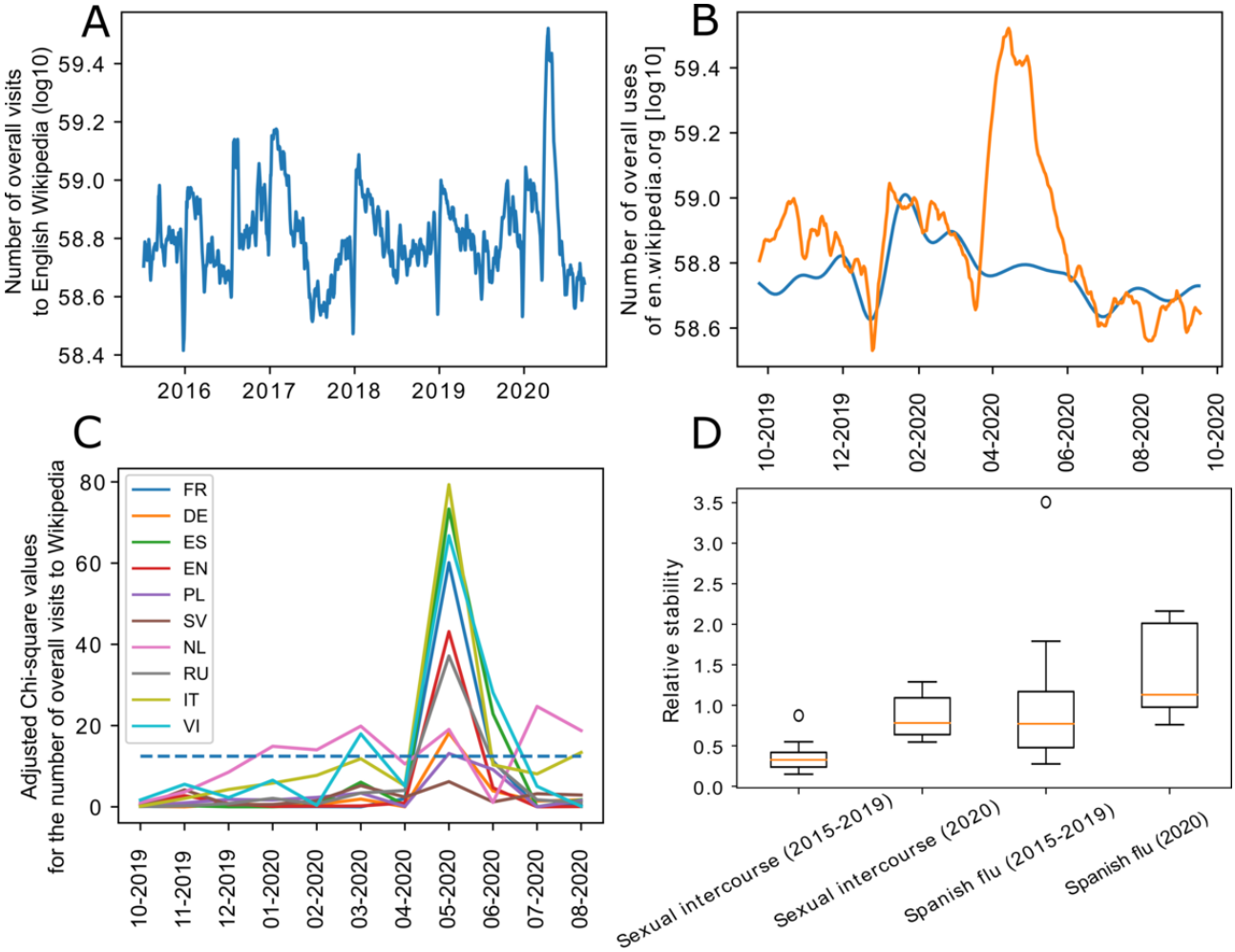
In the 2015-2020 period, there were a total of 2,258,621,012 Wikipedia entries in the top 100 most accessed articles and 775,791,941,485 overall visits to Wikipedia across all selected languages (Wikipedia access across all languages: median 0.41%; range 0.29%-0.46%). Annual access to Wikipedia significantly and gradually decreased from 2015 to 2020 (r=−0.0513; P=.03). Significant increases in use (P<.001) were observed for the English (r=0.0619), Polish (r=0.1237), Dutch (r=0.0481), Italian (r=0.2197), and Vietnamese (r=0.7480) versions of Wikipedia. Significant decreases in use (P<.001) were observed for the German (r=−0.2654), Spanish (r=−0.1085), and Russian (r=−0.4338) versions of Wikipedia. For the top 100 most accessed Wikipedia medical articles, we observed a 1.5-fold to 3.8-fold increase in access across all language versions of Wikipedia in 2020 (ie, compared to access in previous years).

Surge in Article Visits During Early 2020

To determine the effect of the COVID-19 pandemic on Wikipedia use, we used a graphical representation of total daily Wikipedia access (Figure 1). Due to the suspected increase in Wikipedia use during 2020, we constructed a generalized additive model by using a limited-memory Broyden-Fletcher-Goldfarb-Shanno fitting algorithm that was implemented in the FBProphet tool to predict Wikipedia use in 2020 based on previous access patterns. We based the model on data from the 2015-2019 period to identify expected behaviors in 2020 and compare them to observed access patterns. The global pattern for Wikipedia access was indicative of seasonal behaviors (Figure 1). This pattern changed across all language versions of Wikipedia in March 2020, and it returned to normal in September 2020, as predicted by our models. The Chi-square goodness-of-fit test values for monthly Wikipedia access had P values of <.001 across all eight languages (Figure 1).

To determine how abnormal Wikipedia access altered the relative search stability of articles, we needed to pick a point of reference. We determined that the “Sexual intercourse” article demonstrated high yet stable monthly access within the evaluated period. By comparing the “Sexual intercourse” article to the “Spanish flu” article, we were able to determine that the relative access to these articles before the COVID-19 pandemic changed during the pandemic (“Spanish flu” article: P<.04; “Sexual intercourse” article: P<.001; Figure 1).

Figure 1. The disruption in annual Wikipedia visit patterns. (A) The pattern of general English Wikipedia access from 2015 to 2020 (Multimedia Appendix 3). Data for other language versions of Wikipedia are provided in Multimedia Appendix 4, Supplementary Figures S1a-S9a. (B) The FBProphet prediction model was created based on data from the 2015-2019 period. The model compared expected behaviors in 2020 (ie, the blue line) to observed access (ie, the orange line). Data for the other language versions of Wikipedia are provided in Multimedia Appendix 4, Supplementary Figures S1b-S9b. (C) A summary of monthly general access to Wikipedia across all Wikipedia language versions from September 2019 to September 2020. Solid lines represent Chi-square goodness-of-fit values and the dashed line represents the cutoff value. (D) The stability of the two reference Wikipedia articles across all languages in the 2015-2019 and 2020 periods (ie, the moving SD divided by mean percent access across a 30-day window). DE: German; EN: English; ES: Spanish; FR: French; IT: Italian; NL: Dutch; PL: Polish; RU: Russian; SV: Swedish; VI: Vietnamese.



Effects of the COVID-19 Pandemic on the Annual Top 10 Most Commonly Accessed Articles in Wikipedia

We provided a graphical representation of the annual top 10 most accessed articles by creating word clouds for articles that were published from 2015 to 2020 (Multimedia Appendix 3, Multimedia Appendix 4, Supplementary Figures S10-S17). English Wikipedia word clouds are shown in Multimedia Appendix 3, and those for other language versions of Wikipedia are shown in Multimedia Appendix 4, Supplementary Figures S10-S17. We investigated how the 10 most viewed articles changed each year. In the 2015-2019 period, we observed a set of articles that were constantly among the top 10 most accessed articles each year. These articles were as follows: “Leonardo da Vinci” (appearance frequency: 37/45), “Asperger syndrome” (appearance frequency: 36/45), and “Bipolar disorder” (appearance frequency: 28/45). Compared to previous years, there was a distinctive difference in the top 10 most accessed articles in 2020; the “Pandemic” (appearance frequency: 9/9), “COVID-19 pandemic” (appearance frequency: 9/9), “Coronavirus disease 2019” (appearance frequency: 8/9), and “Spanish flu” (appearance frequency: 8/9) articles became the most frequently accessed articles. We found 8 new articles among the top 10 most accessed articles in 2020 that have never appeared in this list. Of these articles, 2 were created after the SARS-CoV-2 pandemic (ie, the “COVID-19 pandemic” and “Coronavirus disease 2019” articles), while the other 6 were present within Wikipedia before the pandemic (ie, the ”Pandemic,” “Spanish flu,” “Coronavirus,” “Bubonic plague,” ”Influenza,” and “World Health Organization” articles).

Identifying Cross-Language Similarities in Article Access Prior to and During the COVID-19 Pandemic via Unsupervised Clustering

The across-language abnormality that was observed in Wikipedia and Wikipedia Medicine Project access frequency in 2020 was further investigated. By conducting an unsupervised analysis, we identified differences between article access before and during the COVID-19 pandemic. Articles that were published in 2020 and recognized by the kNN algorithm were

considered COVID-19–related articles. These included the following eight articles: “Coronavirus,” “Spanish flu,” “Coronavirus disease 2019,” “COVID-19 pandemic, “Influenza,” “Pandemic,” “Virus,” and “Black Death.” The group of articles that were unrelated to COVID-19 included 25 articles. We selected the four articles from this group that were present in most language versions of Wikipedia (ie, the “Sexual intercourse,” “Leonardo da Vinci,” “Bipolar disorder,” and “Borderline personality disorder” articles).

We compared the DOSs of COVID-19–related articles to the DOSs of articles that were unrelated to COVID-19. We used the “COVID-19 pandemic” article as a point of reference for COVID-19–related articles and the “Sexual intercourse” article as a point of reference for articles unrelated to COVID-19. The analysis showed that the DOSs between articles unrelated to COVID-19 and the “Sexual intercourse” article were significantly higher than the DOSs between COVID-19–related articles and the “COVID-19 pandemic” article (Kruskall-Wallis test and posthoc Dunn test: $P<.001$; Table 1). This indicated that COVID-19–related articles were more closely connected, which we suspected due to the observed similarities in themes.

We performed a Spearman rank correlation analysis on articles to provide additional context. We observed that articles unrelated to COVID-19 had higher median correlation values across languages than those in the COVID-19–related article group (Table 1). In the COVID-19–related article group, the highest median correlation value (ie, across all languages) was found between the “COVID-19 pandemic” and “Coronavirus disease 2019” articles (R=0.7022), and the lowest median correlation value was found between the “COVID-19 pandemic” and “Influenza” articles (R=0.0330). In the group of articles unrelated to COVID-19, correlation values were higher. The highest median correlation value was found between the “Sexual intercourse” and “Bipolar disorder” articles (R=0.8657), and the lowest median correlation value was found between the “Sexual intercourse” and “Leonardo da Vinci” articles (R=0.4631).

We also compared the DOSs between all articles that were analyzed in detail and the “COVID-19 pandemic” article. COVID-19–related articles had significantly lower DOSs compared to articles unrelated to COVID-19 (*U* Mann-Whitney test: $P<.001$; Table 1).

Finally, we investigated how the patterns in articles access (ie, those identified by unsupervised clustering) were associated with the total number of global and region-specific deaths resulting from SARS-CoV-2 infection. To this end, we standardized each measure. We used ordinary least squares linear regression to identify correlations across each month in the March to September 2020 period (Figure 2, Multimedia Appendix 2, Supplementary Table S3).

Upon further investigation, we found that the kNN-derived pattern in COVID-19–related articles’ access was significantly associated with both the total number of global and region-specific deaths resulting from SARS-CoV-2 infection for articles across all Wikipedia language versions (Figure 2). There was a notable difference in the correlation between article access and the total number of deaths resulting from SARS-CoV-2 infection, which appeared linear for region-specific deaths and negatively exponential for global deaths (Figure 2). This was also reflected by the low absolute Spearman rank coefficients between article access and the total number of global deaths resulting from SARS-CoV-2 infection for the months after June 2020 (Figure 2).

Table 1. Degrees of separation (DOSs) between investigated articles and reference articles within and across article clusters for the 2016-2019 and 2020 periods.

Wikipedia articles	Prevalence[a], %	Spearman R[b] (within groups)	DOSs within groups[c] (IQR)	DOSs across articles[d] (IQR)
COVID-19–related article group				
Coronavirus	80	0.2770	1 (1-1)	1 (1-1)
COVID-19 pandemic	77.8	Reference[e]	Reference[e]	Reference[e]
Spanish flu	55.6	0.4451	2 (1.75-2)	2 (1.75-2)
Coronavirus disease 2019	55.6	0.7022	1 (1-1.25)	1 (1-1.25)
Pandemic	44.4	0.4341	1 (1-1)	1 (1-1)
Black death	33.3	0.4332	2 (2-2)	2 (2-2)
Virus	11.1	0.3851	2 (1-2)	2 (1-2)
Influenza	11.1	0.0330	1.5 (1-2)	1.5 (1-2)
Group of articles unrelated to COVID-19				
Sexual intercourse	88.9	Reference[e]	Reference[e]	3 (2-3)
Leonardo da Vinci	88.9	0.4631	2 (1.5-2)	2 (2-2.5)
Bipolar disorder	77.8	0.8657	3 (2-3)	3 (2.75-3)
Borderline personality disorder	77.8	0.8352	2.5 (2-3)	3 (3-3)

[a]Refers to the presence of an article in k-nearest neighbor–determined groups across all available languages.

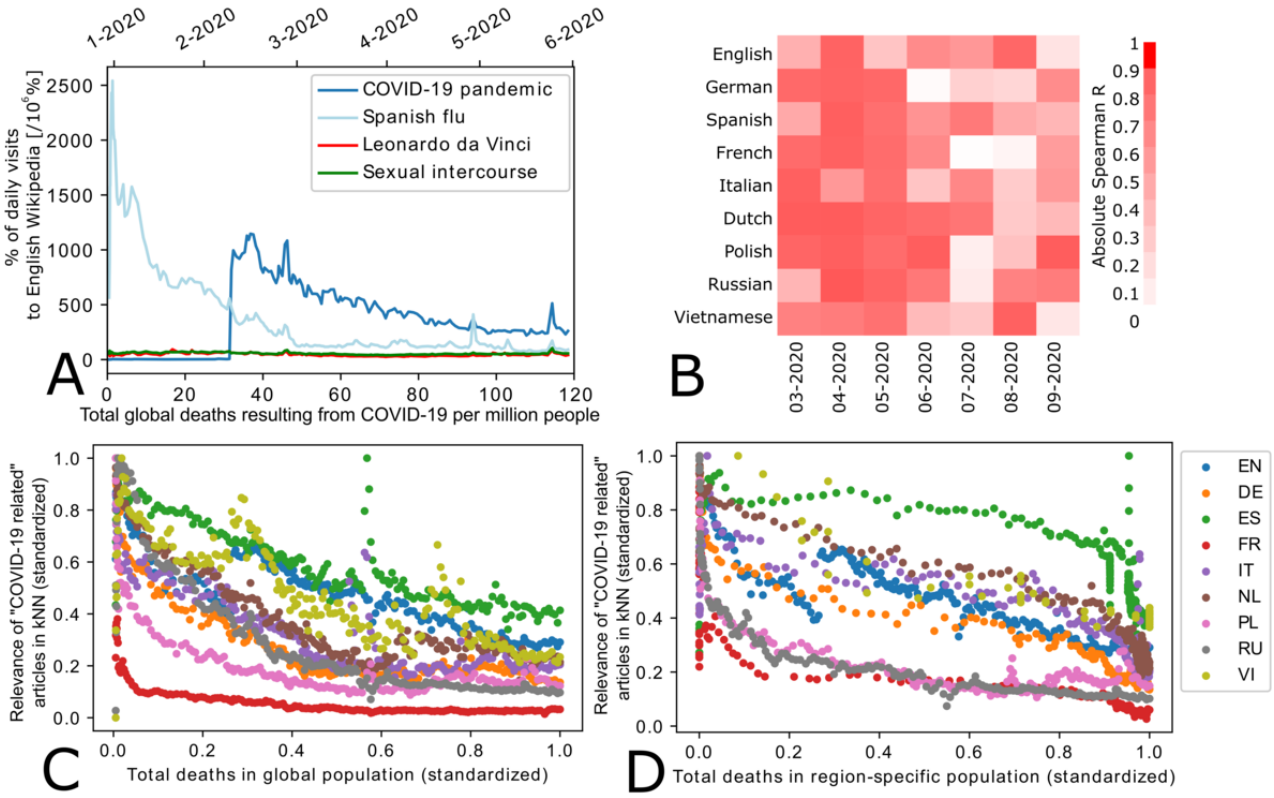
[b]The Spearman R value for an article.

[c]The DOS of an article within a k-nearest neighbor–determined group.

[d]The DOS across all articles that were investigated in detail.

[e]The article of reference for calculating DOSs.

Figure 2. Access to COVID-19–related Wikipedia articles in 2020 and the total number of deaths resulting from SARS-CoV-2 infection. (A) The percentage of daily article access (ie, the "COVID-19 pandemic," "Spanish flu," "Leonardo da Vinci," and "Sexual intercourse" articles) to English Wikipedia and total number of global deaths resulting from SARS-CoV-2 infection (ie, per 1 million people). Data for other language versions of Wikipedia are provided in Multimedia Appendix 4, Supplementary Figures S18-S25. (B) Heatmap of Spearman absolute regression coefficients for COVID-19–related articles and the total number of global deaths resulting from SARS-CoV-2 infection across languages and months. (C) Access to COVID-19–related articles (ie, the kNN-determined relevance values across selected Wikipedia language versions) versus the total number of region-specific deaths resulting from SARS-CoV-2 infection. The graph shows correlations between region-specific deaths and COVID-19–related articles. (D) Access to COVID-19–related articles (ie, the kNN-determined relevance values across selected Wikipedia language versions) versus the total number of global deaths resulting from SARS-CoV-2 infection. DE: German; EN: English; ES: Spanish; FR: French; IT: Italian; kNN: k-nearest neighbor; NL: Dutch; PL: Polish; RU: Russian; SV: Swedish; VI: Vietnamese.



Discussion

Principal Results

Our study shows that the pandemic has significantly influenced patterns in Wikipedia access. There was an apparent surge in interest for infectious disease, which somewhat surprisingly and quickly declined despite the mounting death toll. This is the first study on the impact of the previously recognized change in society's interest, which resulted from the COVID-19 pandemic. In this study, this impact was reflected by changes in Wikipedia access patterns. Moreover, the analysis of different Wikipedia language versions allowed us to confirm that the observed effect of these changes is reflected by both the regional and global impacts of the COVID-19 pandemic. Our study provides proof that an unsupervised clustering method for analyzing data on daily access to medical information could be used to identify interests in global health issues.

Recent studies have indicated that the global lockdown is a possible reason for the change in internet use [28]. We noticed a pronounced increase in the total number of visits to Wikipedia in March 2020. Not only did we observe more frequent visits to Wikipedia, but we also observed a distinctive change in searched topics. The list of the top 10 most accessed articles in 2020 included articles that have not appeared in such lists during previous years. Moreover, the DOS between most of the articles of interest was low, suggesting that navigation was relevant to the observed change. This finding, combined with the correlations between article access, indicated that during the pandemic, people had an increased interest in topics that were not directly related to COVID-19 but were related to pandemics in general. Correlations in access to the "Spanish flu," "Black Death," and "Bubonic plague" articles show that the general public has rapidly gained interest in the topic of previous infectious disease–related health crises. We suspect that these observed correlations are associated with the COVID-19 Wikiproject, which curates articles to provide users with valuable information. The COVID-19 Wikiproject, which is managed by 1200 editors, has resulted in the addition of more than 6500 entries to Wikipedia [29].

Our study shows that similar changes in public interests have already occurred in recent years. We noticed an increase in access to the 2016 "Zika virus" article in 5 of the 9 investigated language versions of Wikipedia. This could be associated with the epidemic that was announced by the WHO in November 2016. This association was also confirmed in the Bragazzi et al [30] and Hickmann et al [31] studies on H1N1-related and Zika virus–related articles and these viruses' respective outbreaks.

Other studies have focused on the increased amount of searches for terms related to anosmia (ie, a characteristic symptom of COVID-19) and have shown its correlation with the SARS-CoV-2 outbreak in many countries [32]. Moreover, trends in the amount of searches for the terms “wash hands,” “hand sanitizer,” and “antiseptic” have successfully predicted the rising number of COVID-19 cases in many countries, which indicates that search trends can potentially be used in epidemiologic surveillance [33]. In our analysis, we noticed an increase in the popularity of a Wikipedia article about the WHO. This result may indicate that the general public has decided to learn more about the organization, making the WHO a focal point for the outbranching of searches. Conversely, this increase in popularity could be linked with President Donald Trump’s attempts to discredit the WHO for the way that they handled the COVID-19 outbreak. These attempts were highly amplified by mainstream media and Twitter [34].

We were interested in identifying the aspect of the pandemic that specifically drove people’s growing interest in COVID-19 from March to June 2020. Our unsupervised clustering analysis showed that the total number of global deaths significantly correlated with the temporal increase in the number of searches for COVID-19–related articles (ie, from March to June 2020). Interestingly, despite the continuous rise in the number of global deaths, the number of visits to COVID-19–related articles decreased after June 2020. This could be interpreted as a gradual decline in global interest toward the pandemic. Data on interest in COVID-19–related articles suggest that there is an increased interest in contagious diseases and that an opportunity to raise awareness through Wikipedia-centered strategies will arise during future outbreaks of infectious diseases. However, the stable correlation between the region-specific number of deaths and the frequency of searches for COVID-19–related articles that persisted until September 2020 is a noteworthy finding. A study by Gozii et al [35] showed that public attention and response are mostly driven by media coverage instead of disease spread. Interestingly, they also showed that the media sharply shifted their focus toward domestic situations as soon as the first death was confirmed in a person’s home country [35]. Therefore, we conclude that reporting about region-specific death reshapes individuals’ perceptions of risk and significantly impacts public interest, thereby affecting the information-seeking behaviors of people from affected areas.

Our study's limitations include the fact that our analysis was restricted to article access data, which did not contain detailed information on user-specific article access patterns or users’ demographic characteristics (eg, age, gender, and educational status). However, we refrained from performing such analyses to maintain the privacy of Wikipedia users. Our study proves that it is possible to identify the global and regional effects of a health crisis on people's behavior. The in-depth analysis of users’ behaviors can be unethical due to its potential for targeting region-specific populations and spreading disinformation, which are ever-present concerns due to the malicious spread of fake news.

Language bias may have also slightly skewed our results, as preferences for English may have resulted in an underreporting of native language–specific searches. Moreover, several idiomatic phrases that are specific to the English vocabulary—“black death” being the prime example—may not directly translate to other languages. To mitigate this bias, we assumed that analyzing data from English Wikipedia would account for most of the global population’s language preferences. Further, we used search data from the other language versions of Wikipedia to validate our results. This strategy yielded a cohesive and surprisingly uniform sample of COVID-19–related search data.

Conclusions

Our results support the idea that Wikipedia can be used as a tool for successfully surveilling trends in public interest. The increase in interest toward COVID-19–related articles was followed by a progressive decline. This shows that the potential optimal window for efficient information dissemination via Wikipedia is the early phase of a pandemic. Wikipedia articles that directly link to major articles about a major health crisis can contribute to the spread of global anxiety and the promotion of prevention behaviors. Therefore, Wikipedia articles should be carefully selected.

Acknowledgments

DJ’s participation was possible thanks to a grant (grant number: 2019/35/B/HS6/01056) from the Polish National Science Centre.

Conflicts of Interest

Author DJ is a non-paid, volunteer member of the Board of Trustees of Wikimedia Foundation, a non-profit publisher of Wikipedia. All the other authors have no conflicts to declare.

Multimedia Appendix 1
Open data.
[XLSX File (Microsoft Excel File), 22109 KB - jmir_v23i4e26331_app1.xlsx]

Multimedia Appendix 2
Supplementary tables S1-S3.
[XLSX File (Microsoft Excel File), 70 KB - jmir_v23i4e26331_app2.xlsx]

Multimedia Appendix 3
Word cloud.
[PDF File (Adobe PDF File), 327 KB - jmir_v23i4e26331_app3.pdf]

Multimedia Appendix 4
Supplementary figures S1-S25.
[PDF File (Adobe PDF File), 1291 KB - jmir_v23i4e26331_app4.pdf]

References

1. Ashrafi-Rizi H, Kazempour Z. Information typology in coronavirus (COVID-19) crisis; a commentary. Arch Acad Emerg Med 2020 Mar 12;8(1):e19 [FREE Full text] [Medline: 32185370]
2. Ashrafi-Rizi H, Kazempour Z. Information diet in Covid-19 crisis; a commentary. Arch Acad Emerg Med 2020 Mar 28;8(1):e30 [FREE Full text] [Medline: 32232215]
3. Liu PL. COVID-19 information seeking on digital media and preventive behaviors: The mediation role of worry. Cyberpsychol Behav Soc Netw 2020 Oct;23(10):677-682. [doi: 10.1089/cyber.2020.0250] [Medline: 32498549]
4. Bragazzi NL, Dai H, Damiani G, Behzadifar M, Martini M, Wu J. How big data and artificial intelligence can help better manage the COVID-19 pandemic. Int J Environ Res Public Health 2020 May 02;17(9):3176 [FREE Full text] [doi: 10.3390/ijerph17093176] [Medline: 32370204]
5. Bento AI, Nguyen T, Wing C, Lozano-Rojas F, Ahn YY, Simon K. Evidence from internet search data shows information-seeking responses to news of local COVID-19 cases. Proc Natl Acad Sci U S A 2020 May 26;117(21):11220-11222 [FREE Full text] [doi: 10.1073/pnas.2005335117] [Medline: 32366658]
6. Sousa-Pinto B, Anto A, Czarlewski W, Anto JM, Fonseca JA, Bousquet J. Assessment of the impact of media coverage on COVID-19-related Google Trends data: Infodemiology study. J Med Internet Res 2020 Aug 10;22(8):e19611 [FREE Full text] [doi: 10.2196/19611] [Medline: 32530816]
7. Soroya SH, Farooq A, Mahmood K, Isoaho J, Zara SE. From information seeking to information avoidance: Understanding the health information behavior during a global health crisis. Inf Process Manag 2021 Mar;58(2):102440 [FREE Full text] [doi: 10.1016/j.ipm.2020.102440] [Medline: 33281273]
8. Laurent MR, Vickers TJ. Seeking health information online: does Wikipedia matter? J Am Med Inform Assoc 2009;16(4):471-479 [FREE Full text] [doi: 10.1197/jamia.M3059] [Medline: 19390105]
9. Jemielniak D. Wikipedia: Why is the common knowledge resource still neglected by academics? Gigascience 2019 Dec 01;8(12):giz139 [FREE Full text] [doi: 10.1093/gigascience/giz139] [Medline: 31794014]
10. Giles J. Internet encyclopaedias go head to head. Nature 2005 Dec 15;438(7070):900-901. [doi: 10.1038/438900a] [Medline: 16355180]
11. Dunn PK, Marshman M, McDougall R. Evaluating Wikipedia as a self-learning resource for statistics: You know they'll use it. Am Stat 2018 Jun 11;73(3):224-231. [doi: 10.1080/00031305.2017.1392360]
12. Heilman JM, Kemmann E, Bonert M, Chatterjee A, Ragar B, Beards GM, et al. Wikipedia: a key tool for global public health promotion. J Med Internet Res 2011 Jan 31;13(1):e14 [FREE Full text] [doi: 10.2196/jmir.1589] [Medline: 21282098]
13. Mesgari M, Okoli C, Mehdi M, Nielsen F, Lanamäki A. “The sum of all human knowledge”: A systematic review of scholarly research on the content of Wikipedia. J Assoc Inf Sci Technol 2014 Dec 02;66(2):219-245. [doi: 10.1002/asi.23172]
14. Shafee T, Masukume G, Kipersztok L, Das D, Häggström M, Heilman J. Evolution of Wikipedia's medical content: past, present and future. J Epidemiol Community Health 2017 Nov;71(11):1122-1129 [FREE Full text] [doi: 10.1136/jech-2016-208601] [Medline: 28847845]
15. Matheson D, Matheson-Monnet C. Wikipedia as informal self-education for clinical decision-making in medical practice. Open Med J 2017 Sep 30;4(1):15-25 [FREE Full text] [doi: 10.2174/1874220301704010015]
16. Jemielniak D, Masukume G, Wilamowski M. The most influential medical journals according to Wikipedia: Quantitative analysis. J Med Internet Res 2019 Jan 18;21(1):e11429 [FREE Full text] [doi: 10.2196/11429] [Medline: 30664451]
17. Konieczny P. Teaching with Wikipedia in a 21st - century classroom: Perceptions of Wikipedia and its educational benefits. J Assoc Inf Sci Technol 2016 Mar 15;67(7):1523-1534. [doi: 10.1002/asi.23616]
18. Jemielniak D, Aibar E. Bridging the gap between wikipedia and academia. J Assoc Inf Sci Technol 2016 Apr 04;67(7):1773-1776. [doi: 10.1002/asi.23691]
19. Breckler S. Don’t like Wikipedia? Change it. American Psychological Association. 2010 Dec. URL: https://www.apa.org/science/about/psa/2010/12/wikipedia-change [accessed 2020-12-06]
20. Mahzarin B. Harnessing the power of Wikipedia for scientific psychology: A Call to Action. Association for Psychological Science. 2011 Feb 11. URL: https://www.psychologicalscience.org/observer/harnessing-the-power-of-wikipedia-for-scientific-psychology-a-call-to-action [accessed 2020-12-06]
21. James R. WikiProject Medicine: creating credibility in consumer health. J Hosp Librariansh 2016 Oct 18;16(4):344-351. [doi: 10.1080/15323269.2016.1221284]

22. The World Health Organization and Wikimedia Foundation expand access to trusted information about COVID-19 on Wikipedia. World Health Organization. 2020 Oct 22. URL: https://www.who.int/news/item/22-10-2020-the-world-health-organization-and-wikimedia-foundation-expand-access-to-trusted-information-about-covid-19-on-wikipedia [accessed 2020-12-06]
23. Azzam A, Bresler D, Leon A, Maggio L, Whitaker E, Heilman J, et al. Why medical schools should embrace Wikipedia: Final-year medical student contributions to Wikipedia articles for academic credit at one school. Acad Med 2017 Feb;92(2):194-200 [FREE Full text] [doi: 10.1097/ACM.0000000000001381] [Medline: 27627633]
24. Jemielniak D, Wilamowski M. Cultural diversity of quality of information on Wikipedias. J Assoc Inf Sci Technol 2017 Aug 07;68(10):2460-2470. [doi: 10.1002/asi.23901]
25. Reavley NJ, Mackinnon AJ, Morgan AJ, Alvarez-Jimenez M, Hetrick SE, Killackey E, et al. Quality of information sources about mental disorders: a comparison of Wikipedia with centrally controlled web and printed sources. Psychol Med 2012 Aug;42(8):1753-1762. [doi: 10.1017/S003329171100287X] [Medline: 22166182]
26. Coronavirus pandemic (COVID-19) - statistics and research. Our World in Data. URL: https://ourworldindata.org/coronavirus [accessed 2020-09-17]
27. Link two articles on Wikipedia. Degrees of Wikipedia. URL: https://degreesofwikipedia.com/ [accessed 2021-03-10]
28. Brodeur A, Clark AE, Fleche S, Powdthavee N. COVID-19, lockdowns and well-being: Evidence from Google Trends. J Public Econ 2021 Jan;193:104346 [FREE Full text] [doi: 10.1016/j.jpubeco.2020.104346] [Medline: 33281237]
29. Benjakob O. Why Wikipedia is immune to coronavirus. Haaretz. 2020 Apr 08. URL: https://www.haaretz.com/us-news/.premium.MAGAZINE-why-wikipedia-is-immune-to-coronavirus-1.8751147 [accessed 2020-12-06]
30. Bragazzi NL, Alicino C, Trucchi C, Paganino C, Barberis I, Martini M, et al. Global reaction to the recent outbreaks of Zika virus: Insights from a Big Data analysis. PLoS One 2017 Sep 21;12(9):e0185263. [doi: 10.1371/journal.pone.0185263] [Medline: 28934352]
31. Hickmann KS, Fairchild G, Priedhorsky R, Generous N, Hyman JM, Deshpande A, et al. Forecasting the 2013-2014 influenza season using Wikipedia. PLoS Comput Biol 2015 May 14;11(5):e1004239. [doi: 10.1371/journal.pcbi.1004239] [Medline: 25974758]
32. Walker A, Hopkins C, Surda P. Use of Google Trends to investigate loss-of-smell-related searches during the COVID-19 outbreak. Int Forum Allergy Rhinol 2020 Jul;10(7):839-847 [FREE Full text] [doi: 10.1002/alr.22580] [Medline: 32279437]
33. Lin YH, Liu CH, Chiu YC. Google searches for the keywords of "wash hands" predict the speed of national spread of COVID-19 outbreak among 21 countries. Brain Behav Immun 2020 Jul;87:30-32 [FREE Full text] [doi: 10.1016/j.bbi.2020.04.020] [Medline: 32283286]
34. Gramer R, Lynch C. Trump seeks to halve U.S. funding for World Health Organization as coronavirus rages. Foreign Policy. 2020 Feb 10. URL: https://foreignpolicy.com/2020/02/10/trump-world-health-organization-funding-coronavirus-state-department-usaid-budget-cuts/ [accessed 2021-03-16]
35. Gozzi N, Tizzani M, Starnini M, Ciulla F, Paolotti D, Panisson A, et al. Collective response to media coverage of the COVID-19 pandemic on Reddit and Wikipedia: Mixed-methods analysis. J Med Internet Res 2020 Oct 12;22(10):e21597 [FREE Full text] [doi: 10.2196/21597] [Medline: 32960775]

Abbreviations

DOS: degree of separation
kNN: k-nearest neighbor
PHP: hypertext preprocessor
WHO: World Health Organization

Edited by R Kukafka, C Basch; submitted 07.12.20; peer-reviewed by M Martini, E Marzilli; comments to author 12.01.21; revised version received 21.01.21; accepted 18.02.21; published 12.04.21.

Please cite as:
Chrzanowski J, Sołek J, Fendler W, Jemielniak D
Assessing Public Interest Based on Wikipedia's Most Visited Medical Articles During the SARS-CoV-2 Outbreak: Search Trends Analysis
J Med Internet Res 2021;23(4):e26331
URL: https://www.jmir.org/2021/4/e26331
doi:10.2196/26331
PMID:33667176

Original Paper

COVID-19 Vaccine Hesitancy in Canada: Content Analysis of Tweets Using the Theoretical Domains Framework

Janessa Griffith[1,2,3], BA, MSc; Husayn Marani[4], BHSc, MSc; Helen Monkman[5], BSc, MA, PhD

[1]Women's College Hospital Institute for Health Systems Solutions and Virtual Care, Women's College Hospital, Toronto, ON, Canada

[2]Institute of Medical Science, Faculty of Medicine, University of Toronto, Toronto, ON, Canada

[3]Health Information Management, Faculty of Health Sciences, Douglas College, Vancouver, BC, Canada

[4]Institute of Health Policy, Management and Evaluation, Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada

[5]School of Health Information Science, Faculty of Human and Social Development, University of Victoria, Victoria, BC, Canada

Corresponding Author:
Janessa Griffith, BA, MSc
Women's College Hospital Institute for Health Systems Solutions and Virtual Care
Women's College Hospital
76 Grenville Street
Toronto, ON, M5S 1B2
Canada
Phone: 1 6479650943
Email: janessa.griffith@wchospital.ca

Abstract

Background: With the approval of two COVID-19 vaccines in Canada, many people feel a sense of relief, as hope is on the horizon. However, only about 75% of people in Canada plan to receive one of the vaccines.

Objective: The purpose of this study is to determine the reasons why people in Canada feel hesitant toward receiving a COVID-19 vaccine.

Methods: We screened 3915 tweets from public Twitter profiles in Canada by using the search words "vaccine" and "COVID." The tweets that met the inclusion criteria (ie, those about COVID-19 vaccine hesitancy) were coded via content analysis. Codes were then organized into themes and interpreted by using the Theoretical Domains Framework.

Results: Overall, 605 tweets were identified as those about COVID-19 vaccine hesitancy. Vaccine hesitancy stemmed from the following themes: concerns over safety, suspicion about political or economic forces driving the COVID-19 pandemic or vaccine development, a lack of knowledge about the vaccine, antivaccine or confusing messages from authority figures, and a lack of legal liability from vaccine companies. This study also examined mistrust toward the medical industry not due to hesitancy, but due to the legacy of communities marginalized by health care institutions. These themes were categorized into the following five Theoretical Domains Framework constructs: knowledge, beliefs about consequences, environmental context and resources, social influence, and emotion.

Conclusions: With the World Health Organization stating that one of the worst threats to global health is vaccine hesitancy, it is important to have a comprehensive understanding of the reasons behind this reluctance. By using a behavioral science framework, this study adds to the emerging knowledge about vaccine hesitancy in relation to COVID-19 vaccines by analyzing public discourse in tweets in real time. Health care leaders and clinicians may use this knowledge to develop public health interventions that are responsive to the concerns of people who are hesitant to receive vaccines.

(J Med Internet Res 2021;23(4):e26874) doi:10.2196/26874

KEYWORDS

vaccine hesitancy; vaccine; COVID-19; immunization; Twitter; infodemiology; infoveillance; social media; behavioral science; behavior; Canada; content analysis; framework; hesitancy

Introduction

The approval of the Pfizer-BioNTech and Moderna vaccines sent waves of excitement and relief across the world. However, some people remain hesitant about receiving a vaccine for COVID-19 [1,2]. The World Health Organization noted in 2019 that one of the greatest threats to global health was vaccine hesitancy [3]. Emerging international evidence on COVID-19 vaccine hesitancy suggests that there is a range of reasons for this reluctance, including doubts about the safety and efficacy of the vaccine, political or pharmaceutical mistrust, belief in natural immunity, and the belief that the virus is mild or not life-threatening [4-6].

For herd immunity to any communicable disease to be effective, a considerable portion of the population needs to be vaccinated or have antibodies present from being recently infected. Achieving herd immunity is difficult when a large portion of the public is not vaccinated. For herd immunity to be effective for measles and polio, 95% and 80% of the population need to be vaccinated, respectively [7]. The exact percentage required for herd immunity to COVID-19 is difficult to estimate [7].

A Statistics Canada survey conducted in September 2020 (before a vaccine was approved) indicated that 75% of Canadians were either likely or somewhat likely to receive a vaccination [8]. An Angus Reid Institute [4] study conducted between December 8 and 11, 2020 found that 48% of Canadians sampled wanted to be vaccinated immediately if a vaccine was available, and 31% wanted to be vaccinated but preferred to wait. Additionally, 7% of respondents indicated that they were unsure if they would receive a vaccination, and 14% indicated that they would not get vaccinated [4].

In the context of influenza vaccinations, there remains a broad, ethical imperative to respect others' agency over personal health decisions (eg, choosing to not get vaccinated). However, from a public health ethics perspective, the decision to not be vaccinated creates a conflict between population safety and personal liberty [9]. As of yet, COVID-19 vaccination has not been deemed mandatory by any nation, but conversations about whether such a public mandate should exist are emerging [10]. Whether vaccines are mandated, it is worthwhile for public institutions to understand how to change behaviors concerning vaccine hesitancy to ensure that informed decision-making practices are being exercised.

Previous research has suggested that behavioral change interventions are more successful when they are grounded in theory [11]. Thus, we selected a behavioral change framework to guide this study. The Theoretical Domains Framework (TDF) was selected because of its ability to help identify the barriers and facilitators to behavior change while taking into account social and environmental factors [12]. Other public health interventions have used the TDF. For example, Garbutt et al [13] used this framework to improve human papillomavirus vaccine uptake in primary care settings. The use of such theories can facilitate the development of comprehensive health education programs [11], but this requires correctly identifying the attributes of individuals and their surroundings, which influence behavioral patterns [14]. As Bandura [15] and other behavioral theorists have posited, social norms, social relationships, and social networks have a substantial and persistent influence on behaviors [15]. It is worth understanding public discourse about vaccine hesitancy in order to develop interventions that are responsive to the needs of the population and effectively address their concerns.

In the past decade, there has been a particular interest in the utility of Twitter as a tool for monitoring and surveilling public health [16], detecting trends [17], conducting research, and disseminating information [18,19]. A systematic review of using Twitter data for health research found that most studies were in the overlapping fields of public health (23%) and infectious disease (20%) [18]. With 187 million active users worldwide as of January 2021 [20], Twitter has become a powerful social network for disseminating important public health information.

Since the start of the COVID-19 pandemic, social networking platforms like Facebook and YouTube have become stricter with their oversight of the spread of COVID-19 misinformation by deleting false information and providing hyperlinks to government websites containing credible and validated information on COVID-19. Twitter took a similar screening approach in May 2020 [21], yet the scale, spread, and speed of information sharing has made this process challenging. Further, at the start of the pandemic, Twitter introduced a system for verifying COVID-19 experts (indicated with a blue checkmark), including physicians, epidemiologists, scientists, and academics, to provide credible information concerning COVID-19 [22]. Yet, there continues to be influential individuals who have also been verified by Twitter and have enough public credibility to contradict expert opinions or present false information.

We can combat the spread of misinformation by creating targeted approaches to changing behaviors and promoting the understanding of vaccines. Thus, the purpose of this study was to identify the reasons behind vaccine hesitancy among people in Canada by conducting a content analysis of tweets through the lens of behavioral science. Our findings can be used to develop behavior change strategies and policies that are responsive to target populations.

Methods

Study Design

Twitter is a social media platform that allows users to microblog and socially network. Each user is allowed up to 280 characters in a post (called a tweet). Users can post text, pictures, videos, or links to websites. Users who have registered for an account can tweet, like, and comment on another user's tweet and repost tweets (called a retweet). Registered users can also follow accounts and send private messages to each other. Unregistered users can read tweets, retweets, and comments but cannot engage in any interactions [23].

Twitter was selected because of its ability to capture real-time data [19]. Other studies have used Twitter to capture data on vaccine hesitancy. One study compared survey results about vaccine hesitancy in 2018 (before the COVID-19 pandemic) to data captured from Twitter and found that the data were similar to each other [24]. The study argued that Twitter could

potentially be used instead of surveys in some contexts and similar results would be obtained [24]. Another study went as far as saying that Twitter is a "sentinel tool" for identifying public opinions on vaccinations [25]. Thus, Twitter was selected as the site of data collection because it offers a publicly available repository of discourse data (ie, tweets) that are captured in a specific point in time for a specific geographic area.

This study did not require research ethics approval, as it was based on data that were publicly available. Other Canadian-based studies [26] have forgone ethical review by using publicly available Twitter data, as some sources are anonymous or unidentifiable. Only the Twitter user's username (ie, handle), city or town, and tweet content were extracted. This paper only presents aggregated data. Moreover, no interaction occurred between the authors of this study and any of the Twitter users.

Data Collection

After the researcher (JG) was approved for a developer account on Twitter, she received credentials for accessing Twitter's application programming interface. By using a Jupyter environment, the researcher created a Python program to access Twitter's application programming interface. Twitter allows access to tweets up to 1 week after they are posted. Thus, the researcher collected data from two time periods (December 18 and 23, 2020) to access 2 weeks' worth of tweets. Tweets that contained the words "COVID" and "vaccine" were extracted. Similar to a library search, tweets were returned based on variations of these words, such as "COVID-19," "COVID19," "vaccination," and "vaccinate."

Data were extracted from tweets from December 10, 2020, to December 23, 2020. These dates were selected because they followed the Pfizer-BioNTech vaccine approval announcement in Canada (December 9, 2020) and included the dates for the first vaccine administration in Canada (December 14, 2020) and the approval of the Moderna vaccine in Canada (December 23, 2020). This date range also accounted for the time frame when the highest number of searches for terms that included both "COVID" and "vaccine" occurred on Google, which perhaps indicated a spike in interest [27]. Thus, our data reflects a time period when receiving a COVID-19 vaccine was close to becoming a reality. Figure 1 provides a graph that shows when data were extracted and when COVID-19–related events occurred in Canada.

To only include tweets from Canada, the researchers used five geographic radiuses that covered most of Canada. However, several small areas were unintentionally omitted (Figure 2). It was not possible to know how many tweets were missed.

Demographic data beyond users' locations (ie, city or town) were not collected. It was possible to obtain estimates for other demographic information, such as age and gender, from third-party companies. However, this study was operating within the confines of publicly available data so as to disseminate the findings sooner.

Figure 1. A graph depicting Google Trends data for the combined search terms "covid" and "vaccine" aligning with vaccine approval and administration dates in Canada. Tweets that were posted between December 10 and 23, 2020 were eligible for analysis. This date range aligned with the time when the highest peaks in related Google search activity occurred in Canada. This figure indicates that the number of searches on Google for the combined words "COVID" and "vaccine" was highest in December 9, 2020. All other searches were relative to this highest peak. For example, on December 14, 2020, roughly 70% of related searches occurred in December 9, 2020 [28]. It was not possible to obtain more detailed numbers.

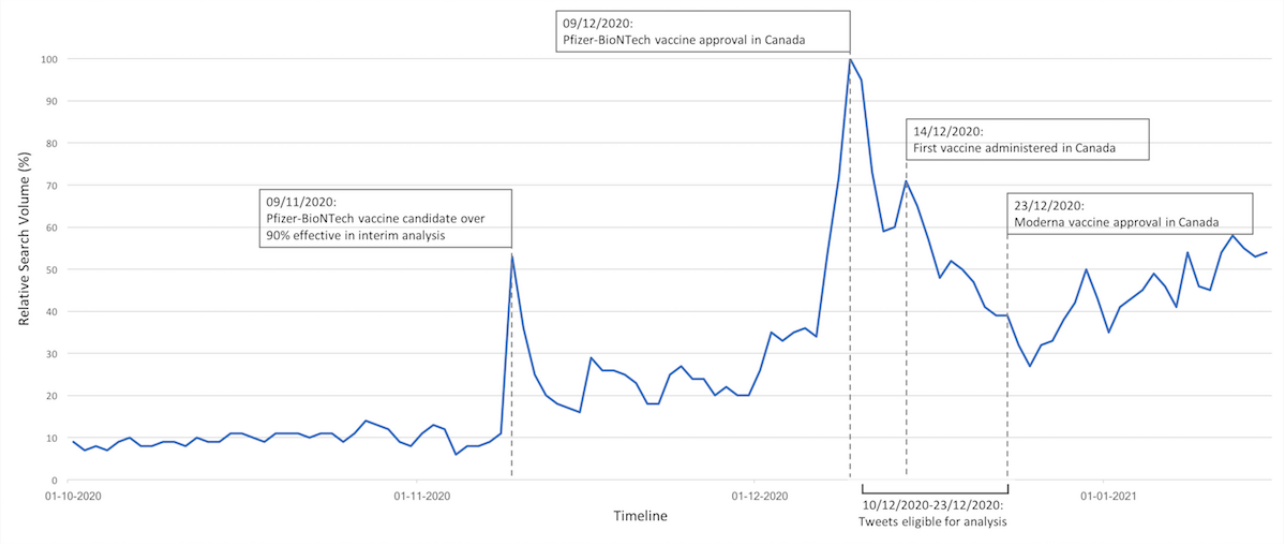
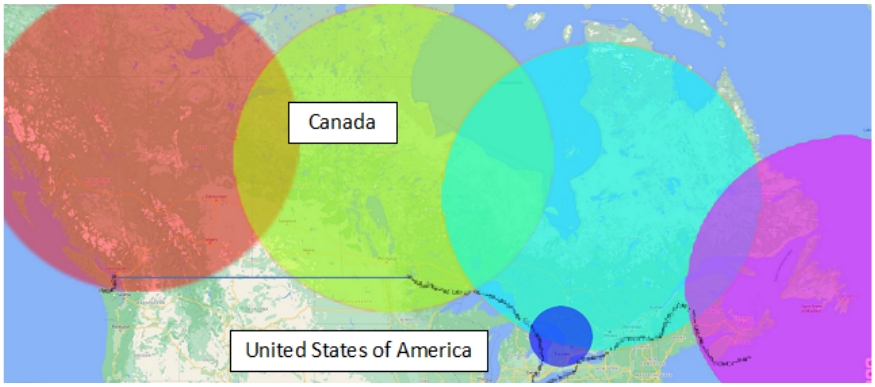


Figure 2. A map of where Twitter users were located. Tweets from outside of Canada (ie, those from the United States) were excluded.



Data Analysis

The results were exported to a comma-separated values file and were analyzed in Microsoft Excel. Tweets were randomized (ie, reordered) in Excel so that tweets were not included based on date. As we expected, the number of tweets extracted was insurmountably high for manual analysis. Therefore, we randomly selected 20% of the tweets to be screened for eligibility. This exceeded the number of randomly selected tweets in other studies, which only included 10% of returned tweets for screening [28]. Double screening was performed for 10% of the tweets to ensure consistency. Manual analysis was selected because this study was exploratory in nature; it was unclear what themes might emerge a priori. As such, training an automated analysis program was unfeasible.

Eligible tweets included any tweets from a Canadian location that contained an expression of hesitancy toward COVID-19 vaccines. These included tweets that provided links to articles or other media that expressed hesitancy toward any COVID-19 vaccine. Eligible tweets also included those with graphics that expressed sentiments of COVID-19 vaccine hesitancy. Tweets that expressed positive or unclear sentiments toward COVID-19 vaccines were excluded. Tweets captured from the United States (given the country's geographic proximity to Canada) were also excluded. As data were extracted on two dates, several duplicate tweets were present. These were identified and deleted in Excel.

All tweets that were deemed eligible after screening were analyzed (ie, qualitatively coded) by 2 authors (JG and HMVM). These researchers had expertise in qualitative coding. Additionally, 10% of the eligible tweets were double-coded to ensure consistency.

In Excel, a content analysis was performed on all eligible tweets. The majority of health studies that use Twitter data (56%) have conducted content analyses [18]. Content analysis was performed as described by Sutton et al [28]; the content of each tweet was systematically reviewed by at least 1 researcher. The researcher(s) then coded the content of tweets according to their meaning. The resulting codes were then organized into thematic categories. Each eligible tweet could be coded into one or more themes.

Once themes emerged from the content analysis, they were mapped onto the TDF. The TDF was selected because it applies a theory-based approach to understanding behavior and has been used extensively in implementation science research. The TDF consists of the following 14 domains: knowledge; skills; social and professional roles and identities; beliefs about capabilities; optimism; beliefs about consequences; reinforcement; intentions; goals; memory, attention, and decision processes; environmental context and resources; social influences; emotion; and behavioral regulation. It has been used in other research pertaining to seasonal flu [29] and human papillomavirus vaccine hesitancy [13] to identify barriers to vaccine uptake and plan for implementation interventions.

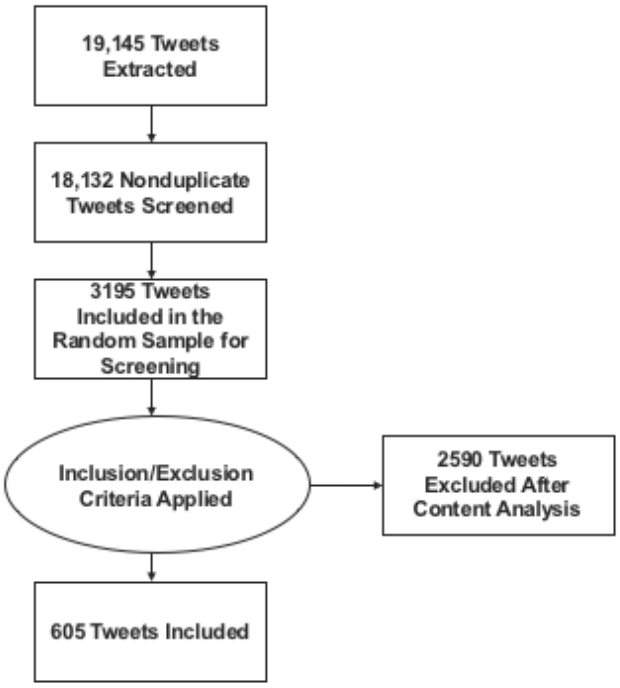
Results

Tweet Characteristics and Themes

In total, 18,132 tweets were returned as search results. Overall, 3915 tweets were screened for eligibility. These tweets represented 21.6% of the total number of tweets. It took approximately 1 hour to manually screen 100 tweets. The 10% (400/3915) of tweets that were double-screened resulted in a Cohen κ coefficient of 0.89, indicating an almost perfect agreement. After screening, 605 tweets met the inclusion criteria. This was represented in a modified PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) diagram (Figure 3).

Through content analysis, the included tweets were grouped into the following major themes concerning vaccine hesitancy: safety, political skepticism, influence from authority figures, a lack of knowledge, and legal liability. The final theme included medical legacies. This theme was different from the other categories of vaccine hesitancy. The themes were not mutually exclusive. Examples of tweets were not provided with the presentation of the themes to preserve the anonymity of Twitter users. In the following subsections, each theme will be described.

Figure 3. Modified PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) diagram of the data extraction process.



Safety

Overall, 48.3% (292/605) of tweets were about safety. These were largely centered around the worry that the vaccine would cause more harm than good. These tweets also expressed concerns that the COVID-19 vaccine was developed more quickly than other vaccines and that the COVID-19 vaccine was not tested to the same rigorous extent as other vaccines. Apprehension over severe side effects was also noted from tweets, including those that reported on nurses fainting and vaccine trial participants experiencing Bell palsy.

Political Skepticism

Another major theme found in 32.4% (196/605) of tweets was skepticism toward the political motivations behind vaccine development. Several Twitter users presented conspiracy theories about the COVID-19 vaccine being a vehicle for exerting political control over citizens. Other participants felt that the vaccine was not tested enough due to political pressures to reopen the economy. Several Twitter users in Canada were also highly influenced by politics in the United States; they cited rumors about the White House threatening the leadership of the US Food and Drug Administration to rush vaccine approval or face forced resignation. Tweets also indicated concern over the influence of big, government-backed pharmaceutical companies ("Big Pharma") that were motivated by profits instead of the desire to help people.

Deficits in Medical and Epidemiologic Literacy Concerning the Benefits of Vaccination

Many tweets (159/605, 26.3%) indicated a lack of knowledge about vaccines among Twitter users. For example, several users expressed the idea that if those who contracted COVID-19 had a ≥99% survival rate, then they should not have to receive a vaccine that is said to be 95% effective. Additionally, Twitter users questioned why anyone else should be concerned if they do not receive the vaccine, indicating a lack of understanding of herd immunity. Twitter users also reported concerns about how the vaccine would alter human DNA. Several Twitter users also felt that a lack of a vaccine for cancer, heart disease, and AIDS was proof that a new virus could not be cured. Additionally, Twitter users viewed COVID-19 as a mild disease; therefore, their interest in undergoing vaccination was low.

Authority Figures

Another theme we found was mistrust toward the COVID-19 vaccine resulting from Canadian and international authority figures not taking the vaccine (51/605, 8.4%). For example, several tweets highlighted users' mistrust toward the CEO of Pfizer and political figureheads in Canadian politics like Doug Ford (the elected provincial leader of Ontario), as they were not taking the vaccine. However, later tweets criticized public figures such as Dr Bonnie Henry (the Provincial Health Officer of British Columbia) and Alexandra Ocasio-Cortez (a member of the US House of Representatives) for receiving the vaccine before frontline workers and older adults.

Legal Liability

To a smaller extent (19/605, 3.1%), Twitter users also expressed mistrust toward vaccines that was based on reports of not being able to take legal action against drug companies if a person experiences any side effects. Additionally, news of the Federal Vaccine Injury Compensation Program in Canada resulted in further skepticism toward vaccine safety.

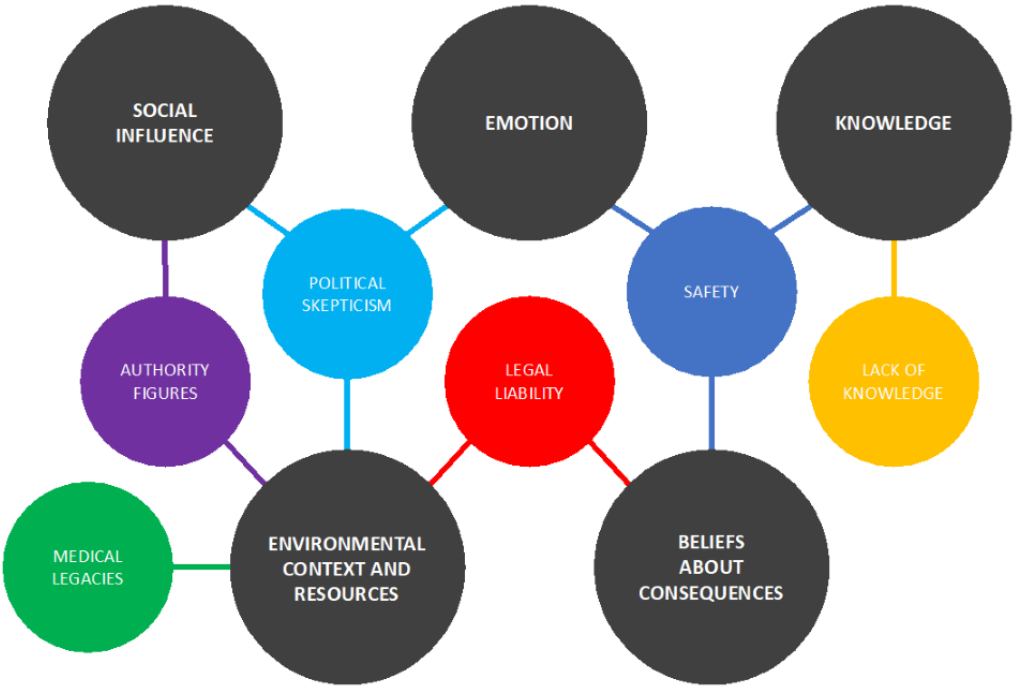
Medical Legacies

The final theme was unlike all of the other themes of vaccine hesitancy in this paper—the legacy of harm caused by health care institutions that have traditionally targeted the Black, Indigenous, and people of color (BIPOC) community and the lesbian, gay, bisexual, transgender, queer+ (LGBTQ+) community. Tweets (24/605, 4%) in this theme highlighted the lack of trust toward the COVID-19 vaccine resulting from how marginalized groups, such as the BIPOC and LGBTQ+ communities, have been historically targeted by the medical community. For example, the Tuskegee syphilis experiments were referenced in several tweets. Moreover, the first people who were vaccinated in the United States were Black health care workers, and several Twitter users viewed this as forced participation in medical experiments. Additionally, a poster promoting COVID-19 vaccination was viewed as paralleling the stigmatization of people who take pre-exposure prophylaxis, a medication for people living with HIV.

Theoretical Domains Framework

Themes were mapped to the TDF and categorized into the following five domains: knowledge, beliefs about consequences, environmental context and resources, social influence, and emotion. The mapping of themes to TDF domains was an interpretive and consensus-driven exercise that was conducted by two study authors (JG and HM). Disagreement was reconciled by a third study author (HMVM). Figure 4 displays a representation of the themes that were mapped to the TDF. We provide insight into this framework in the *Discussion* section. Overall, themes were not mutually exclusive; themes were classified according to several TDF domains.

Figure 4. Themes were categorized based on the TDF. The TDF domains are represented by the dark-gray circles. The themes from the content analysis (smaller colored circles) were mapped to relevant TDF domains. TDF: Theoretical Domains Framework.



Discussion

Principal Results

Through content analysis and TDF application, this study identified the reasons behind vaccine hesitancy among Twitter users in Canada. The major themes that emerged included concerns over safety, suspicion about political or economic forces driving the COVID-19 pandemic or vaccine development, a lack of knowledge about the COVID-19 vaccine, messages from authority figures, and a lack of legal liability from vaccine companies. An additional theme regarding the historical impact of medical mistrust among marginalized communities was also presented. These themes were categorized into the following five TDF constructs: knowledge, beliefs about consequences, environmental context and resources, social influence, and emotion. Thus, efforts to overcome vaccine hesitancy should focus on targeting these constructs.

Although evidence concerning vaccine hesitancy toward the COVID-19 vaccine is still emerging, our findings are consistent with previous studies. A study from Israel found that COVID-19 vaccine hesitancy was related to concerns about safety and efficacy and the belief that the disease is mild [5]. This was similar to our study, wherein concerns about safety was the top reason for vaccine hesitancy. The efficacy of the vaccine and the belief that the virus is mild were grouped into the lack of knowledge theme, which was another top reason for vaccine hesitancy in our study. Another study surveyed individuals from Canada and the United States in May 2020 and reported that vaccine hesitancy correlated with a lack of trust about a vaccine's benefit, concerns about safety (ie, unknown future health consequences), commercial profiteering, and a belief in natural immunity [6]. Of note, these respondents were more likely to receive a vaccine if there was evidence of rigorous testing and safety measures [6]. Both of these studies were

conducted prior to the development and implementation of a COVID-19 vaccine. As such, their results were hypothetical.

This study identified the particular reasons why people in Canada may be hesitant to receive a vaccine, so that implementation scientists who are responsible for vaccine rollouts can become responsive to these concerns. Although the analyzed tweets were from Canada, we believe that the tweets' themes may be generalizable to other contexts. To our knowledge, no other study has analyzed tweets to determine the reasons behind COVID-19 vaccine hesitancy. This study's contribution is especially important because the timing of our study coincided with the approval of the first two vaccines (ie, the Pfizer-BioNTech and Moderna vaccines) and the first vaccine administration in Canada.

Our results relate to vaccine hesitancy in general (ie, past research on non–COVID-19 vaccines), as prior related research has provided similar findings. For example, the influence of the media and people's knowledge about vaccines, past experiences, perceptions of risk, and trust have all been documented [30]. However, hesitancy toward the COVID-19 vaccines presents new, unprecedented challenges; namely, the global COVID-19 pandemic is unlike any pandemic that has been experienced in the past century, herd immunity depends on vaccine participation on a global scale, and new SARS-CoV-2 strains can emerge if the virus has opportunities (ie, time and vectors) to mutate. Additionally, the long-term health consequences of COVID-19 are unknown [31].

Our recommendation for the organizations responsible for implementing vaccination programs is to create behavioral interventions that are responsive to the concerns presented in this study. The mapping of these themes to the TDF provided us with preliminary insights into how to best target these behavioral interventions. For example, safety was a top concern that was found in the tweets, and we mapped safety to both knowledge and beliefs about consequences. Thus, targeting vaccine literacy may be beneficial, and this can be done by explaining how vaccines work, why they are safe, and how no steps were missed in the expedient process of COVID-19 vaccine development. However, trust in politicians and pharmaceutical companies is a vaccine hesitancy factor that is difficult to target because both groups are necessarily involved in vaccine rollouts. One approach to targeting this concern might be to have trusted physicians speak to their patients about why it is important to be immunized. This approach falls under the domain of emotion in the TDF.

Although providing details on interventions for responding to vaccine hesitancy was beyond the scope of this study, Table 1 provides example suggestions for interventions based on each TDF domain.

More research is necessary to determine whether addressing these concerns is effective in overcoming vaccine hesitancy.

Table 1. Reasons for vaccine hesitation fell under several Theoretical Domains Framework (TDF) constructs (left column). The rightmost column provides examples of intervention suggestions for responding to vaccine hesitancy in relation to the TDF construct.

TDF constructs	Content analysis theme	Example suggestions
Knowledge	Lack of knowledge	• Introduce campaigns that educate the public about using clear language in media that are commonly used to digest content (eg, social media).
Social influence	Authority	• Have nonpolitical, respected older adult Canadian celebrities take the vaccine as an example. • Such celebrities could be retired athletes or musicians.
Environmental context and resources	Political skepticism	• Emphasize that vaccines are rooted in science and not politics. This is a difficult quality to understand. • In action, this could be done by having messages come from trusted physicians instead of politicians.
Emotion and beliefs about consequences	Safety	• Highlight examples of instances when the vaccine has worked. • Reiterate the safety of the vaccine. • Reiterate the fact that the steps in the scientific development of the vaccine were not missed.

Limitations

As of January 2021, roughly 6.45 million (~17%) Canadians use Twitter [32]. Therefore, the perspectives on vaccine hesitancy presented in this paper are not wholly representative of the perspectives of all people in Canada. All users included in this study represent people in Canada with broadband internet access, which, as the COVID-19 pandemic has illustrated, is an important determinant of health [33]. As such, we likely missed the perspectives of those who face challenges when accessing the internet. It is also possible that nonhuman Twitter users (bots) were represented in our sample. Previous research has found that Twitter bots have manipulated public opinion and fueled cascades of negative emotions related to topics about COVID-19 [34]. Without any way to systematically identify and exclude these tweets, we suspect that several such tweets were included in our analysis. We also searched for English-only tweets due to limitations in language expertise among this study's authors. A more comprehensive content analysis that is representative of all people in Canada should include tweets that are written in other languages. This limitation may have

resulted in themes not being identified, including those related to culturally specific concerns.

It was not possible to collect demographic data such as age, gender, and ethnicity while also preserving users' anonymity. Thus, we were unable to analyze the demographic characteristics of Twitter users who expressed vaccine hesitancy.

Although the search strategy could have been expanded to include many more terms related to vaccination (eg, "shot," "jab," "immunization," etc), the search results would have been insurmountable for conducting our manual analysis process. Additionally, terms related to hoax beliefs were not included; the inclusion of such terms would have likely produced more results. Although saturation was achieved for our search, we may have missed themes that used alternative language to express vaccine hesitancy.

Of note, the examples of interventions presented in Table 1 are merely suggestions. A behavioral scientist may have more informed suggestions about how to combat vaccine hesitancy according to the TDF.

Finally, the tweets related to the medical legacies discussed in this paper should not be viewed as tweets about vaccine hesitancy or conflated with those under the categories of safety, a lack of knowledge, political skepticism, messages from authority figures, and legal liability. As Mosby and Sridrovich [35] have emphasized, health care providers need to understand the history of "racially segregated health care and medical experimentation." Additionally, Boyd [36] stated that the "hyper-focus on hesitancy implicitly blames Black communities for their undervaccination, and it obscures opportunities to address the primary barrier to COVID-19 vaccination: access." Building trust in the medical system goes far beyond the suggestions presented in this paper.

Conclusions

Overall, this study identified the reasons why people in Canada may feel hesitant toward receiving a COVID-19 vaccine. These reasons fell under the following themes: safety concerns, suspicions about political or economic forces, a lack of knowledge, messages from authority figures, and a lack of legal liability from vaccine companies. Additionally, other tweets revealed the historical impact of medical mistrust among marginalized communities, which should not be viewed as hesitancy or as the result of the reasons identified in this paper. Overall, behavioral, implementation, and public health scientists can use theory-based approaches like the TDF to design interventions that are tailored to address the concerns that people have and improve the uptake of the COVID-19 vaccine, thereby increasing the chances of achieving the threshold necessary for herd immunity.

Acknowledgments

JG conceived the study idea and was involved in all study activities, including study data extraction and analysis and manuscript preparation. HM supported the literature extraction and manuscript preparation. HMVM supported the coding of tweets and manuscript revision. This study was conducted without financial support.

Conflicts of Interest

None declared.

References

1. Detoc M, Bruel S, Frappe P, Tardy B, Botelho-Nevers E, Gagneux-Brunon A. Intention to participate in a COVID-19 vaccine clinical trial and to get vaccinated against COVID-19 in France during the pandemic. Vaccine 2020 Oct 21;38(45):7002-7006 [FREE Full text] [doi: 10.1016/j.vaccine.2020.09.041] [Medline: 32988688]
2. Harrison EA, Wu JW. Vaccine confidence in the time of COVID-19. Eur J Epidemiol 2020 Apr;35(4):325-330 [FREE Full text] [doi: 10.1007/s10654-020-00634-3] [Medline: 32318915]
3. Ten threats to global health in 2019. World Health Organization. URL: https://www.who.int/news-room/spotlight/ten-threats-to-global-health-in-2019 [accessed 2020-12-31]
4. Angus RI. More Canadians willing to roll up their sleeves right away as national COVID-19 vaccine rollout begins. Angus Reid Institute. 2020 Dec 14. URL: https://angusreid.org/canada-COVID-vaccine-december/ [accessed 2020-12-31]
5. Dror AA, Eisenbach N, Taiber S, Morozov NG, Mizrachi M, Zigron A, et al. Vaccine hesitancy: the next challenge in the fight against COVID-19. Eur J Epidemiol 2020 Aug;35(8):775-779. [doi: 10.1007/s10654-020-00671-y] [Medline: 32785815]
6. Taylor S, Landry CA, Paluszek MM, Groenewoud R, Rachor GS, Asmundson GJG. A proactive approach for managing COVID-19: The importance of understanding the motivational roots of vaccination hesitancy for SARS-CoV2. Front Psychol 2020 Oct 19;11:575950 [FREE Full text] [doi: 10.3389/fpsyg.2020.575950] [Medline: 33192883]
7. Coronavirus disease (COVID-19): Herd immunity, lockdowns and COVID-19. World Health Organization. 2020 Dec 30. URL: https://www.who.int/news-room/q-a-detail/herd-immunity-lockdowns-and-covid-19 [accessed 2020-12-31]
8. Majority of Canadians intend to get the COVID-19 vaccine, September 2020. Statistics Canada. URL: https://www150.statcan.gc.ca/n1/daily-quotidien/201217/dq201217c-eng.htm [accessed 2020-12-31]
9. Tilburt JC, Mueller PS, Ottenberg AL, Poland GA, Koenig BA. Facing the challenges of influenza in healthcare settings: the ethical rationale for mandatory seasonal influenza vaccination and its implications for future pandemics. Vaccine 2008 Sep 12;26 Suppl 4:D27-D30. [doi: 10.1016/j.vaccine.2008.07.068] [Medline: 19230155]

10. Gostin LO, Salmon DA, Larson HJ. Mandating COVID-19 vaccines. JAMA 2021 Feb 09;325(6):532-533. [doi: 10.1001/jama.2020.26553] [Medline: 33372955]
11. Glanz K, Bishop DB. The role of behavioral science theory in development and implementation of public health interventions. Annu Rev Public Health 2010;31:399-418. [doi: 10.1146/annurev.publhealth.012809.103604] [Medline: 20070207]
12. Atkins L, Francis J, Islam R, O'Connor D, Patey A, Ivers N, et al. A guide to using the Theoretical Domains Framework of behaviour change to investigate implementation problems. Implement Sci 2017 Jun 21;12(1):77 [FREE Full text] [doi: 10.1186/s13012-017-0605-9] [Medline: 28637486]
13. Garbutt JM, Dodd S, Walling E, Lee AA, Kulka K, Lobb R. Theory-based development of an implementation intervention to increase HPV vaccination in pediatric primary care practices. Implement Sci 2018 Mar 13;13(1):45 [FREE Full text] [doi: 10.1186/s13012-018-0729-6] [Medline: 29534761]
14. Jackson C. Behavioral science theory and principles for practice in health education. Health Educ Res 1997 Mar;12(1):143-150 [FREE Full text] [doi: 10.1093/her/12.1.143]
15. Bandura A. Social Learning Theory. Englewood Cliffs, NJ: Prentice Hall; 1977.
16. Denecke K, Krieck M, Otrusina L, Smrz P, Dolog P, Nejdl W, et al. How to exploit twitter for public health monitoring? Methods Inf Med 2013;52(4):326-339. [doi: 10.3414/ME12-02-0010] [Medline: 23877537]
17. Parker J, Wei Y, Yates A, Frieder O, Goharian N. A framework for detecting public health trends with Twitter. In: Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining. 2013 Presented at: ASONAM '13: Advances in Social Networks Analysis and Mining 2013; August 2013; Niagara, Ontario, Canada p. 556-563. [doi: 10.1145/2492517.2492544]
18. Sinnenberg L, Buttenheim AM, Padrez K, Mancheno C, Ungar L, Merchant RM. Twitter as a tool for health research: A systematic review. Am J Public Health 2017 Jan;107(1):e1-e8. [doi: 10.2105/AJPH.2016.303512] [Medline: 27854532]
19. Scanfeld D, Scanfeld V, Larson EL. Dissemination of health information through social networks: twitter and antibiotics. Am J Infect Control 2010 Apr;38(3):182-188 [FREE Full text] [doi: 10.1016/j.ajic.2009.11.004] [Medline: 20347636]
20. Leading countries based on number of Twitter users as of October 2020. Statista. URL: https://www.statista.com/statistics/242606/number-of-active-twitter-users-in-selected-countries/ [accessed 2020-12-31]
21. Roth Y, Pickles N. Updating our approach to misleading information. Twitter. 2020 May 11. URL: https://blog.twitter.com/en_us/topics/product/2020/updating-our-approach-to-misleading-information.html [accessed 2020-12-31]
22. Lunden I. Twitter prioritizes blue-check verifications to confirm experts on COVID-19 and the novel coronavirus. TechCrunch. 2020 Mar 21. URL: https://techcrunch.com/2020/03/21/twitter-prioritizes-blue-check-verifications-to-confirm-experts-on-covid-19-and-the-novel-coronavirus/ [accessed 2020-12-31]
23. New user FAQ. Twitter. URL: https://help.twitter.com/en/new-user-faq [accessed 2020-12-31]
24. Nowak SA, Chen C, Parker AM, Gidengil CA, Matthews LJ. Comparing covariation among vaccine hesitancy and broader beliefs within Twitter and survey data. PLoS One 2020 Oct 08;15(10):e0239826. [doi: 10.1371/journal.pone.0239826] [Medline: 33031405]
25. Tavoschi L, Quattrone F, D'Andrea E, Ducange P, Vabanesi M, Marcelloni F, et al. Twitter as a sentinel tool to monitor public opinion on vaccination: an opinion mining analysis from September 2016 to August 2017 in Italy. Hum Vaccin Immunother 2020 May 03;16(5):1062-1069 [FREE Full text] [doi: 10.1080/21645515.2020.1714311] [Medline: 32118519]
26. Al-Rawi A, Siddiqi M, Morgan R, Vandan N, Smith J, Wenham C. COVID-19 and the gendered use of emojis on Twitter: Infodemiology study. J Med Internet Res 2020 Nov 05;22(11):e21646 [FREE Full text] [doi: 10.2196/21646] [Medline: 33052871]
27. covid vaccine. Google Trends. URL: https://trends.google.com/trends/explore?date=today%203-m&geo=CA&q=covid%20vaccine [accessed 2020-01-23]
28. Sutton J, Vos SC, Olson MK, Woods C, Cohen E, Gibson CB, et al. Lung cancer messages on Twitter: Content analysis and evaluation. J Am Coll Radiol 2018 Jan;15(1 Pt B):210-217. [doi: 10.1016/j.jacr.2017.09.043] [Medline: 29154103]
29. Kenny E, McNamara Á, Noone C, Byrne M. Barriers to seasonal influenza vaccine uptake among health care workers in long-term care facilities: A cross-sectional analysis. Br J Health Psychol 2020 Sep;25(3):519-539. [doi: 10.1111/bjhp.12419] [Medline: 32320125]
30. Dubé E, Laberge C, Guay M, Bramadat P, Roy R, Bettinger J. Vaccine hesitancy: an overview. Hum Vaccin Immunother 2013 Aug;9(8):1763-1773 [FREE Full text] [doi: 10.4161/hv.24657] [Medline: 23584253]
31. Long-term effects of COVID-19. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/coronavirus/2019-ncov/long-term-effects.html [accessed 2021-01-24]
32. Distribution of Twitter users in Canada as of January 2021, by gender. Statista. URL: https://www.statista.com/statistics/649539/canada-twitter-gender-distribution/ [accessed 2021-04-06]
33. Benda NC, Veinot TC, Sieck CJ, Ancker JS. Broadband Internet Access Is a Social Determinant of Health!. Am J Public Health 2020 Aug;110(8):1123-1125. [doi: 10.2105/AJPH.2020.305784] [Medline: 32639914]
34. Shi W, Liu D, Yang J, Zhang J, Wen S, Su J. Social bots' sentiment engagement in health emergencies: A topic-based analysis of the COVID-19 pandemic discussions on Twitter. Int J Environ Res Public Health 2020 Nov 23;17(22):8701 [FREE Full text] [doi: 10.3390/ijerph17228701] [Medline: 33238567]

35. Mosby I, Swidrovich J. Medical experimentation and the roots of COVID-19 vaccine hesitancy among Indigenous Peoples in Canada. CMAJ 2021 Mar 15;193(11):E381-E383 [FREE Full text] [doi: 10.1503/cmaj.210112] [Medline: 33627413]
36. Boyd R. Black people need better vaccine access, not better vaccine attitudes. The New York Times. 2021 Mar 05. URL: https://www.nytimes.com/2021/03/05/opinion/us-covid-black-people.html [accessed 2021-04-01]

Abbreviations

BIPOC: Black, Indigenous, and people of color
LGBTQ+: lesbian, gay, bisexual, transgender, queer+
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
TDF: Theoretical Domains Framework

Edited by R Kukafka, C Basch; submitted 31.12.20; peer-reviewed by R Marcilly, C Cheng; comments to author 18.01.21; revised version received 25.01.21; accepted 19.02.21; published 13.04.21.

Please cite as:
Griffith J, Marani H, Monkman H
COVID-19 Vaccine Hesitancy in Canada: Content Analysis of Tweets Using the Theoretical Domains Framework
J Med Internet Res 2021;23(4):e26874
URL: https://www.jmir.org/2021/4/e26874
doi:10.2196/26874
PMID:33769946

Original Paper

Spatial-Temporal Relationship Between Population Mobility and COVID-19 Outbreaks in South Carolina: Time Series Forecasting Analysis

Chengbo Zeng[1,2,3], MS; Jiajia Zhang[1,3,4], PhD; Zhenlong Li[1,3,5], PhD; Xiaowen Sun[1,3,4], MS; Bankole Olatosi[1,3,6], PhD; Sharon Weissman[1,3,7], MD; Xiaoming Li[1,2,3], PhD

[1]South Carolina SmartState Center for Healthcare Quality, Arnold School of Public Health, University of South Carolina, Columbia, SC, United States

[2]Department of Health Promotion, Education, and Behavior, Arnold School of Public Health, University of South Carolina, Columbia, SC, United States

[3]Big Data Health Science Center, University of South Carolina, Columbia, SC, United States

[4]Department of Epidemiology and Biostatistics, Arnold School of Public Health, University of South Carolina, Columbia, SC, United States

[5]Geoinformation and Big Data Research Lab, Department of Geography, College of Arts and Sciences, University of South Carolina, Columbia, SC, United States

[6]Department of Health Services, Policy, and Management, Arnold School of Public Health, University of South Carolina, Columbia, SC, United States

[7]School of Medicine, University of South Carolina, Columbia, SC, United States

Corresponding Author:
Chengbo Zeng, MS
South Carolina SmartState Center for Healthcare Quality
Arnold School of Public Health
University of South Carolina
915 Greene Street
Columbia, SC, 29208
United States
Phone: 1 8034775298
Email: czeng@email.sc.edu

Abstract

Background: Population mobility is closely associated with COVID-19 transmission, and it could be used as a proximal indicator to predict future outbreaks, which could inform proactive nonpharmaceutical interventions for disease control. South Carolina is one of the US states that reopened early, following which it experienced a sharp increase in COVID-19 cases.

Objective: The aims of this study are to examine the spatial-temporal relationship between population mobility and COVID-19 outbreaks and use population mobility data to predict daily new cases at both the state and county level in South Carolina.

Methods: This longitudinal study used disease surveillance data and Twitter-based population mobility data from March 6 to November 11, 2020, in South Carolina and its five counties with the largest number of cumulative confirmed COVID-19 cases. Population mobility was assessed based on the number of Twitter users with a travel distance greater than 0.5 miles. A Poisson count time series model was employed for COVID-19 forecasting.

Results: Population mobility was positively associated with state-level daily COVID-19 incidence as well as incidence in the top five counties (ie, Charleston, Greenville, Horry, Spartanburg, and Richland). At the state level, the final model with a time window within the last 7 days had the smallest prediction error, and the prediction accuracy was as high as 98.7%, 90.9%, and 81.6% for the next 3, 7, and 14 days, respectively. Among Charleston, Greenville, Horry, Spartanburg, and Richland counties, the best predictive models were established based on their observations in the last 9, 14, 28, 20, and 9 days, respectively. The 14-day prediction accuracy ranged from 60.3%-74.5%.

Conclusions: Using Twitter-based population mobility data could provide acceptable predictions of COVID-19 daily new cases at both the state and county level in South Carolina. Population mobility measured via social media data could inform proactive measures and resource relocations to curb disease outbreaks and their negative influences.

(J Med Internet Res 2021;23(4):e27045) doi:10.2196/27045

KEYWORDS

COVID-19; mobility; incidence; South Carolina

Introduction

Since the first confirmed case of COVID-19 in the United States on January 21, 2020, countrywide COVID-19 outbreaks have surged. As of March 5, 2021, there were 28,580,198 cumulative confirmed cases and 517,224 COVID-19–related deaths in the United States [1]. South Carolina, a state located in the southeastern United States, had its first confirmed cases on March 6, 2020. From March to May 2020, the trend of daily new cases was flat, with an average daily increase in cases of less than 500. However, the daily new cases in South Carolina have risen sharply since June 2020. On July 14, 2020, COVID-19 cases in South Carolina surpassed 60,000, with more than 2200 daily new cases, the second highest increase in one day in the United States [2]. Between August and October 2020, the transmission rate slowed down with the further implementation of nonpharmaceutical interventions (NPIs), such as dine-in service restrictions and face-covering requirements, but increased steadily after October. By March 5, 2021, there were 448,275 reported cases and 7697 deaths in South Carolina [3].

Given the rapid transmission of COVID-19 and limited options in terms of medical interventions, forecasting is of critical importance as it could predict the spread of disease, estimate the impacts of NPIs, and inform further decision making regarding public health interventions [4]. During the COVID-19 pandemic, decision makers in the United States need to balance the net losses arising from social interruptions, economic damage, and indirect effects on health caused by NPIs with the direct health benefits of disease control [5]. Accurate and reasonable forecasting of COVID-19 could minimize the disease burden in health care settings and the loss of health and life in different phases of reopening plans [5,6].

Existing literature has suggested that population mobility may reflect the influences (both positive and negative) of NPIs, reopening actions, and public holidays [7-9]. For instance, in the early stages of the COVID-19 pandemic, the governor of South Carolina issued a series of NPIs, such as shelter-in-place and the closure of schools and nonessential businesses, to reduce social interaction. These NPIs showed positive effects in suppressing the statewide spread of COVID-19. Later, in May 2020, reopening policies and public holidays diluted the implementation of NPIs, leading to increased social interactions and statewide COVID-19 spread [10,11]. At present, it may be difficult to directly measure the real-time impact of reopening policies, public holidays, and NPI implementation fidelity. Therefore, population mobility may be a proximal indicator allowing for real-time COVID-19 transmission forecasting.

Social media platforms, such as Twitter, collect geospatial information and closely monitor changes in population mobility [12,13]. Indeed, the tremendous volume of user-generated geoinformation from social media enables the real-time or near real-time surveillance of population mobility and provides timely data on how population mobility changes in response to different phases of the COVID-19 outbreak, policy reactions, and public holidays [14-16]. Several studies have leveraged mobility data from social media (eg, Google, Facebook, Twitter) to investigate the relationship between population mobility and COVID-19 transmission [9,11,17-19]. These studies identified a consistently positive relationship between population mobility and COVID-19 incidence. However, few studies used population mobility as a predictor to forecast further outbreaks and to evaluate prediction accuracy in addition to performing correlation analysis. A study by Wang and Yamamoto [19] predicted COVID-19 daily new cases in Arizona using disease surveillance data, the Google Community Mobility report, and partial differential equations. They found an acceptable prediction accuracy for the next 3 days, but the time window of prediction did not cover the duration of viral incubation (ie, 14 days). Furthermore, this study only split Arizona into three regions (ie, central, northern, and southern) rather than examining prediction accuracy at both the state and county level. In fact, there may be geospatial differences in population mobility due to the plausible differential implementation fidelity of NPIs and reactions to reopening policies by county [20,21]. Additionally, there may be geospatial differences in the estimation of population mobility on social media as the number of users and their demographic characteristics may differ by county. All these differences may result in variations in prediction accuracy at the county level, and further studies are needed in this regard.

Prior research has predicted COVID-19 incidence using disease surveillance data and several different time series methods. Most of the studies successfully incorporated the association of the current incidence with the previous incidence using time series methods such as autoregressive, moving average, autoregressive integrated moving average (ARIMA), and Holt-Winters [22]. Some studies used generalized linear regression with continuous outcomes (eg, rate and count), without including time series [23]. However, there were few studies that simultaneously considered time-varying population mobility. Recently, Liboschik and colleagues [24] suggested that count time series following generalized linear models could overcome the limitations of classic time series methods. Based on the generalized linear model methodology, a suitable distribution for count data and appropriate link function could be specified, and the effect of the time-varying covariate could be tested and integrated into forecasting. In this study, we adopted the Poisson count time series model and time-varying population mobility data extracted from Twitter, which may increase the accuracy of COVID-19 prediction.

To address these knowledge gaps, by leveraging disease surveillance data and Twitter-based population mobility, this study aimed to construct Poisson count time series models of COVID-19 daily new cases, investigate the relationship between them, and evaluate the prediction accuracy of daily new cases for the next two-week window at both the state and county level in South Carolina.

Methods

COVID-19 Incidence Data

Cumulative confirmed cases of COVID-19 through November 11, 2020, at both the state and county level in South Carolina were collected from The New York Times data set, which was deposited in GitHub [25]. The data set was compiled using data from state and local governments and health departments, ensuring its accuracy. Within the study period (March 6, 2020 [date of first COVID diagnosis in South Carolina] to November 11, 2020 [251st day]), daily new cases were calculated by subtracting the cumulative confirmed cases of the previous day from the total cases for the entire state and its five counties with the largest numbers of cumulative confirmed cases (ie, Charleston, Greenville, Horry, Spartanburg, and Richland). The study protocol was approved by the Institutional Review Board at the University of South Carolina.

Population Mobility

Population mobility was determined using the number of people (Twitter users) with a moving distance greater than 0.5 miles per day in South Carolina and the selected counties. The methodology of extracting daily population movement (origin-destination flows) from geotagged tweets is discussed elsewhere [26,27]. Briefly, geotagged tweets during the study periods were collected and used for calculation. Only users who posted at least twice per day or posted tweets on at least two consecutive days were included in the calculation. Daily travel distance was calculated for each user based on the derived origin-destination flows and used to generate a variable of how many people moved each day (with a travel distance greater than 0.5 miles). This method of capturing population mobility using Twitter has been previously validated [16,26].

Statistical Analysis

First, daily new cases of COVID-19 and population mobility at both the state and county level were described using line charts in R (version 3.6.3; R Foundation for Statistical Computing; “ggplot” package). Daily new cases and mobility were also described using five quantiles (ie, minimum, 25th percentile, 50th percentile, 75th percentile, and maximum) for each month.

Second, a Poisson count time series model was used to model the impact of population mobility on the daily new cases of COVID-19 at the state level. Time series models were built at various time windows. For the first-round selection, a total of 17 time windows (by 7-day increments) were considered, including 1-7 days, 1-14 days,…, and 1-119 days. The daily new cases from the first to the 234th day were used as the training data set, and those from the next 3 days (days 235-237) were used as a testing data set for the purpose of model evaluation. With the smallest prediction error (equation 1) and good interpretation, the predictive model with the best time window was selected. After the best time window in the first round selection was determined, second- and third-round selections were conducted to narrow down the time window and obtain the final model with the smallest prediction error. The final model was used to predict the COVID-19 daily new cases for the next 3, 7, and 14 days (days 238-251). The cumulative difference (equation 2) between observed and predicted cases and mean absolute percentage accuracy (equation 3) for each time frame were reported [19]. The equations used are as follows:

In equations 1-3, *d* represents the day; *n* is the next 3, 7, or 14 days; *o* is the observed value, *p* is the predicted value, and *x* represents the daily new cases.

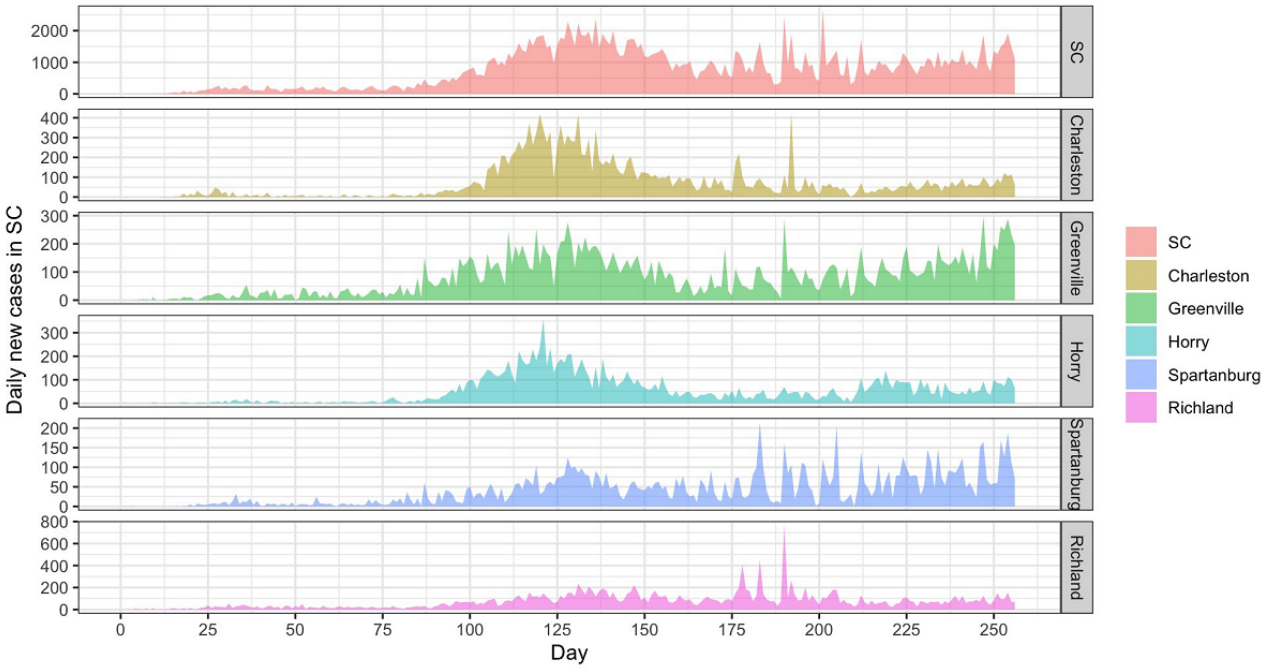
Finally, a similar analytic procedure was performed to construct the final model at the county level for each of the top five counties (ie, Charleston, Greenville, Horry, Spartanburg, and Richland) in South Carolina. A Poisson count time series model was conducted using an R package (“tscount”). Table S1 in Multimedia Appendix 1 provides a detailed description of the data acquisition process, scripts for analysis and figures, and a link to data resources.

Results

Descriptive Statistics

Figure 1 shows the changes in COVID-19 daily new cases at both the state and county level. By October 31, 2020, there were 176,612 cumulative confirmed COVID-19 cases in South Carolina. The cumulative confirmed cases in Charleston, Greenville, Horry, Spartanburg, and Richland were 17,384, 18,021, 12,591, 9290, and 17,531, respectively. At the state level, the daily new cases from March to the end of May were less than 500. From June to the middle of July, the number of daily new cases rose, with 2217 new patients with confirmed COVID-19 on July 14. After that, the transmission rate decreased, with most daily new case counts staying under 1500. However, since October 2020, the daily new cases have steadily increased.

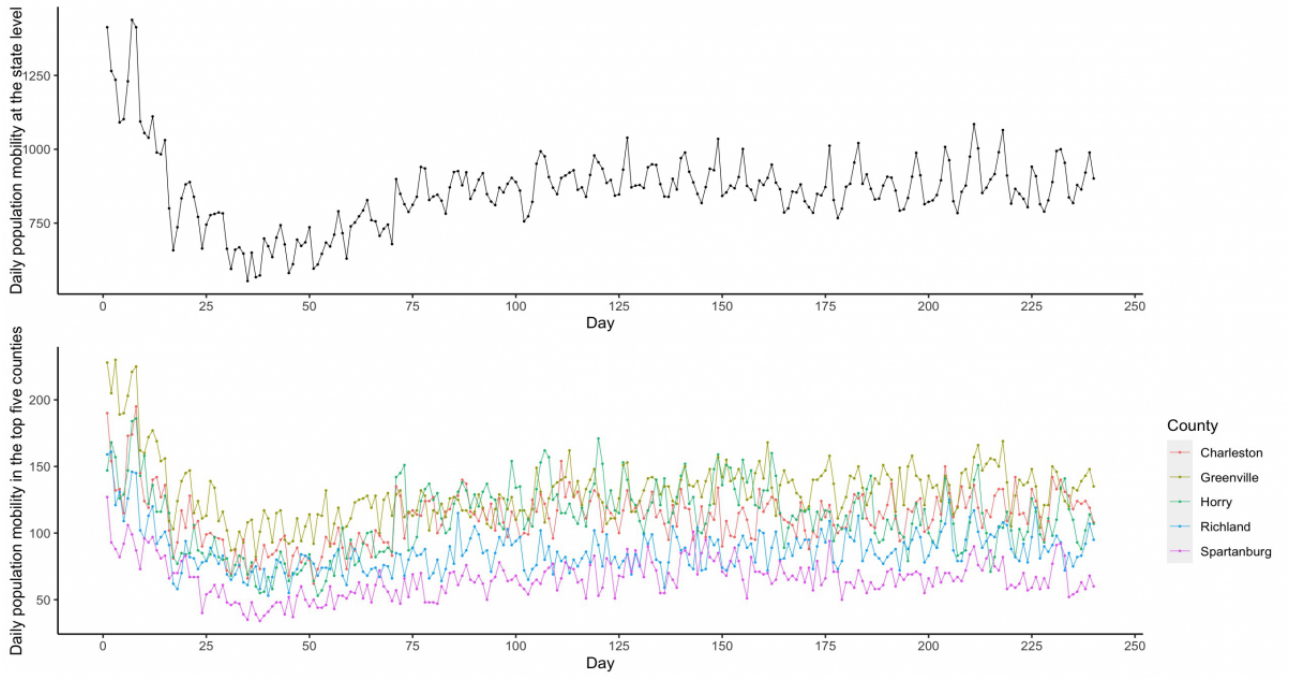
Figure 1. Daily COVID-19 new cases at both state and county level in South Carolina. SC: South Carolina.



At the county level, the top five counties showed a similar trend of COVID-19 outbreaks and accounted for more than 40% of the total cases in South Carolina. The daily new cases increased earlier in Greenville than in the other four counties (ie, Charleston, Horry, Spartanburg, and Richland).

Trends for population mobility at both the state and county level were similar. The number of people in South Carolina (Twitter users in our data) with a moving distance of more than 0.5 miles decreased from 1400 to 550 between March 6 and April 9, 2020. Although there were slight increases from the middle of April to that of June, the numbers were consistently around 1000 after this timeframe. At the county level, each of the five counties had less than 200 people with a moving distance greater than 0.5 miles after the middle of March. Figure 2 shows the changes in population mobility at both the state and county level. Table S2 in Multimedia Appendix 1 presents the descriptive statistics of population mobility and COVID-19 new cases at both the state and county level.

Figure 2. Daily population mobility at both state and county level in South Carolina.



Model Selection of Time Series Analyses

Following the model selection procedure, a Poisson count time series model of COVID-19 incidence at the state level was constructed using daily new cases and population mobility. Population mobility was positively associated with state-level COVID-19 daily new cases (β=.818, 95% CI .761-.876), and the model using the past 7 days (1-7 days) as the time window had the smallest prediction error (Table 1). The prediction error of new cases in the next 3 days (days 235-237) was 0.294.

At the county level, a similar modelling procedure was employed. Population mobility was consistently and positively associated with new cases of COVID-19 across the top five counties. The best time windows for Charleston, Greenville, Horry, Spartanburg, and Richland were 9, 14, 28, 20, and 9 days, respectively. Table 1 displays the detailed results of the final model, the correlation analysis, and the 3-day prediction error at both the state and county level.

Table 1. The impacts of population mobility on COVID-19 outbreaks in South Carolina.

Parameters	State level	County level				
		Charleston	Greenville	Horry	Spartanburg	Richland
Model training						
Time windows (days)	1-7	1-9	1-14	1-28	1-20	1-9
Coefficient of population mobility (95% CI)	0.818 (0.761-0.876)	0.486 (0.338-0.634)	0.278 (0.165-0.390)	0.395 (0.275-0.515)	0.270 (0.118-0.422)	0.157 (0.067-0.246)
Model evaluation (3-day prediction error)	0.294	2.032	0.214	3.146	0.427	0.396
3-day forecasting						
Cumulative difference	42	30	28	40	66	81
Accuracy (%)	98.7	85.1	93.3	69.0	76	72.2
7-day forecasting						
Cumulative difference	670	110	147	45	175	144
Accuracy (%)	90.9	76.7	85.2	85.9	68.3	76.8
14-day forecasting						
Cumulative difference	2858	272	541	217	452	329
Accuracy (%)	81.6	72.1	74.5	72.6	60.3	73.6

COVID-19 Daily New Cases Forecasting

Table 1 also presents the results of forecasting and prediction accuracy. Using the final models with the selected time windows, COVID-19 daily new cases were forecasted for the next 14 days at both the state and county level. At the state level, the 3-day cumulative difference and prediction accuracies were 42 and 98.7%, respectively. As compared to the 3-day prediction accuracy, the 7- and 14-day accuracies reduced to 90.9% and 81.6%, respectively. At the county level, among the top five counties, the 3-day prediction accuracy ranged from 69.0%-93.3%. The prediction accuracy deceased in Charleston, Greenville, and Spartanburg with increased time span. In contrast, the prediction accuracy in Horry and Richland increased in the 7-day prediction but decreased in the 14-day prediction. The 14-day prediction accuracies for Horry and Richland were closer to their values in the 3-day prediction. Table S2 in Multimedia Appendix 1 presents the predicted and observed cases of COVID-19 in the final models.

Discussion

Principal Findings

This study leveraged disease surveillance data and Twitter-based population mobility data to test the relationship between mobility and COVID-19 daily new cases and forecast transmission during the next 14 days at both the state and county level in South Carolina. Results revealed that population mobility was significantly and positively associated with new daily COVID-19 cases. Using the selected models to forecast COVID-19 transmission, we found that although the prediction accuracy at the state level and most of the selected counties decreased as the time span increased, the prediction accuracy remained acceptable. To the best of our knowledge, this is the first study that combined correlation analysis and forecasting together to investigate the impacts of population mobility on COVID-19 transmission at both the state and county level.

Population mobility could reflect the impacts of NPIs, reopening policies, and public holidays, and estimate social movement during the current COVID-19 pandemic. It is closely related to COVID-19 outbreaks, which is in accordance with the findings of prior research [9,11,17-19]. This study adds value to previous studies by examining the impacts of population mobility on COVID-19 incidence at both the state and county level in South Carolina. The results revealed a positive association of population mobility with daily new COVID-19 cases. However, it should be noted that the population mobility data used in our study only reflected the mobility of people who used Twitter, although such mobility data have been validated to be a good proxy of actual human movement during the pandemic

[16,26,27]. Additionally, those Twitter users tended to be young, which might influence how much and what they tweet. The sociodemographic characteristics of Twitter users may be potential confounders, which were not controlled for in our study. Thus, caution is needed when interpreting our findings. Future studies are needed to consider and control for the sociodemographic characteristics of Twitter users.

Using Twitter-based mobility data to predict daily new COVID-19 cases could yield acceptable accuracy, which could also justify the prediction efficacy of this indicator. The high prediction accuracy at the state level was consistent with Wang's finding in Arizona [19]. However, such a high prediction accuracy was not found at the county level. One possible explanation for this finding is that we did not capture or account for the influences of contextual factors (ie, population density) and the roles of mitigating factors (eg, wearing a face mask, practicing social distancing) [18,19,28]. Additionally, the Twitter-based mobility data did not differentiate between social movement at different locations, such as parks, workplaces, and retail locations, which have different impacts on COVID-19 incidence [9]. Finally, in this study, we only captured population mobility at the state and county level, while population mobility at the zip code level might provide a more accurate prediction. Nevertheless, the findings generated from our study confirmed the spatial-temporal relationship between Twitter-based mobility and COVID-19 outbreaks in South Carolina and the acceptable prediction efficacy of population mobility.

Our findings provide empirical evidence to support the application of Poisson count time series and time-varying population mobility data in improving the accuracy of COVID-19 forecasting. Compared with the existing literature, our models yielded acceptable prediction accuracy for two-week forecasting at both the state and county level. Time-varying population mobility could be incorporated into other forecasting models, such as classic time series methods and machine learning [22,24]. Since we are particularly interested in count data, we preferred the Poisson count time series model. When modelling rate, ARIMA and Holt-Winters are more appropriate than the Poisson count time series model. Regarding machine learning, most models are applied to the prediction of binary or categorical variables, and future studies are needed to apply them to predicting the count outcome with time-varying population mobility.

The use of population mobility data has potential implications for future research and practices to curb COVID-19 outbreaks. From a research perspective, research on mobility and COVID-19 could be studied at the state, county, and/or zip code level. In addition, mobility around different locations could provide detailed information regarding COVID-19 transmission, identify the most relevant mobility associated with daily new cases, and inform tailored interventions on social distancing by location to control disease outbreaks. Furthermore, the geospatial difference in the prediction accuracy of population mobility for daily new cases by county suggested that contextual factors—such as demographic characteristics and the implementation fidelity of NPIs at the county level—should be accounted for in future research. Finally, since the incubation and transmission of COVID-19 are closely associated with time-varying factors, such as temperature and weather, such impacts should be accounted for in forecasting studies [29]. Regarding the practice of disease control and prevention, leveraging social media platforms to monitor daily population mobility could improve predictions of further COVID-19 transmission, inform proactive NPIs, and guide the allocation of health care resources to reduce disease morbidity and mortality [30,31].

Conclusions

Population mobility was positively associated with COVID-19 transmission at both the state and county level in South Carolina. Using Twitter-based mobility data could enable acceptable predictions of COVID-19 daily new cases. The use of social media data to monitor population mobility and predict COVID-19 spread could inform proactive measures to curb disease outbreaks and plan coordinated responses.

Acknowledgments

This study was supported by the National Institute of Health (NIH) Research Grant R01AI127203-04S1 by National Institute of Allergy and Infectious Diseases and National Science Foundation (NSF) grant number 2028791.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary data.
[DOCX File , 38 KB - jmir_v23i4e27045_app1.docx]

References

1. Centers for Disease Control and Prevention. Coronavirus Disease 2019 (COVID-19): Cases & Deaths in the US. 2021. URL: https://www.cdc.gov/coronavirus/2019-ncov/index.html [accessed 2021-04-06]
2. Bohatch E. SC sees more than 2,200 coronavirus cases, second highest increase in one day. The State. 2020 Jul 14. URL: https://www.thestate.com/news/coronavirus/article244215817.html [accessed 2021-04-06]

3. South Carolina Department of Health and Environmental Control. SC Demographic Data (COVID-19). 2021. URL: https://scdhec.gov/covid19#coviddata [accessed 2021-04-06]
4. National Center for Immunization and Respiratory Diseases (NCIRD), Division of Viral Diseases. COVID-19 Forecasting: Background Information. 2020 Oct 06. URL: https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/forecasting.html [accessed 2021-04-06]
5. IHME COVID-19 Forecasting Team. Modeling COVID-19 scenarios for the United States. Nat Med 2021 Jan 23;27(1):94-105 [FREE Full text] [doi: 10.1038/s41591-020-1132-9] [Medline: 33097835]
6. National Center for Immunization and Respiratory Diseases (NCIRD), Division of Viral Diseases. Estimated Disease Burden of COVID-19. 2021 Jan 19. URL: https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/burden.html [accessed 2021-04-06]
7. Drake TM, Docherty AB, Weiser TG, Yule S, Sheikh A, Harrison EM. The effects of physical distancing on population mobility during the COVID-19 pandemic in the UK. The Lancet Digital Health 2020 Aug;2(8):e385-e387. [doi: 10.1016/s2589-7500(20)30134-5]
8. Wang S, Liu Y, Hu T. Examining the Change of Human Mobility Adherent to Social Restriction Policies and Its Effect on COVID-19 Cases in Australia. Int J Environ Res Public Health 2020 Oct 29;17(21):7930 [FREE Full text] [doi: 10.3390/ijerph17217930] [Medline: 33137958]
9. Sulyok M, Walker M. Community movement and COVID-19: a global study using Google's Community Mobility Reports. Epidemiol Infect 2020 Nov 13;148:e284. [doi: 10.1017/s0950268820002757]
10. Gibbs H, Liu Y, Pearson CAB, Jarvis CI, Grundy C, Quilty BJ, LSHTM CMMID COVID-19 working group, et al. Changing travel patterns in China during the early stages of the COVID-19 pandemic. Nat Commun 2020 Oct 06;11(1):5012 [FREE Full text] [doi: 10.1038/s41467-020-18783-0] [Medline: 33024096]
11. Xiong C, Hu S, Yang M, Luo W, Zhang L. Mobile device data reveal the dynamics in a positive relationship between human mobility and COVID-19 infections. Proc Natl Acad Sci U S A 2020 Nov 03;117(44):27087-27089 [FREE Full text] [doi: 10.1073/pnas.2010836117] [Medline: 33060300]
12. Hawelka B, Sitko I, Beinat E, Sobolevsky S, Kazakopoulos P, Ratti C. Geo-located Twitter as proxy for global mobility patterns. Cartogr Geogr Inf Sci 2014 May 27;41(3):260-271 [FREE Full text] [doi: 10.1080/15230406.2014.890072] [Medline: 27019645]
13. Martin Y, Li Z, Ge Y. Towards real-time population estimates: introducing Twitter daily estimates of residents and non-residents at the county level. arXiv. Preprint posted online on November 26, 2020. [FREE Full text]
14. Yang C, Sha D, Liu Q, Li Y, Lan H, Guan WW, et al. Taking the pulse of COVID-19: a spatiotemporal perspective. International Journal of Digital Earth 2020 Aug 25;13(10):1186-1211. [doi: 10.1080/17538947.2020.1809723]
15. Li Z, Li X, Porter D, Zhang J, Jiang Y, Olatosi B, et al. Monitoring the Spatial Spread of COVID-19 and Effectiveness of Control Measures Through Human Movement Data: Proposal for a Predictive Model Using Big Data Analytics. JMIR Res Protoc 2020 Dec 18;9(12):e24432 [FREE Full text] [doi: 10.2196/24432] [Medline: 33301418]
16. Huang X, Li Z, Jiang Y, Ye X, Deng C, Zhang J, et al. The characteristics of multi-source mobility datasets and how they reveal the luxury nature of social distancing in the U.S. during the COVID-19 pandemic. International Journal of Digital Earth 2021 Feb 17:1-19. [doi: 10.1080/17538947.2021.1886358]
17. Bisanzio D, Kraemer MU, Bogoch II, Brewer T, Brownstein JS, Reithinger R. Use of Twitter social media activity as a proxy for human mobility to predict the spatiotemporal spread of COVID-19 at global scale. Geospat Health 2020 Jun 15;15(1):2020 [FREE Full text] [doi: 10.4081/gh.2020.882] [Medline: 32575957]
18. Gatalo O, Tseng K, Hamilton A, Lin G, Klein E, CDC MInD-Healthcare Program. Associations between phone mobility data and COVID-19 cases. Lancet Infect Dis 2020 Sep 15:1 [FREE Full text] [doi: 10.1016/S1473-3099(20)30725-8] [Medline: 32946835]
19. Wang H, Yamamoto N. Using a partial differential equation with Google Mobility data to predict COVID-19 in Arizona. Math Biosci Eng 2020 Jul 13;17(5):4891-4904 [FREE Full text] [doi: 10.3934/mbe.2020266] [Medline: 33120533]
20. Kishore N, Kiang MV, Engø-Monsen K, Vembar N, Schroeder A, Balsari S, et al. Measuring mobility to monitor travel and physical distancing interventions: a common framework for mobile phone data analysis. The Lancet Digital Health 2020 Nov;2(11):e622-e628. [doi: 10.1016/s2589-7500(20)30193-x]
21. Chow TE, Choi Y, Yang M, Mills D, Yue R. Geographic pattern of human mobility and COVID-19 before and after Hubei lockdown. Annals of GIS 2020 Nov 11:1-12. [doi: 10.1080/19475683.2020.1841828]
22. Lynch C, Gore R. Short-Range Forecasting of COVID-19 During Early Onset at County, Health District, and State Geographic Levels Using Seven Methods: Comparative Forecasting Study. J Med Internet Res 2021 Mar 23;23(3):e24925 [FREE Full text] [doi: 10.2196/24925] [Medline: 33621186]
23. Chan S, Chu J, Zhang Y, Nadarajah S. Count regression models for COVID-19. Physica A 2021 Feb 01;563:125460 [FREE Full text] [doi: 10.1016/j.physa.2020.125460] [Medline: 33162665]
24. Liboschik T, Fokianos K, Fried R. tscount: An R Package for Analysis of Count Time Series Following Generalized Linear Models. J Stat Soft 2017;82(5):a. [doi: 10.18637/jss.v082.i05]
25. The New York Times. Coronavirus (Covid-19) Data in the United States. 2020. URL: https://github.com/nytimes/covid-19-data [accessed 2021-04-06]

26. Huang X, Li Z, Jiang Y, Li X, Porter D. Twitter reveals human mobility dynamics during the COVID-19 pandemic. PLoS One 2020 Nov 10;15(11):e0241957 [FREE Full text] [doi: 10.1371/journal.pone.0241957] [Medline: 33170889]
27. Li Z, Huang X, Ye X, Li X. ODT Flow Explorer: Extract, Query, and Visualize Human Mobility. arXiv. Preprint posted online on November 26, 2020.
28. Badr HS, Du H, Marshall M, Dong E, Squire MM, Gardner LM. Association between mobility patterns and COVID-19 transmission in the USA: a mathematical modelling study. The Lancet Infectious Diseases 2020 Nov;20(11):1247-1254. [doi: 10.1016/s1473-3099(20)30553-3]
29. Carlson CJ, Gomez ACR, Bansal S, Ryan SJ. Misconceptions about weather and seasonality must not misguide COVID-19 response. Nat Commun 2020 Aug 27;11(1):4312 [FREE Full text] [doi: 10.1038/s41467-020-18150-z] [Medline: 32855406]
30. Kurian SJ, Bhatti AUR, Ting HH, Storlie C, Shah N, Bydon M. Utilization of Mobility Data in the Fight Against COVID-19. Mayo Clin Proc Innov Qual Outcomes 2020 Dec;4(6):733-735 [FREE Full text] [doi: 10.1016/j.mayocpiqo.2020.10.003] [Medline: 33134851]
31. Buckee CO, Balsari S, Chan J, Crosas M, Dominici F, Gasser U, et al. Aggregated mobility data could help fight COVID-19. Science 2020 Apr 10;368(6487):145-146. [doi: 10.1126/science.abb8021] [Medline: 32205458]

Abbreviations

ARIMA: autoregressive integrated moving average
NPI: nonpharmaceutical intervention

Edited by C Basch; submitted 08.01.21; peer-reviewed by R Gore, L Jordan; comments to author 09.02.21; revised version received 08.03.21; accepted 23.03.21; published 13.04.21.

Please cite as:
Zeng C, Zhang J, Li Z, Sun X, Olatosi B, Weissman S, Li X
Spatial-Temporal Relationship Between Population Mobility and COVID-19 Outbreaks in South Carolina: Time Series Forecasting Analysis
J Med Internet Res 2021;23(4):e27045
URL: https://www.jmir.org/2021/4/e27045
doi:10.2196/27045
PMID:33784239

Original Paper

Machine Learning Applied to Clinical Laboratory Data in Spain for COVID-19 Outcome Prediction: Model Development and Validation

Juan L Domínguez-Olmedo[1], PhD; Álvaro Gragera-Martínez[2], MSc, MBA, PharmD; Jacinto Mata[1], PhD; Victoria Pachón Álvarez[1], PhD

[1]Higher Technical School of Engineering, University of Huelva, Huelva, Spain

[2]Juan Ramón Jiménez University Hospital, Huelva, Spain

Corresponding Author:
Juan L Domínguez-Olmedo, PhD
Higher Technical School of Engineering
University of Huelva
Fuerzas Armadas Ave
Huelva, 21007
Spain
Phone: 34 959217371
Email: juan.dominguez@dti.uhu.es

Abstract

Background: The COVID-19 pandemic is probably the greatest health catastrophe of the modern era. Spain's health care system has been exposed to uncontrollable numbers of patients over a short period, causing the system to collapse. Given that diagnosis is not immediate, and there is no effective treatment for COVID-19, other tools have had to be developed to identify patients at the risk of severe disease complications and thus optimize material and human resources in health care. There are no tools to identify patients who have a worse prognosis than others.

Objective: This study aimed to process a sample of electronic health records of patients with COVID-19 in order to develop a machine learning model to predict the severity of infection and mortality from among clinical laboratory parameters. Early patient classification can help optimize material and human resources, and analysis of the most important features of the model could provide more detailed insights into the disease.

Methods: After an initial performance evaluation based on a comparison with several other well-known methods, the extreme gradient boosting algorithm was selected as the predictive method for this study. In addition, Shapley Additive Explanations was used to analyze the importance of the features of the resulting model.

Results: After data preprocessing, 1823 confirmed patients with COVID-19 and 32 predictor features were selected. On bootstrap validation, the extreme gradient boosting classifier yielded a value of 0.97 (95% CI 0.96-0.98) for the area under the receiver operator characteristic curve, 0.86 (95% CI 0.80-0.91) for the area under the precision-recall curve, 0.94 (95% CI 0.92-0.95) for accuracy, 0.77 (95% CI 0.72-0.83) for the F-score, 0.93 (95% CI 0.89-0.98) for sensitivity, and 0.91 (95% CI 0.86-0.96) for specificity. The 4 most relevant features for model prediction were lactate dehydrogenase activity, C-reactive protein levels, neutrophil counts, and urea levels.

Conclusions: Our predictive model yielded excellent results in the differentiating among patients who died of COVID-19, primarily from among laboratory parameter values. Analysis of the resulting model identified a set of features with the most significant impact on the prediction, thus relating them to a higher risk of mortality.

(J Med Internet Res 2021;23(4):e26211) doi:10.2196/26211

KEYWORDS

COVID-19; electronic health record; machine learning; mortality; prediction

Introduction

The COVID-19 pandemic is one of the most prominent health catastrophes of the modern era. This is not exclusive to the health field, as the far-reaching economic and social consequences of this crisis are still unquantifiable [1]. The disease primarily affects the respiratory system, causing respiratory failure, and in certain patients, results in severe inflammatory syndrome. This is mediated by proinflammatory cytokines and can lead to marked systemic complications, which may be fatal in many cases [2].

The lack of knowledge of this virus led the World Health Organization, together with the US Center for Disease Control and Prevention, to define a profile of high-risk patients; this included factors such as age over 65 years, living in nursing homes, and having at least one of the following health problems: chronic lung disease, severe heart disease, obesity, diabetes, kidney failure, liver disease, or an immunocompromised status. The result has been a highly uneven response to the pandemic, both with respect to treatment and the diagnostic and prognostic criteria for the disease [3].

The exponential increase in COVID-19 cases over a short period, lack of experience and knowledge of the virus, and the shortcomings in health resources and health care personnel (many of whom were infected) have caused hospitals to become saturated, especially intensive care units, which have received a very high number of patients every day, many of whom required long stays. Pressure on the health care system after the first wave of the pandemic led to a search for different resources in order to help understand and accurately predict how each patient would react on interacting with the virus. The availability of tools to enable us to classify at-risk patients is crucial because microbiological diagnostics are slow, the PCR test takes more than 4 hours, and emergency physicians do not usually receive the results for up to 24 hours after collecting the sample. Furthermore, the treatments are based on vital support that is not always effective, and potentially gives rise to a large number of adverse events; furthermore, drug availability is sometimes limited. Developing tools that allow us to classify patients at the risk of complications, such those with a prothrombotic status, or an increase in the number of inflammation parameters in a blood sample, would help alleviate the saturation of the health system, optimize resources, and save time in resolving clinical complications [4].

Hence, we developed a model to predict the mortality risk from the laboratory parameters obtained during patients' hospital stay [5]. With this model, we aimed to evaluate how laboratory parameters are related to the risk of a more (or less) severe disease, so that when a patient presents at the emergency department at a hospital, the mortality risk can be predicted on the basis of the blood parameters.

Methods

Data Description

This study is based on anonymized clinical data obtained from a private hospital group in Spain (HM Hospitales), with centers primarily in the Autonomous Communities of Madrid and Galicia and in Barcelona. This group made its data available to the scientific community for research purposes. Using these electronic case histories, we accessed data on individuals suspected with COVID-19 admitted to their centers between March and June 2020. From all the data tables provided, we selected the following: (1) a main table containing specific data on hospitalization and patients (2547 records) and (2) a laboratory data table with the results of the various tests requested for each patient during hospitalization and those presenting at the emergency department (584,136 records).

In the table, an "Outcome" feature is present, with 5 possible values: "Death," "Home," "Transfer to hospital," "Transfer to sociosanitary center" and "Voluntary discharge." This Outcome feature is the aim of the predictive model developed in this study.

Data Preprocessing

Before developing the model, and as a prior step in any machine learning procedure, the information in the 2 tables was preprocessed as follows:

1. Only those patients with a confirmed diagnosis of COVID-19, and whose "Outcome" feature was either "Home" or "Death," were selected.
2. Data from both tables were combined in accordance with the patient ID. Since the patients can present a variable number of measurements for each laboratory parameter, the mean value was calculated and assigned to each of them.
3. Owing to the large number of missing values, we decided to filter records and features in order to handle data without missing values. Some machine learning algorithms can function by directly using data with missing values, and imputation methods can also be used. In this study, for the sake of uniformity and simplicity, the following procedure was used: first, those features having missing values in >10% of all records were eliminated; thereafter, only those records that had value in all the remaining features were selected.
4. Features such as "Sex" and "Outcome" were properly encoded as binary values. No other preprocessing such as normalization or scalarization was applied to the data.

Machine Learning Techniques

A range of machine learning methods to obtain predictive models have been developed, such as those based on logistic regression, linear discriminant analysis, instance-based learning, artificial neural networks, decision trees, and ensemble learning. This study applied the gradient boosting method to develop a predictive model.

Gradient Boosting

Gradient boosting is a machine learning technique used to resolve regression and classification problems and yields a predictive model through an ensemble of weak prediction models, usually decision trees. As in other boosting methods, it builds the model incrementally by incorporating weak prediction models, but it optimizes an arbitrary differentiable loss function. Finally, the prediction for a new case is obtained

by aggregating the predictions of all the individual decision trees that constitute the model. By combining many trees, nonlinearity and interactions between predictor features are achieved [6].

Extreme gradient boosting (XGBoost) is a relatively new gradient boosting implementation that has achieved excellent results in many classification tasks. It is an open-source software library that provides a gradient boosting framework designed to be highly efficient and flexible [7]. It has also been successfully applied in medicine; for example, for the prediction of diabetes risk [8], hypertension [9], drug responses [10], or kidney injury [11].

Shapley Additive Explanations

A fundamental feature of studies performed with machine learning techniques is the interpretability of the results. In medicine, this feature is essential for health care professionals to draw conclusions and take decisions based on the results obtained from machine learning algorithms. Doshi-Velez and Kim [12] defined interpretability as the "ability to explain or to present in understandable terms to a human." This renders interpretability in machine learning a favorable model characteristic.

Recently, the Shapley Additive Explanations (SHAP) framework has been applied to interpret derived machine learning models [13]. SHAP is based on the game theory [14] and helps evaluate feature contributions toward model prediction, identifying the features that most prominently influence the prediction. SHAP values are associated with each feature's marginal contribution when aggregated to the model. The XGBoost method has an additional advantage when SHAP is used, in that being based on decision trees we can use TreeSHAP, a fast variant of SHAP for tree-based machine learning [15].

Model Training and Evaluation

In order to obtain a mortality predictive model ("Outcome" feature), a gradient boosting model was trained using previously described data. Input features were "Age," "Sex," and each of the laboratory values (mean values) in accordance with the data preprocessing described above. For this, the XGBoost model was developed using the existing implementation for Python.

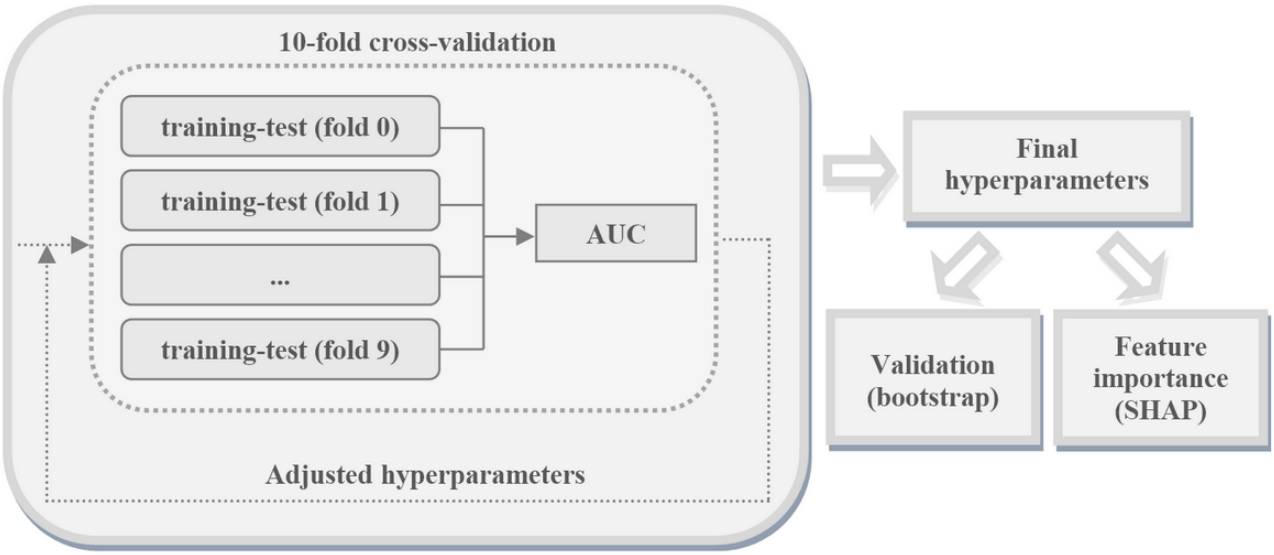
To initially assess the performance of the XGBoost algorithm in relation to other models in the literature, a comparison was made with 8 representative classifiers in machine learning: decision tree, K-nearest neighbors, linear discriminant analysis, logistic regression, multilayer perceptron, Gaussian naive Bayes, random forest, and support vector machines. For this, the corresponding implementation in the Python Scikit-learn library [16] was used. The metrics analyzed were the area under the receiver operator characteristic curve (AUROC), the area under the precision-recall curve (AUPRC), accuracy, and F-score (F1). To assess a value for these metrics, bootstrap validation was used.

For each classifier, the most relevant model parameters (hyperparameters) were adjusted by selecting the best values after an iterative tuning procedure, and leaving the rest with their default values. Hyperparameter values were identified using hyperopt, a Python library for distributed hyperparameter optimization [17]; the metric and the algorithm used in the optimization were AUROC and the 3-structured Parzen estimator. To estimate the AUROC value, k-fold stratified cross-validation (k=10) was employed. Thus, each tuning cycle involved 10 training-test executions using different nonoverlapping test data (each with 10% of the total records). Through cross-validation, the variance of the estimates can be reduced, and the estimation of the generalization performance was improved [18].

Once the results were analyzed, suitable behavior was confirmed in most of the metrics obtained using XGBoost. To further improve its performance, the final model parameters were adjusted using a more exhaustive tuning procedure. Among the variety of parameters available in XGBoost, the ones considered more relevant were selected for tuning. The 6 selected parameters influence the number of gradient boosted trees and their structure (n_estimators, max_depth, and min_child_weight), and the learning process (learning_rate, subsample, and colsample_bytree).

Following this parameter tuning phase, the final model was assessed through bootstrapping. The performance metrics were as follows: AUROC, AUPRC, accuracy, F1, Youden's index, sensitivity, and specificity. Finally, the relative importance of the features in the model was obtained using SHAP (Figure 1).

Figure 1. Procedure for obtaining the model parameters, validation, and feature importance. AUC: area under the curve, SHAP: Shapley Additive Explanations.



Results

Study Population and Features

Following the initial data preprocessing phase, the combination of the data in the 2 tables produced a data set composed of 1823 records and 33 features. All the data correspond to patients with a confirmed diagnosis of COVID-19. Tables 1 and 2 show prevalence and clinical laboratory values, respectively. The median age of all patients was 68 (IQR 57-79) years, and 1114 (61.1%) were male. The "Death" outcome had a prevalence of approximately 14% in the resulting subset of patients after data preprocessing.

Table 1. Prevalence for the "Age," "Sex," and "Outcome" features.

Feature	Patients, n (%)
Age (years)	
0-25	7 (0.4)
25-50	235 (12.9)
50-75	942 (51.7)
75-100	635 (34.8)
100-125	4 (0.2)
Sex	
Male	1114 (61.1)
Female	709 (38.9)
Outcome	
Home	1561 (85.6)
Death	262 (14.4)

Table 2. Clinical laboratory values for the features in the data set.

Feature (units)	Median (IQR[a])	Reference value
Alanine transaminase (U/L)	31.7 (19.2-55.7)	<40
Aspartate transaminase (U/L)	31.8 (22.2-47.3)	<40
Anisocytosis coefficient (%)	13.0 (11.9-14.1)	11.5-14.5
Basophils (%)	0.3 (0.2-0.5)	0-1
Basophil count (10^{-3}/μL)	0.02 (0.01-0.03)	0-0.1
C-reactive protein (mg/L)	52.8 (24.1-94.0)	<5
Creatinine (mg/dL)	0.8 (0.7-1.0)	0.6-1.0
D-Dimer (ng/mL)	885 (492-1883)	<500
Eosinophils (%)	0.8 (0.2-1.6)	2-7
Eosinophil count (10^{-3}/μL)	0.05 (0.01-0.10)	0.1-0.6
Glucose (mg/dL)	110 (97-132)	70-105
Hematocrit (%)	39.5 (36.5-42.5)	40-54
Hemoglobin (g/dL)	13.3 (12.1-14.3)	13.5-17.5
Lactate dehydrogenase (U/L)	507 (402-654)	120-230
Leukocyte count (10^{-3}/μL)	7.0 (5.5-9.2)	4.4-11.3
Lymphocytes (%)	18.4 (12.1-25.5)	20-48
Lymphocyte count (10^{-3}/μL)	1.2 (0.9-1.6)	1.2-3.4
Mean corpuscular hemoglobin (pg)	29.7 (28.6-30.8)	28-33
Mean corpuscular hemoglobin concentration (g/dL)	33.5 (32.7-34.2)	33-36
Mean corpuscular volume (fL)	88.4 (85.5-91.5)	80-95
Mean platelet volume (fL)	10.3 (9.7-11.0)	7.4-10.4
Monocytes (%)	8.1 (6.0-10.5)	1-11
Monocyte count (10^{-3}/μL)	0.6 (0.4-0.7)	0.1-1
Neutrophils (%)	71.0 (62.5-80.1)	40-75
Neutrophil count (10^{-3}/μL)	4.9 (3.6-7.1)	1.5-7.5
Platelet count (10^{-3}/μL)	250 (195-317)	150-450
Potassium (mmol/L)	4.3 (4.0-4.6)	3.5-5.1
Erythrocyte count (10^{-6}/μL)	4.5 (4.1-4.9)	4.1-5.9
Sodium (mmol/L)	138 (136-140)	135-145
Urea (mg/dL)	38 (29-54)	5-50

[a]IQR: Q1-Q3 values.

Model Performance

In the initial evaluation of XGBoost's performance, a comparison with several well-known classifiers was carried out. Table 3 shows the results of this comparison. XGBoost yielded the best results for 3 measures and the second-best results for the F1. These results reaffirm the choice of XGBoost as the predictive method for this study. Figure 2 displays the resulting receiver operator characteristic and precision-recall curves for XGBoost, and Multimedia Appendix 1 shows the corresponding ones for the other methods.

Table 3. Comparison of the outcomes of methods after bootstrap validation.

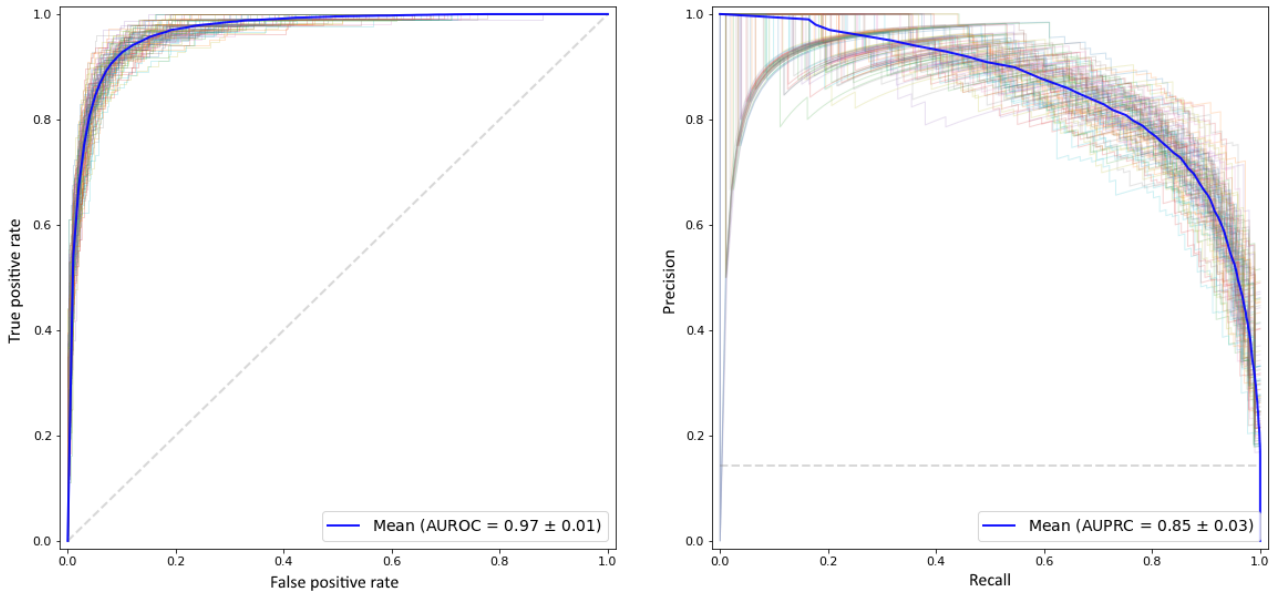
Method	AUROC[a], mean (95% CI)	AUPRC[b], mean (95% CI)	Accuracy, mean (95% CI)	F1[c], mean (95% CI)
Decision tree	0.89 (0.84-0.92)	0.67 (0.58-0.74)	0.89 (0.85-0.92)	0.60 (0.52-0.68)
K-nearest neighbors	0.87 (0.85-0.90)	0.55 (0.46-0.64)	0.88 (0.86-0.90)	0.41 (0.29-0.50)
Linear discriminant analysis	0.96 (0.94-0.97)	0.85 (0.80-0.90)	0.94 (0.92-0.95)	0.75 (0.70-0.82)
Logit	0.96 (0.94-0.98)	0.84 (0.79-0.89)	0.94 (0.92-0.95)	0.76 (0.70-0.82)
Multilayer perceptron	0.95 (0.93-0.97)	0.79 (0.71-0.86)	0.93 (0.91-0.94)	0.73 (0.65-0.79)
Naive Bayes	0.94 (0.91-0.96)	0.74 (0.66-0.82)	0.91 (0.89-0.92)	0.68 (0.62-0.76)
Random forest	0.96 (0.95-0.98)	0.84 (0.76-0.90)	0.93 (0.91-0.95)	0.73 (0.67-0.79)
Support vector machines	0.91 (0.88-0.94)	0.62 (0.53-0.71)	0.87 (0.85-0.88)	0.21 (0.11-0.31)
XGBoost	0.97 (0.96-0.98)	0.85 (0.79-0.91)	0.94 (0.92-0.95)	0.76 (0.71-0.81)

[a]AUROC: area under the receiver operating characteristic curve.

[b]AUPRC: area under the precision-recall curve.

[c]F1: F-score.

Figure 2. The receiver operator characteristic curve (left) and precision-recall curve (right) in the XGBoost model after bootstrap validation. AUROC: area under the receiver operator characteristic curve, AUPRC: area under the precision-recall curve, XGBoost: extreme gradient boosting.



In an attempt to improve XGBoost's performance, the final model parameters were adjusted through a more exhaustive tuning procedure using the Python hyperopt library. Setting the number of iterations (max_eval) at 8000 yielded the hyperparameter values presented in Table 4. The remaining hyperparameters retained their default values. The model used 110 decision trees, with a maximum depth of 3.

Table 4. Final values of the tuned hyperparameters in the extreme gradient boosting model.

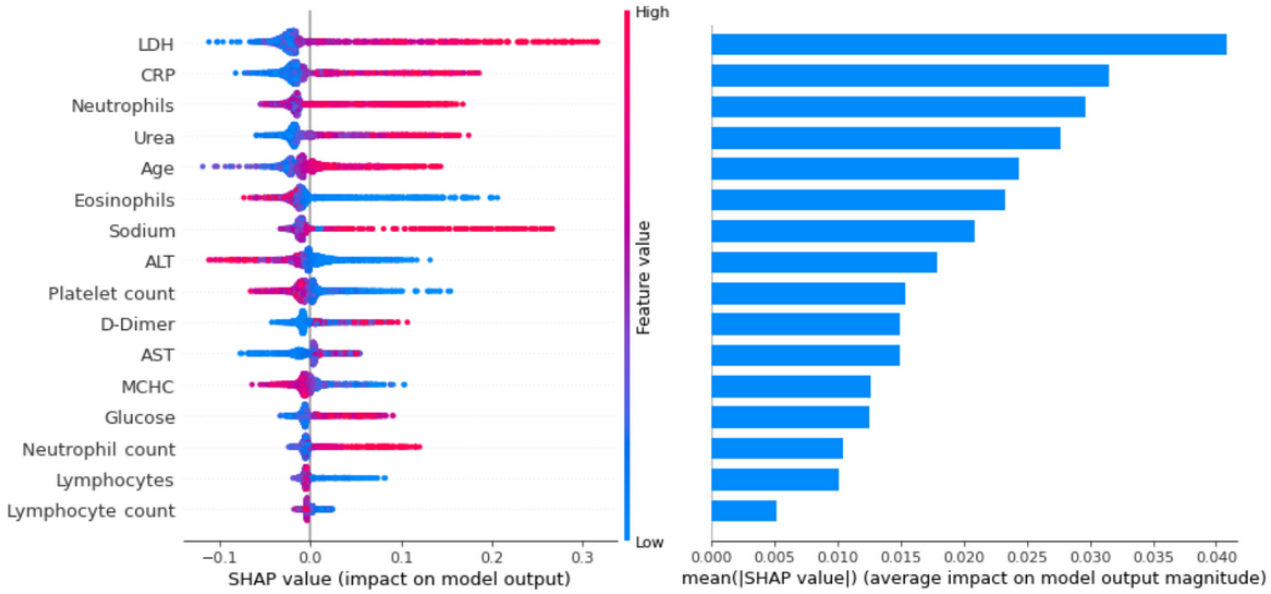
Hyperparameter	Value
Number of gradient-boosted trees	110
Maximum tree depth	3
Minimum sum of instance weight needed in a child	5
Boosting learning rate	0.094
Subsample ratio of the training instances	0.928
Subsample ratio of columns	0.474

With these hyperparameter values, bootstrap validation was again used to obtain model results for the performance metrics, after 300 bootstrap iterations. Through this process, the values obtained were 0.97 (95% CI 0.96- 0.98) for AUROC, 0.86 (95% CI 0.80- 0.91) for AUPRC, 0.94 (95% CI 0.92-0.95) for accuracy, and 0.77 (95% CI 0.72- 0.83) for the F1. We observed a slight improvement owing to more processing in the hyperparameter search process (8000 vs 1000 iterations). Furthermore, the associated sensitivity and specificity values were calculated using the receiver operator characteristic curve values to determine the cut-point that maximizes the Youden index. These calculations yielded a value of 0.85 (95% CI 0.80-0.90) for the Youden index, 0.93 (95% CI 0.89-0.98) for sensitivity, and 0.91 (95% CI 0.86-0.96) for specificity.

Feature Importance

After applying the tuned XGBoost model to the total data set, the SHAP values associated with this model were calculated. Each feature's overall performance can be determined on the basis of these SHAP values in accordance with their average impact on model output. Figure 3 shows SHAP summary plots for the 16 most important features. Based on the mean absolute SHAP values, 5 features, including lactate dehydrogenase, C-reactive protein, neutrophil (%), urea, and age, had a greater average impact on model output. Among these, the feature's highest values (red) are generally associated with a higher SHAP value and, by extension, to a greater likelihood of the "Death" outcome. In other cases, for example eosinophil (%) and alanine aminotransferase, the feature's lowest values (blue) are associated with a greater risk of the "Death" outcome.

Figure 3. SHAP summary plots for the 16 most important features in accordance with their mean absolute values. Beeswarm plot (left), where each dot corresponds to an individual patient, showing the impact of the feature on the model's prediction for that patient. The graph on the right shows the average impact on model output. ALT: alanine transaminase, AST: aspartate transaminase, CRP: C-reactive protein, LDH: lactate dehydrogenase, MCHC: mean corpuscular hemoglobin concentration, SHAP: Shapley Additive Explanations.



Plots developed using SHAP values are displayed in Multimedia Appendix 1; these highlight the relationship between these features and the mortality risk. Every dot represents an individual patient. Furthermore, Multimedia Appendix 1 contains boxplots that describe value distribution between recovered and dead patients for the same features.

Discussion

Principal Findings

COVID-19 mortality is strongly linked to 2 events. There are patients who develop a severe inflammatory syndrome, which results in uncontrolled activation of the immune system and a massive release of proinflammatory cytokines, which translates into an increase in acute-phase reactants such as C-reactive protein, interleukin-6, ferritin, cell destruction markers such as lactate dehydrogenase, and an increase in proinflammatory cells such as neutrophils. This severe inflammatory syndrome has been described as a cause of mortality in most patients with complications arising from a SARS-CoV-2 infection. In such patients, lactate dehydrogenase is associated with an increase in cell destruction, which results in a reduction in lymphocytes, rupture of the lung parenchyma due to inflammation, cell damage, cell remodeling, and lung fibrosis [19,20]. Our study data are concurrent with this trend, with lactate dehydrogenase, C-reactive protein, and neutrophils having the greatest impact on mortality among these patients. Another important complication described in these patients is acute renal failure [21]; our data show that the laboratory parameter that most influences mortality in relation to renal function is urea, a marker of renal function at the prerenal level, which indicates whether renal filtering is effective. Urea levels tend to increase when patients are dehydrated or experience excessive fluid loss [3].

From among clinical laboratory findings, it is essential to establish a biochemical panel of acute-phase reactants that facilitate the identification of patients susceptible to an acute inflammatory syndrome. In this case, we propose lactate dehydrogenase and C-reactive protein as the best candidates

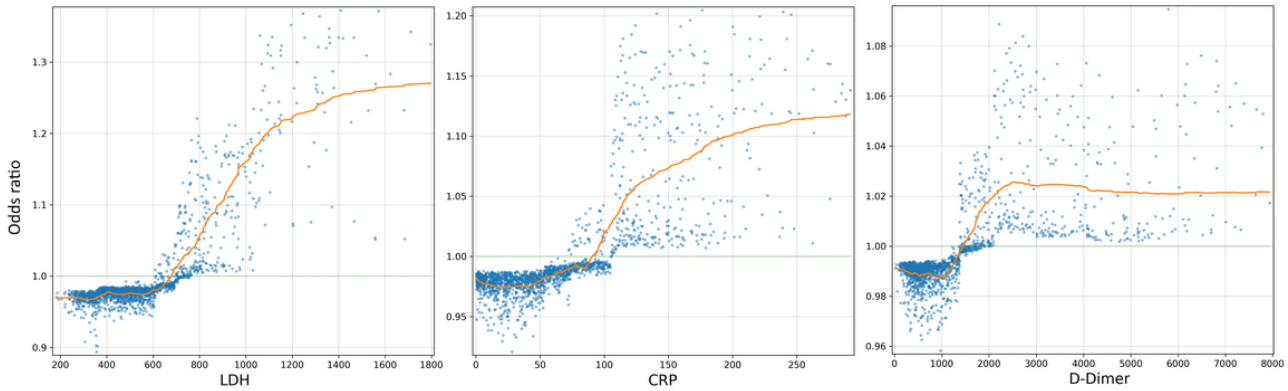
according to the data obtained, to which interleukin-6 ferritin should be added, at the very least.

Another complication that results in high mortality in these patients is coagulation disorders. COVID-19 results in a systemic hypercoagulation state, producing pulmonary thromboembolisms, ischemic strokes, and other disorders, and a markedly large number patients experience severe complications. This complication can be assessed on the basis of 2 laboratory parameters: D-Dimer and platelets. As a degradation product of a previously formed clot, the increase in this parameter will thus be proportional to the number of previously formed clots. In the first step of the coagulation process, a reduction in the number of platelets would indicate that clots are being formed. Accordingly, a risk factor would be an increase in D-dimer levels and a reduction in the platelet count [22].

Approximately 30% of patients with COVID-19 complications have hypercoagulation disorders; hence, it is important to be able to predict these complications in order to establish prophylactic anticoagulant treatment as early as possible in patients in whom this blood disorder is identified. Some studies have compared the hypercoagulation status resulting from COVID-19 to that appearing in patients with an antiphospholipid syndrome, who present with the same complications and in whom the treatment is identical [23]. Of note, we have established a strong relationship between coagulation parameters and mortality in the predictive model we developed in this study.

Figure 4 shows the most interesting parameters—from the clinical point of view—and their relation to mortality. The 3 graphs have a common relationship; that is, from a certain value, the curve that relates the value of the variable to mortality increases significantly. At this point, the medical intervention could change the clinical course of patients since, as seen in the graph, very high values in these tests represent a higher mortality risk, while low levels relate to a more favorable prognosis.

Figure 4. Plots developed using SHAP values, displaying the relationship between laboratory values—including LDH, CRP, and D-Dimer—and mortality risk. Every dot represents an individual patient. Higher values of these features indicate an increase in the mortality risk, and lower ones are associated with a more favorable prognosis. CRP: C-reactive protein, LDH: lactate dehydrogenase, SHAP: Shapley Additive Explanations.



At the beginning of the pandemic, one of the main risk factors by which patients were classified was age; as expected, higher morbidity and mortality rates prevail among older individuals. In our predictive model, age ranks in fifth position, which is important, but mortality is still more prominent among those patients who develop a severe inflammatory syndrome. Therefore, if we relate age as an independent variable to the main biochemical markers of severe inflammation, we can estimate patient mortality on the basis of their age and a clinical laboratory value (Figure 5). On the other hand, Figure 6 shows the difference in different clinical laboratory values from among patients who die or are discharged from hospital. We observed a clear difference between different laboratory values depending on each group.

Figure 5. Partial dependence plots representing the model output associated with age and other features (LDH, CRP, and urea). Red zones indicate a greater influence on mortality risk. CRP: C-reactive protein, LDH: lactate dehydrogenase.

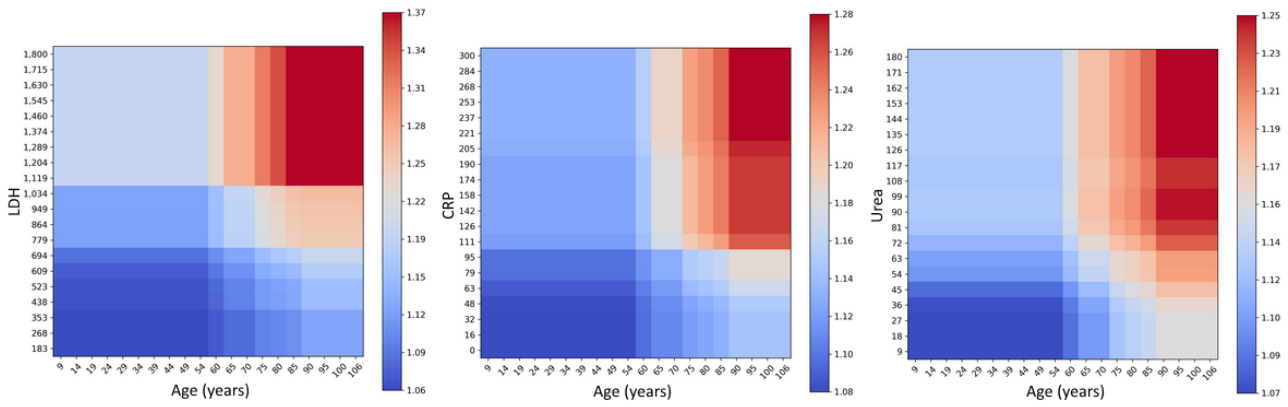
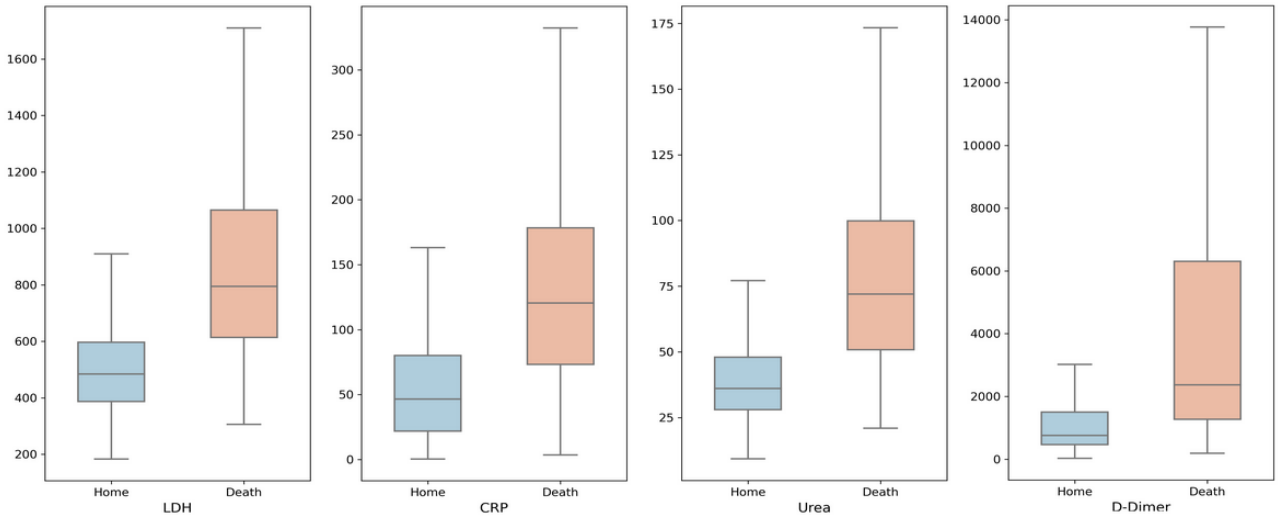


Figure 6. Boxplots describing value distribution between recovered and dead patients on the basis of the laboratory values of LDH, CRP, Urea, and D-Dimer. CRP: C-reactive protein, LDH: lactate dehydrogenase.



It is not easy to establish strict criteria for mortality in patients with COVID-19, as they are influenced by other unquantified variables and environmental factors. The comorbidities in these patients prior to them contracting COVID-19 are very important when managing these patients and predicting complications [24]. Patients with chronic pathologies such as hypertension or diabetes have a higher number of complications and rate of mortality than those who do not; however, the underlying reason remains unclear. It has been hypothesized that these patients have higher expression levels of angiotensin-converting enzyme 2 receptors through which the virus penetrates the cells to replicate; such patients are candidates for a stronger and more severe disease. Furthermore, Fang et al [25] reported that polymorphisms in the gene that encodes this receptor increases the severity of the disease.

Carrasco-Sánchez et al [26] collected data from approximately 20,000 patients in Spain and reported that mortality can be predicted among those patients who arrive at an emergency department and are then found to have high blood glucose levels, during their hospital stay, provided they are not in a critical condition. Blood glucose is thus one of the most prominent predictors of patient mortality, which is concurrent with our hypothesis. Therefore, glycemic control among patients before and during their hospital stay is essential to increase their survival.

Limitations

Clinically, this study has a series of limitations. First, this study has a small patient cohort; previous similar studies have included a markedly larger patient cohort [21,26]. Second, we did not record the comorbidities of these patients; therefore, we cannot assess their role in relation to other variables and their potential to predict a patient's mortality. Finally, it is very important to perform a battery of laboratory tests, which facilitates the evaluation of an inflammatory syndrome with more parameters, such as interleukin-6 and ferritin.

Conclusions

This study aimed to develop a model to predict the mortality of patients with COVID-19, which can assess mortality from laboratory values with a high degree of accuracy. The use of machine learning techniques, in this case the XGBoost predictive method, has yielded excellent results for several performance metrics. The analysis of the resulting model enables us to identify a set of features with a markedly high prediction potential, which can be useful for improving care decisions and increasing patient survival.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary figures.
[DOCX File , 6339 KB - jmir_v23i4e26211_app1.docx]

References

1. Mahase E. Coronavirus covid-19 has killed more people than SARS and MERS combined, despite lower case fatality rate. BMJ 2020 Feb 18;368:m641. [doi: 10.1136/bmj.m641] [Medline: 32071063]
2. Ranney ML, Griffeth V, Jha AK. Critical Supply Shortages - The Need for Ventilators and Personal Protective Equipment during the Covid-19 Pandemic. N Engl J Med 2020 Apr 30;382(18):e41. [doi: 10.1056/NEJMp2006141] [Medline: 32212516]

3. Yan L, Zhang H, Goncalves J, Xiao Y, Wang M, Guo Y, et al. An interpretable mortality prediction model for COVID-19 patients. Nat Mach Intell 2020 May 14;2(5):283-288. [doi: 10.1038/s42256-020-0180-7]
4. Schwab P, DuMont Schütte A, Dietz B, Bauer S. Clinical Predictive Models for COVID-19: Systematic Study. J Med Internet Res 2020 Oct 06;22(10):e21439 [FREE Full text] [doi: 10.2196/21439] [Medline: 32976111]
5. Barda N, Riesel D, Akriv A, Levy J, Finkel U, Yona G, et al. Developing a COVID-19 mortality risk prediction model when individual-level data are not available. Nat Commun 2020 Sep 07;11(1):4439 [FREE Full text] [doi: 10.1038/s41467-020-18297-9] [Medline: 32895375]
6. Friedman JH. Stochastic gradient boosting. Comput Data Stat Anal 2002 Feb;38(4):367-378. [doi: 10.1016/S0167-9473(01)00065-2]
7. Chen T, Guestrin CE. XGBoost: A Scalable Tree Boosting System. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York, NY: Association for Computing Machinery; 2016 Aug Presented at: KDD '16: The 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; August 2016; San Francisco, CA p. 785-794. [doi: 10.1145/2939672.2939785]
8. Wang L, Wang X, Chen A, Jin X, Che H. Prediction of Type 2 Diabetes Risk and Its Effect Evaluation Based on the XGBoost Model. Healthcare (Basel) 2020 Jul 31;8(3):247 [FREE Full text] [doi: 10.3390/healthcare8030247] [Medline: 32751894]
9. Chang W, Liu Y, Xiao Y, Yuan X, Xu X, Zhang S, et al. A Machine-Learning-Based Prediction Method for Hypertension Outcomes Based on Medical Data. Diagnostics (Basel) 2019 Nov 07;9(4):178 [FREE Full text] [doi: 10.3390/diagnostics9040178] [Medline: 31703364]
10. Mo X, Chen X, Li H, Li J, Zeng F, Chen Y, et al. Early and Accurate Prediction of Clinical Response to Methotrexate Treatment in Juvenile Idiopathic Arthritis Using Machine Learning. Front Pharmacol 2019;10:1155 [FREE Full text] [doi: 10.3389/fphar.2019.01155] [Medline: 31649533]
11. Hsu C, Liu C, Tain Y, Kuo C, Lin Y. Machine Learning Model for Risk Prediction of Community-Acquired Acute Kidney Injury Hospitalization From Electronic Health Records: Development and Validation Study. J Med Internet Res 2020 Aug 04;22(8):e16903 [FREE Full text] [doi: 10.2196/16903] [Medline: 32749223]
12. Doshi-Velez F, Kim B. Towards A Rigorous Science of Interpretable Machine Learning. arXiv. Preprint posted online February 28, 2017. [FREE Full text]
13. Lundberg S, Erion G, Chen H, DeGrave A, Prutkin JM, Nair B, et al. From Local Explanations to Global Understanding with Explainable AI for Trees. Nat Mach Intell 2020 Jan;2(1):56-67 [FREE Full text] [doi: 10.1038/s42256-019-0138-9] [Medline: 32607472]
14. Kuhn HW, Tucker AW, editors. Contributions to the Theory of Games (AM-28), Volume II. Princeton, NJ: Princeton University Press; 1953.
15. Lundberg S, Erion G, Lee S. Consistent Individualized Feature Attribution for Tree Ensembles. arXiv. Preprint posted online February 12, 2018.
16. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine Learning in Python. J Mach Learn Res 2011;12:2825-2830 [FREE Full text]
17. Bergstra J, Yamins D, Cox D. Making a science of model search: hyperparameter optimization in hundreds of dimensions for vision architectures. In: Proceedings of the 30th International Conference on International Conference on Machine Learning - Volume 28. 2013 Presented at: The 30th International Conference on International Conference on Machine Learning; June 16-21, 2013; Atlanta, GA p. I-115-I-123.
18. Arlot S, Celisse A. A survey of cross-validation procedures for model selection. Statist Surv 2010 Jan 1;4:40-79. [doi: 10.1214/09-SS054]
19. Drent M, Cobben N, Henderson R, Wouters E, van Dieijen-Visser M. Usefulness of lactate dehydrogenase and its isoenzymes as indicators of lung damage or inflammation. Eur Respir J 1996 Aug;9(8):1736-1742 [FREE Full text] [doi: 10.1183/09031936.96.09081736] [Medline: 8866602]
20. Henry BM, de Oliveira MHS, Benoit S, Plebani M, Lippi G. Hematologic, biochemical and immune biomarker abnormalities associated with severe illness and mortality in coronavirus disease 2019 (COVID-19): a meta-analysis. Clin Chem Lab Med 2020 Jun 25;58(7):1021-1028. [doi: 10.1515/cclm-2020-0369] [Medline: 32286245]
21. Vaid A, Somani S, Russak AJ, De Freitas JK, Chaudhry FF, Paranjpe I, et al. Machine Learning to Predict Mortality and Critical Events in a Cohort of Patients With COVID-19 in New York City: Model Development and Validation. J Med Internet Res 2020 Nov 06;22(11):e24018 [FREE Full text] [doi: 10.2196/24018] [Medline: 33027032]
22. Klok F, Kruip M, van der Meer N, Arbous M, Gommers D, Kant K, et al. Incidence of thrombotic complications in critically ill ICU patients with COVID-19. Thromb Res 2020 Jul;191:145-147 [FREE Full text] [doi: 10.1016/j.thromres.2020.04.013] [Medline: 32291094]
23. Zhang Y, Xiao M, Zhang S, Xia P, Cao W, Jiang W, et al. Coagulopathy and Antiphospholipid Antibodies in Patients with Covid-19. N Engl J Med 2020 Apr 23;382(17):e38 [FREE Full text] [doi: 10.1056/NEJMc2007575] [Medline: 32268022]
24. Kim N, Ha E, Moon JS, Lee Y, Choi EY. Acute Hyperglycemic Crises with Coronavirus Disease-19: Case Reports. Diabetes Metab J 2020 Apr;44(2):349-353 [FREE Full text] [doi: 10.4093/dmj.2020.0091] [Medline: 32347027]

25. Fang L, Karakiulakis G, Roth M. Are patients with hypertension and diabetes mellitus at increased risk for COVID-19 infection? Lancet Respir Med 2020 Apr;8(4):e21 [FREE Full text] [doi: 10.1016/S2213-2600(20)30116-8] [Medline: 32171062]
26. Carrasco-Sánchez FJ, López-Carmona MD, Martínez-Marcos FJ, Pérez-Belmonte LM, Hidalgo-Jiménez A, Buonaiuto V, SEMI-COVID-19 Network. Admission hyperglycaemia as a predictor of mortality in patients hospitalized with COVID-19 regardless of diabetes status: data from the Spanish SEMI-COVID-19 Registry. Ann Med 2021 Dec;53(1):103-116 [FREE Full text] [doi: 10.1080/07853890.2020.1836566] [Medline: 33063540]

Abbreviations

AUPRC: area under the precision-recall curve
AUROC: area under the receiving operator characteristic curve
SHAP: Shapley Additive Explanations
XGBoost: Extreme Gradient Boosting

Edited by C Basch; submitted 07.12.20; peer-reviewed by V Rivas Santos, W Wu; comments to author 23.12.20; revised version received 29.12.20; accepted 08.03.21; published 14.04.21.

Please cite as:
Domínguez-Olmedo JL, Gragera-Martínez Á, Mata J, Pachón Álvarez V
Machine Learning Applied to Clinical Laboratory Data in Spain for COVID-19 Outcome Prediction: Model Development and Validation
J Med Internet Res 2021;23(4):e26211
URL: https://www.jmir.org/2021/4/e26211
doi:10.2196/26211
PMID:33793407

Original Paper

Rise in Use of Digital Mental Health Tools and Technologies in the United States During the COVID-19 Pandemic: Survey Study

Dara H Sorkin[1], PhD; Emily A Janio[1], MPH; Elizabeth V Eikey[2,3], PhD; Margaret Schneider[4], PhD; Katelyn Davis[5], MSc; Stephen M Schueller[6,7], PhD; Nicole A Stadnick[8,9,10], MPH, PhD; Kai Zheng[7], PhD; Martha Neary[6], MSc; David Safani[11], MBA, MD; Dana B Mukamel[1], PhD

[1]Department of Medicine, University of California, Irvine, Irvine, CA, United States
[2]Herbert Wertheim School of Public Health and Human Longevity Science, University of California, San Diego, San Diego, CA, United States
[3]The Design Lab, University of California, San Diego, San Diego, CA, United States
[4]Department of Public Health, University of California, Irvine, Irvine, CA, United States
[5]Department of Pathology, University of California, San Francisco, San Francisco, CA, United States
[6]Department of Psychological Science, University of California, Irvine, Irvine, CA, United States
[7]Department of Informatics, University of California, Irvine, Irvine, CA, United States
[8]Department of Psychiatry, University of California, San Diego, San Diego, CA, United States
[9]UC San Diego Dissemination and Implementation Science Center, San Diego, CA, United States
[10]Child and Adolescent Services Research Center, San Diego, CA, United States
[11]Department of Psychiatry, University of California, Irvine, Irvine, CA, United States

Corresponding Author:
Dara H Sorkin, PhD
Department of Medicine
University of California, Irvine
100 Theory
Suite 120
Irvine, CA, 92697
United States
Phone: 1 949 824 0149
Email: dsorkin@uci.edu

Abstract

Background: Accompanying the rising rates of reported mental distress during the COVID-19 pandemic has been a reported increase in the use of digital technologies to manage health generally, and mental health more specifically.

Objective: The objective of this study was to systematically examine whether there was a COVID-19 pandemic–related increase in the self-reported use of digital mental health tools and other technologies to manage mental health.

Methods: We analyzed results from a survey of 5907 individuals in the United States using Amazon Mechanical Turk (MTurk); the survey was administered during 4 week-long periods in 2020 and survey respondents were from all 50 states and Washington DC. The first set of analyses employed two different logistic regression models to estimate the likelihood of having symptoms indicative of clinical depression and anxiety, respectively, as a function of the rate of COVID-19 cases per 10 people and survey time point. The second set employed seven different logistic regression models to estimate the likelihood of using seven different types of digital mental health tools and other technologies to manage one's mental health, as a function of symptoms indicative of clinical depression and anxiety, rate of COVID-19 cases per 10 people, and survey time point. These models also examined potential interactions between symptoms of clinical depression and anxiety, respectively, and rate of COVID-19 cases. All models controlled for respondent sociodemographic characteristics and state fixed effects.

Results: Higher COVID-19 case rates were associated with a significantly greater likelihood of reporting symptoms of depression (odds ratio [OR] 2.06, 95% CI 1.27-3.35), but not anxiety (OR 1.21, 95% CI 0.77-1.88). Survey time point, a proxy for time, was associated with a greater likelihood of reporting clinically meaningful symptoms of depression and anxiety (OR 1.19, 95% CI 1.12-1.27 and OR 1.12, 95% CI 1.05-1.19, respectively). Reported symptoms of depression and anxiety were associated with a greater likelihood of using each type of technology. Higher COVID-19 case rates were associated with a significantly greater likelihood of using mental health forums, websites, or apps (OR 2.70, 95% CI 1.49-4.88), and other health forums, websites, or

apps (OR 2.60, 95% CI 1.55-4.34). Time was associated with increased odds of reported use of mental health forums, websites, or apps (OR 1.20, 95% CI 1.11-1.30), phone-based or text-based crisis lines (OR 1.20, 95% CI 1.10-1.31), and online, computer, or console gaming/video gaming (OR 1.12, 95% CI 1.05-1.19). Interactions between COVID-19 case rate and mental health symptoms were not significantly associated with any of the technology types.

Conclusions: Findings suggested increased use of digital mental health tools and other technologies over time during the early stages of the COVID-19 pandemic. As such, additional effort is urgently needed to consider the quality of these products, either by ensuring users have access to evidence-based and evidence-informed technologies and/or by providing them with the skills to make informed decisions around their potential efficacy.

(J Med Internet Res 2021;23(4):e26994) doi:10.2196/26994

KEYWORDS

COVID-19; digital technologies; mHealth; mental health; anxiety; depression; MTurk; e-mental health; digital health; distress; self-management

Introduction

Background

On March 11, 2020, the World Health Organization designated the COVID-19 outbreak a global pandemic, which, among other things, has led to unprecedented hazards to mental health globally [1]. Individual states within the United States began to implement measures to contain the spread of the virus, including limiting travel, mandating physical distancing, and limiting nonessential medical visits. By April 2020, nearly 200,000 cases of COVID-19 and more than 5000 deaths had been reported in the United States [2].

In addition to the unpredictability and uncertainty of the pandemic itself, policy efforts to mitigate risk, such as stay-at-home orders and/or social distancing, introduced a number of additional stressors including social isolation, inactivity, loss of income, and lack of access to basic services, to name but a few [3]. This may be why, mirroring the increase in COVID-19 cases and deaths, there has been an increase in mental distress [1,4-6]. For example, data from the US Census Bureau reported that adults assessed as part of a nationally representative survey in April and May 2020 were more than three times as likely to screen positive for depressive disorders, anxiety disorders, or both, relative to a comparable sample in 2019 [7]. In a similar vein, the Centers for Disease Control and Prevention reported significantly elevated levels of adverse mental health conditions, substance use, and suicidal ideation resulting from the COVID-19 pandemic, with these mental health conditions disproportionately affecting specific populations, such as young adults, Hispanic persons, Black persons, essential workers, unpaid caregivers of adults, and those receiving treatment for preexisting psychiatric conditions [8].

As medical organizations and hospital systems, including those that address mental health, quickly pivoted to digital platforms such as videoconferencing/teleconferencing and patient-provider SMS text messaging [9], the documented growth in and reliance on eHealth/telehealth use emerged as a viable solution to continue providing health and mental health services [10-12]. Indeed, it has been suggested that the COVID-19 pandemic is serving as a "black swan" moment for mental health care—an unforeseen event that will permanently shift mental health care provision toward online prevention, treatment, and care [12]. Given the increase in mental health issues resulting from the pandemic, digital platforms also offer the potential to provide scalable, nonconsumable mental health resources, as they may be less reliant on trained providers [13]. A nonconsumable treatment is one that, once used, retains its therapeutic potential. Unlike a dose of medication, which will not benefit another person once used, digital resources can be used by many people and continue to be helpful [14].

In tandem with this increase in the use of telehealth services and digital platforms, technology companies have also reported increased demand for digital mental health products and therapeutics since the start of the COVID-19 pandemic [15]. However, there have been no empirical studies examining rates of use of these digital mental health products and therapeutics during the COVID-19 pandemic. Examination of marketplace trends through app analytics platforms (eg, App Annie) indicate that downloads and engagement have increased since the onset of COVID-19 [16]. However, it is unclear if reported rates of growth relate to the pandemic or represent an already documented gradual trend of engagement [17]. In light of the transformative changes that have occurred in the delivery of health and mental health care resulting from the COVID-19 pandemic, understanding potential changes in consumer behavior pertaining to the use of digital mental health tools and other technologies during the COVID-19 pandemic is needed.

Current Study

To this end, this study examined the following questions:

1. To what extent was the likelihood of having symptoms indicative of clinical depression and anxiety associated with the county COVID-19 case rate and time?
2. Were individuals with moderate to high self-reported depressive or anxious symptomatology more likely to use digital mental health tools and other technologies compared to individuals who endorsed low depression or anxiety symptom experiences?
3. If so, did the differences in use of these digital mental health tools and other technologies among those with high versus low symptom endorsement vary according to the COVID-19 case rate?

Methods

Participants and Procedures

Participants were recruited from Amazon Mechanical Turk (MTurk), an online crowdsourcing platform commonly used in behavioral science studies [18]. Survey data collection occurred across all 50 states and Washington DC during 4 one-week periods in April, May, June, and July 2020, starting on approximately the 6th of each month. At the first time point, we aimed to recruit approximately 1250 people. At each of the following time points, we aimed to recruit approximately 1750 people.

MTurk allows researchers to collect a large amount of quality data quickly and for relatively little cost [19-21]. On MTurk, requesters are people who post or request tasks (eg, surveys) to be completed, whereas workers are people who are paid for task completion. Requesters can customize the tasks to be available to certain MTurk workers. MTurk workers are able to read descriptions of tasks and select the tasks they are interested in.

For the purpose of this study, study participants had to meet three eligibility criteria: (1) be at least 18 years old, (2) currently reside in the United States, and (3) not have completed the survey in a prior wave. If eligible, participants completed the survey via Qualtrics and received US $6 via MTurk as compensation upon completion. The protocol was approved by the Institutional Review Board (IRB#2019-5406) at the University of California, Irvine.

Measures

Symptoms Indicative of Clinical Depression and Anxiety

The Patient Health Questionnaire-9 (PHQ-9) [22] is a self-report measure used to assess depressive symptoms. It consists of the nine criteria upon which the diagnosis of major depressive disorder is based, according to the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV). The questionnaire uses a 4-point Likert scale (0=not at all, 1=several days, 2=more than half the days, 3=nearly every day) to gauge responses to questions about participants' mental health over the previous 2-week period. Scores on the PHQ-9 can range from 0-27, with 0-4 suggesting no depression symptoms, 5-9 mild symptoms, 10-14 moderate symptoms, 15-19 moderately severe symptoms, and 20-27 more severe depression [23], with a cut-off score of 10 or above being used in this study to indicate the presence of clinically elevated levels of depression symptoms (0=minimal or mild depressive symptomatology and 1=moderate to severe depressive symptomatology). This cut-off has been shown to have high sensitivity (88%) and specificity (88%) [22]. In the present sample, reliability was strong (unweighted α=.92).

The Generalized Anxiety Disorder-7 Questionnaire (GAD-7) [24] was used to assess generalized anxiety disorder. Respondents rate answers (0=not at all, 1=several days, 2=more than half the days, 3=nearly every day) to 7 questions assessing anxiety symptoms experienced over the past two weeks, with the total score ranging from 0-21; scores from 0-4 suggest minimal symptoms, 5-9 mild symptoms, 10-14 moderate symptoms, and 15-21 severe symptoms. A cut-off score of 10 or above was used in this study to indicate the presence of elevated levels of anxiety symptoms (0=minimal to mild anxiety symptomatology and 1=moderate to severe anxiety symptomatology). This cut-off has high sensitivity (89%) and specificity (82%) [24]. Reliability in the present study was high (unweighted α=.92).

Digital Mental Health Tools and Other Technologies to Support Mental Health

Participants were asked how frequently (never, rarely, sometimes, often, or always) they used 20 different types of technology in the last 7 days to manage or support their mental health. Responses were dichotomized (0=never, rarely, sometimes and 1=often or always). The 20 technology types included the following: (1) mental health online forums or communities (eg, Mental Health Forum, BeyondBlue), (2) mental health websites (eg, NAMI, StudentsAgainstDepression.org), (3) mental health apps (eg, 7 Cups, Headspace, Moodpath), (4) phone-based or text-based crisis lines (eg, Crisis Text Line, Suicide Prevention Lifeline), (5) other health online forums or communities (eg, MyFitnessPal forum), (6) other health websites (eg, WebMD, WHO), and (7) other health apps (eg, LoseIt, MapMyRun, Sleep Cycle, Flo). Additionally, participants were also asked about their use of the following technologies to manage their mental health: (8) social media (eg, Facebook, Instagram, Twitter, Snapchat), (9) blogs, (10) online, computer, or console gaming/video gaming, (11) online calendar, checklist, or planner, (12) Word, Notepad, or Google Docs, (13) spreadsheet or Google Sheets, (14) email, (15) texting or messaging software, and (16) video conferencing software. Participants were also able to select the following options: (17) other general online forums or communities, which may include mental health communities (eg, Reddit); and writing in (18) other types of websites, (19) other apps, and (20) other types of technologies not listed [25].

The technologies were collapsed into seven categories by theme: (1) mental health forums, websites, or apps, (2) phone-based or text-based crisis lines, (3) other health forums, websites, or apps, (4) social media and blogs, (5) online, computer, or console gaming/video gaming, (6) online calendar, checklist, planner, Word, Notepad, Google Docs, spreadsheet, or Google Sheets, and (7) email, texting or messaging software, or video conferencing software [25].

County COVID-19 Case Rate

The COVID-19 Data Repository maintained by the Center for Systems Sciences and Engineering at Johns Hopkins University has provided the COVID-19 case count of each county in the United States on a daily basis since January 22, 2020 [2]. Using participants' zip codes, the case count in each participant's county on the date each participant began the survey was merged with the data collected from MTurk. The county case count was converted into the rate of cases per 10 people by dividing the count by each county's total population and multiplying this number by 10. Total county population was obtained from the US Census Bureau's American Community Survey 2018 5-year estimates [26].

Survey Time Period

Survey response windows that reflected the states' population sizes were made available during 4 week-long periods in April, May, June, and July of 2020. As such, this variable also serves as a proxy indicator of time. The time point variable was treated as a continuous variable in the analyses.

Covariates

Additional variables were included as covariates. These encompassed standard demographic characteristics, including age, sex (1=male, 2=female), race/ethnicity (1=non-Hispanic White, 2=Latino, 3=non-Hispanic Asian, 4=non-Hispanic Black or African American, 5=other), marital status (1=married or living with a partner, 2=single or not living with a partner, 3=separated, divorced, or widowed), employment status (1=no change in employment status due to COVID-19, 2=reduced hours due to COVID-19, 3=lost job due to COVID-19), income level, and education level (1=completed high school or less, 2=some college or more). Covariates also included a variable representing state fixed effects.

Analytic Sample

A total of 6704 survey responses were collected. After omitting ineligible people due to duplicate or missing MTurk worker identification numbers, the total sample included 5907 participants. Of this group, 2.22% had missing data for at least one question from the PHQ-9 and 1.64% had missing data for at least one question from the GAD-7. Participants' average PHQ-9 and GAD-7 scores were imputed if at least 50% of the scale questions were answered. Less than 5.51% of sociodemographic data was missing. Binary and continuous variables underwent mean imputation. An additional category for missing data was created for all other sociodemographic covariates. Where participants had a missing zip code, but provided their state of residence, the average county case count and the average county population in the participant's state were used to calculate the county case rate per 10 people. After mean imputation was executed and missing county case rates were replaced with average rates, the analytic sample included 5899 individuals.

Statistical Analyses

Summary statistics were calculated to describe the sociodemographic characteristics of and prevalence of anxiety and depression symptoms in the sample. The use of digital mental health tools and other technologies in the sample was also described. Means, standard errors, and proportions were reported.

To examine the extent to which the likelihood of having symptoms indicative of clinical depression and anxiety was associated with the county COVID-19 case rate and time, we estimated two separate logistic regression models, one for each mental health outcome. Adjusted models included age, sex, race/ethnicity, marital status, employment status, income level, education level, and state fixed effects.

To evaluate whether individuals with moderate to high self-reported depressive or anxious symptomatology were more likely to use digital mental health tools and other technologies compared to individuals who endorsed low depression or anxiety symptom experience, we estimated seven different logistic regression models, one for each of the technology types defined above. Symptoms of anxiety, symptoms of depression, county case rate per 10 people, and time point were the independent variables. Adjusted models included age, sex, race/ethnicity, marital status, education level, income level, and state fixed effects. To descriptively assess the use of digital mental health tools and other technologies over time, the predicted probabilities of employing each type of tool or technology were plotted at each time point, while holding the covariates at their mean values. We also plotted the total number of respondents using each type of technology at each time point.

Finally, to understand if differences in use among those with high versus low symptom endorsement varied according to the rate of COVID-19 cases, we further tested the inclusion of two interactions examining the following in these adjusted models: (1) county-level COVID-19 case rate with depression and (2) county-level COVID-19 case rate with anxiety.

Probability weights were generated to account for oversampling and undersampling of participants in each state. Use of probability weights allowed for the proportion of participants from each state to reflect the proportion of the US population that each state population comprises. Weights were generated using data from the American Community Survey 2018 5-year estimates of US state populations. Survey weights were employed when estimating summary statistics and multivariable regression. Additionally, a Šidák-corrected cut-off P value (P<.0037) was employed to identify significant findings and account for the increased type I error rate that resulted from the use of multiple models [27]. All analyses were conducted in Stata 16 (StataCorp LLC) [28].

Results

Table 1 shows the weighted means and standard errors for continuous variables, as well as the weighted proportions of individuals within each categorical variable group. Overall, the average age was 36.99, and most of the sample was male, non-Hispanic White, and married or living in a marital-like relationship. Most of the sample did not experience a change in employment status due to the COVID-19 pandemic, had an income level greater than US $50,000, and had greater than a high school education. Close to half of the sample experienced symptoms indicative of clinical levels of depression (45.45%) or anxiety (40.04%).

Table 1. Characteristics of analytic sample (N=5899).

Sociodemographic characteristics	Values[a]
County case rate per 10 people, mean (SE)	0.069 (0.0023)
Age, mean (SE)	36.99 (0.17)
Symptoms of depression, %	45.45
Symptoms of anxiety, %	40.04
Sex, %	
Male	57.58
Female	42.42
Race/ethnicity, %	
Non-Hispanic White	61.07
Latino	11.69
Asian	6.54
Black/African American	7.85
Other	12.85
Marital status, %	
Married/living in a marital-like relationship	62.20
Single/never married	31.76
Separated, divorced, or widowed	6.04
Employment status, %	
No change in employment status due to COVID-19	64.42
Reduced hours due to COVID-19	26.88
Lost job due to COVID-19	8.70
Income level (US $), %	
0-10,000	4.67
10,001-20,000	7.42
20,001-30,000	11.71
30,001-40,000	11.91
40,001-50,000	11.85
50,001-60,000	12.14
60,001-70,000	8.75
70,001-80,000	8.60
80,001-90,000	5.24
90,001-100,000	6.14
>100,000	11.56
Education level, %	
High school or less	19.35
More than high school	80.65
Use of digital mental health tools and other technologies to manage mental health, %	
Mental health forums, websites, or apps	27.39
Phone-based or text-based crisis lines	17.74
Other health forums, websites, or apps	34.48
Social media or blogs	47.73
Online, computer, or console gaming/video gaming	33.55

Sociodemographic characteristics	Values[a]
Online calendar, checklist, planner, Word, Notepad, Google Docs, spreadsheet, or Google Sheets	44.55
Email, SMS texting or messaging software, or video conferencing software	55.50

[a]The proportion of missing data ranged from 0.0%-5.51%.

Table 2 shows the results of the two adjusted models regressing the prevalence of clinically meaningful symptoms of depression and anxiety, respectively, on the county COVID-19 case rate per 10 people, time point, and the covariates. As shown, COVID-19 case rate was significantly associated with an increase in the likelihood of experiencing clinically meaningful symptoms of depression (odds ratio [OR] 2.06, 95% CI 1.27-3.35), but not anxiety (OR 1.21, 95% CI 0.77-1.88). Additionally, the likelihood of experiencing symptoms of depression or symptoms of anxiety increased over time (OR 1.19, 95% CI 1.12-1.27 and OR 1.12, 95% CI 1.05-1.19, respectively). Results associated with each covariate are shown in Multimedia Appendix 1.

Table 2. Association between symptoms indicative of clinical levels of depression and anxiety and rates of COVID-19 cases and time: estimates based on two separate logistic models (N=5899)[a].

Variables	Model 1: Symptoms of depression, odds ratio (95% CI)	Model 2: Symptoms of anxiety, odds ratio (95% CI)
County-level COVID-19 case rate per 10 people	2.06[b] (1.27-3.35)	1.21 (0.77-1.88)
Survey time point	1.19[c] (1.12-1.27)	1.12[c] (1.05-1.19)

[a]Models adjusted for the following covariates: age, sex, race/ethnicity, marital status, employment status, income level, education level, and fixed state effects.

[b]$P<.0037$ (Šidák-corrected P value).

[c]$P<.001$.

The adjusted association between depression, anxiety, county case rate per 10 people, time point, and use of each of the seven categories of digital mental health tools and other technologies is presented in Table 3. Results pertaining to each covariate are shown in Multimedia Appendix 1. Both symptoms of depression and anxiety were significantly associated with the likelihood of using each type of tool and technology. Those with depressive symptoms were three to six times more likely to use digital mental health tools than those without depressive symptoms. Those with symptoms of anxiety were two to three times more likely to use digital mental health tools than those without symptoms of anxiety. Specifically, those with depressive symptoms or symptoms of anxiety were more likely than those without symptoms to use mental health forums, websites, or apps (OR 6.01, 95% CI 4.70-7.70 and OR 2.95, 95% CI 2.37-3.66, respectively), phone-based or text-based crisis lines (OR 4.98, 95% CI 3.66-6.77 and OR 2.85, 95% CI 2.22-3.66, respectively), and other health forums, websites, or apps (OR 3.44, 95% CI 2.81-4.20 and OR 2.55, 95% CI 2.11-3.10, respectively).

The findings pertaining to digital mental health tools mirrored the results from models assessing use of other technologies not necessarily related to health. Specifically, those with symptoms indicative of depression and anxiety were more likely, as compared to those without symptoms, to engage in use of social media and blogs (OR 1.56, 95% CI 1.31-1.86 and OR 1.80, 95% CI 1.51-2.14, respectively); online, computer, or console gaming/video gaming (OR 1.63, 95% CI 1.36-1.95 and OR 1.67, 95% CI 1.40-1.99, respectively); online calendar, checklist, planner, Word, Notepad, Google Docs, spreadsheet, or Google Sheets (OR 1.91, 95% CI 1.60-2.28 and OR 2.09, 95% CI 1.75-2.50, respectively); and email, texting or messaging software, or video conferencing software (OR 1.66, 95% CI 1.39-1.98 and OR 1.82, 95% CI 1.52-2.17, respectively). However, in general, even though significant, the odds ratios for these other technologies tended to be much smaller, with the largest being 2.09 and all others below 2.00, while all the odds ratios for the mental health technologies were well above 2.00, ranging from 2.55 to 6.01.

With regard to the association between county-level COVID-19 case rate and digital mental health tool and other technology use, a higher county case rate per 10 people was associated with increased likelihood of using digital mental health tools, specifically mental health forums, websites, or apps (OR 2.70, 95% CI 1.49-4.88) and other health forums, websites, or apps (OR 2.60, 95% CI 1.55-4.34). Additionally, over time, there was an increase in the likelihood of using mental health forums, websites, or apps (OR 1.20, 95% CI 1.11-1.30); phone-based or text-based crisis lines (OR 1.20, 95% CI 1.10-1.31); other health forums, websites, or apps (OR 1.12, 95% CI 1.05-1.20); and online, computer, or console gaming/video gaming (OR 1.12, 95% CI 1.05-1.19). The predicted probabilities of using each type of technology at each time point while holding the covariates at their mean values are graphed in Multimedia Appendix 2. Multimedia Appendix 3 shows the total number of respondents using each type of technology at each time point.

Table 3. Associations between use of digital mental health tools and other technologies and prevalence of mental illness symptoms and the rate of COVID-19 cases: estimates based on seven separate logistic models[a].

Variables	Model 1: Mental health forums, websites, or apps (n=5849)	Model 2: Phone-based or text-based crisis lines (n=5831)	Model 3: Other health forums, websites, or apps (n=5854)	Model 4: Social media and blogs (n=5788)	Model 5: Online, computer, or console gaming/video gaming (n=5866)	Model 6: Online calendar, checklist, planner, Word, Notepad, Google Docs, spreadsheet, or Google Sheets (n=5835)	Model 7: Email, texting or messaging software, or video conferencing software (n=5849)
	OR[b] (95% CI)	OR (95% CI)	OR (95% CI)	OR (95% CI)	OR (95% CI)	OR (95% CI)	OR (95% CI)
Symptoms of depression	6.01[c] (4.70-7.70)	4.98[c] (3.66-6.77)	3.44[c] (2.81-4.20)	1.56[c] (1.31-1.86)	1.63[c] (1.36-1.95)	1.91[c] (1.60-2.28)	1.66[c] (1.39-1.98)
Symptoms of anxiety	2.95[c] (2.37-3.66)	2.85[c] (2.22-3.66)	2.55[c] (2.11-3.10)	1.80[c] (1.51-2.14)	1.67[c] (1.40-1.99)	2.09[c] (1.75-2.50)	1.82[c] (1.52-2.17)
County-level COVID-19 case rate per 10 people	2.70[d] (1.49-4.88)	1.81[e] (1.02-3.19)	2.60[c] (1.55-4.34)	1.49 (0.95-2.36)	1.65[e] (1.07-2.56)	2.04[e] (1.26-3.30)	1.77[e] (1.09-2.89)
Survey time point	1.20[c] (1.11-1.30)	1.20[c] (1.10-1.31)	1.12[d] (1.05-1.20)	1.08[e] (1.02-1.15)	1.12[c] (1.05-1.19)	1.08[e] (1.01-1.15)	1.08[e] (1.02-1.14)

[a]Models adjusted for the following covariates: age, sex, race/ethnicity, marital status, employment status, income level, education level, and fixed state effects.

[b]OR: odds ratio.

[c]*P*<.001.

[d]*P*<.0037 (Šidák-corrected *P* value).

[e]*P*<.05.

Lastly, models assessing the interaction of county-level COVID-19 case rate and symptoms indicative of anxiety or depression did not yield significant findings (analyses available on request). For this reason, interaction terms were not subject to further analyses.

Discussion

To our knowledge, this is the first study to examine rates of use of digital mental health tools and other technologies to manage one's mental health among a sample of people during the early stages of the COVID-19 pandemic. Similar to other studies [29], rates of COVID-19 cases were associated with increased rates of depressive symptoms. Both depressive and anxiety symptoms and COVID-19 county case rates were associated with the largest increased likelihood of using digital mental health tools, specifically mental health forums, websites, and apps; phone-based/text-based crisis lines; and other health forums, websites, and apps, reflecting the greatest increase in likelihood of use for products designed specifically for health and mental health. Interestingly, we also found a general increase in the likelihood of using other technologies to support one's mental health, including gaming and work products, among those experiencing symptoms of mental illness. This increase in the use of other technologies may reflect the transition to working from home that occurred for many people.

We also sought to examine if the likelihood of using these tools and technologies was the highest among people who were both living in counties with high COVID-19 case rates *and* exhibiting symptoms indicative of clinically meaningful levels of depression or anxiety. The findings suggested, in fact, that there was no additional significant likelihood of increased use among respondents with depressive or anxious symptoms living in counties with high COVID-19 case rates compared to individuals with similar symptom levels in counties with low COVID-19 rates. Of course, it is likely that other factors, such as the impacts of policy measures aimed at mitigating risks (eg, physical distancing, working from home, or school closures) may have had a greater impact on mental health symptoms, resulting in people turning to digital tools for mental health support. In the absence of in-person connections or services, digital tools appear to play an important role in combatting the mental health impact of the pandemic.

We also found that the odds of having used these technologies generally increased over time. This increase in use of digital mental health tools and other technologies may be due, in part, to increased access to some of these technologies. For example, cities like New York City [30] have made technologies freely available to their residents and/or shared recommendations for particular products that have been vetted. Additionally, health insurance companies, such as Kaiser Permanente and others, have also made mental health products freely available for their enrollees [31]. It might also be due to the fact that people are spending more time interacting with technologies and these may be convenient ways to receive mental health support.

Our findings that people are increasingly using various technologies—both those specifically designed for mental health support as well as those that are not—are consistent with studies that explore the various ways people use technology to self-manage their mental health [32-34]. These findings also support recent calls for understanding how the technology

ecosystem might impact mental health and lead to opportunities for prevention and intervention tools [35]. Some efforts, such as Google's integration of mental health screening into their search engine, have been launched [36]. In light of our findings, it appears as though there is consumer interest for such resources. Furthermore, an important consideration for future work is to ensure that consumers find effective and safe resources, and have the proper support for using such resources appropriately.

Indeed, despite this increased use and interest, evidence-based and safe resources are rarely available for consumers. One study has suggested that only 2.08% of publicly available psychosocial wellness and stress management mobile apps have published, peer-reviewed evidence of feasibility and/or efficacy [37]. Furthermore, few products provide sufficient information to gauge their safety and privacy [38], and even among those that do, many share information with third parties in ways that might not be disclosed in those policies [39]. Although various efforts have been launched that either evaluate or offer evaluative frameworks for mental health apps (including One Mind PsyberGuide [40], ORCHA [41], and the American Psychiatric Association app evaluation framework [42]), no widely accepted or coordinated effort at regulation and evaluation exists in the United States, despite multiple calls for such models (see National Institutes of Health, National Advisory Work Group [43]). Indeed, better regulation and better access to information at point-of-access could be a significant improvement in helping guide consumers [44], who are demonstrating a clear interest in such resources, and might serve an important need in light of the COVID-19 pandemic.

Several study limitations should be noted. First, this survey was completed by MTurk workers who regularly use the computer, and therefore have the necessary technology, mobility, and digital literacy to participate [45]. This might contribute to the rates of technology use reported and the ability to use technologies—both those designed for mental health and otherwise—to support their mental health. Additionally, it is likely that people with the lowest socioeconomic status, those that are the most isolated, and those with limited access to technology were omitted from the study. These individuals may represent those most affected by COVID-19 and thus may have the greatest mental health needs [46]. Second, the survey was only available in English, and thus the findings may not hold among groups with limited English proficiency who also might be less likely to use these technologies, especially because many are not designed for diverse populations [47]. Third, this study only included people living in the United States, and thus, findings cannot be generalized to other countries [1]. Fourth, there exist a number of scales pertaining to examining the self-reported impact of COVID-19 on mental health [48-51]. These works were primarily published during the development and implementation of this survey, and as such, they were not included in this study. Fifth, there exists the possibility that at least some of the survey responses were bot-generated. To test for the potential influence of bots, we executed post hoc diagnostics previously suggested by Chmielewski and Kucker [52]. Removing those responses suspected to be bot-generated did not substantially change the findings. Thus, we reported on those findings generated from the full analytic sample. However, we also have included the results garnered after removing tagged responses (Multimedia Appendix 4).

Finally, this study focused on examining the extent to which people used mobile health technologies throughout the early period of the pandemic. The findings do not shed light on the effectiveness of these particular strategies for managing mental health. For many, engagement in digital mental health tool and other technology use has been critical for enhancing social connectedness, managing stress and anxiety, and providing greatly needed entertainment [53]. Although the vast majority of the increase in digital technology use is adaptive, it is important to note that there are likely negative impacts of this growth, including the spread of false information [54-56]. Furthermore, there exist subgroups of vulnerable individuals that are at risk of developing problematic usage patterns [53]. Excessive engagement in specific online activities such as video gaming, social media use, and shopping has been linked with severe problems and can elevate the risk of disordered or addictive use [57,58].

In conclusion, this study provides an important description of the prevalence of symptoms of anxiety and depression and digital mental health tool and other technology use during the early stages of the COVID-19 pandemic. Specifically, we found evidence of an increased likelihood of experiencing depressive symptoms both when considering differences between county COVID-19 case rates and changes over time. Furthermore, COVID-19 case rate and time were generally associated with an increased likelihood of using digital mental health tools and other technologies. Lastly, those experiencing symptoms of depression or anxiety were more likely than those without such symptoms to use tools and technologies to manage their mental health.

As the pandemic is expected to continue well into the spring and summer of 2021, and as depression [59,60], anxiety [7,61], and suicidal ideation rates continue to climb [1,8,62], the importance of providing easy access to tools and technologies to manage one's mental health will become even more important. The current shortage of mental health professionals demands the need to explore more scalable solutions that might be able to be adapted and deployed to meet the needs of various populations. The findings from this study—as well as future research that may address more specific issues designed to understand how digital mental health tools and other technologies can be more accessible and effective for the populations who need them—could inform public health efforts.

Acknowledgments

This work was funded by Help@Hand (agreement number 417-ITS-UCI-2019), a project overseen by the California Mental Health Service Authority (CalMHSA). CalMHSA reviewed the manuscript for confidentiality. The information or content and

conclusions presented here are those of the authors, and should not be construed as the official position or policy of, nor should any endorsements be inferred by, the participating Help@Hand Counties and/or CalMHSA.

Authors' Contributions

DHS is the first and corresponding author and contributed to the study concept and design, drafting of the manuscript, and critical revision of the manuscript. EAJ conducted analyses, contributed to the drafting of the manuscript, and critically revised the manuscript. EVE, MS, KD, SMS, NAS, and KZ contributed to the study concept and design, provided feedback on manuscript drafts, and critically revised the manuscript. DS and MN provided feedback on manuscript drafts and critically revised the manuscript. DM is the senior author and contributed to the study concept and design, provided feedback on manuscript drafts, and critically revised the manuscript.

Conflicts of Interest

SMS is a member of the Scientific Advisory Board for Headspace for which he receives compensation.

Multimedia Appendix 1
Regression models employing full sample.
[DOCX File , 51 KB - jmir_v23i4e26994_app1.docx]

Multimedia Appendix 2
Predicted probability of using a type of digital mental health tool or other technology at each time point when covariates are at their mean values.
[PNG File , 43 KB - jmir_v23i4e26994_app2.png]

Multimedia Appendix 3
Number of participants using a type of digital mental health tool or other technology at each time point.
[PNG File , 49 KB - jmir_v23i4e26994_app3.png]

Multimedia Appendix 4
Regression models using the sample generated after removing responses suspected to be bot-generated.
[DOCX File , 51 KB - jmir_v23i4e26994_app4.docx]

References

1. Xiong J, Lipsitz O, Nasri F, Lui LM, Gill H, Phan L, et al. Impact of COVID-19 pandemic on mental health in the general population: A systematic review. J Affect Disord 2020 Dec 01;277:55-64 [FREE Full text] [doi: 10.1016/j.jad.2020.08.001] [Medline: 32799105]
2. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. The Lancet Infectious Diseases 2020 May;20(5):533-534. [doi: 10.1016/s1473-3099(20)30120-1]
3. Moreno C, Wykes T, Galderisi S, Nordentoft M, Crossley N, Jones N, et al. How mental health care should change as a consequence of the COVID-19 pandemic. The Lancet Psychiatry 2020 Sep;7(9):813-824. [doi: 10.1016/s2215-0366(20)30307-2]
4. Xiang Y, Yang Y, Li W, Zhang L, Zhang Q, Cheung T, et al. Timely mental health care for the 2019 novel coronavirus outbreak is urgently needed. The Lancet Psychiatry 2020 Mar;7(3):228-229. [doi: 10.1016/s2215-0366(20)30046-8]
5. Ayers JW, Leas EC, Johnson DC, Poliak A, Althouse BM, Dredze M, et al. Internet Searches for Acute Anxiety During the Early Stages of the COVID-19 Pandemic. JAMA Intern Med 2020 Aug 24;180(12):1706. [doi: 10.1001/jamainternmed.2020.3305] [Medline: 32832984]
6. Stijelja S, Mishara BL. COVID-19 and Psychological Distress-Changes in Internet Searches for Mental Health Issues in New York During the Pandemic. JAMA Intern Med 2020 Oct 05;180(12):1703. [doi: 10.1001/jamainternmed.2020.3271] [Medline: 33016982]
7. Twenge JM, Joiner TE. U.S. Census Bureau-assessed prevalence of anxiety and depressive symptoms in 2019 and during the 2020 COVID-19 pandemic. Depress Anxiety 2020 Oct 15;37(10):954-956 [FREE Full text] [doi: 10.1002/da.23077] [Medline: 32667081]
8. Czeisler, Lane RI, Petrosky E, Wiley JF, Christensen A, Njai R, et al. Mental Health, Substance Use, and Suicidal Ideation During the COVID-19 Pandemic - United States, June 24-30, 2020. MMWR Morb Mortal Wkly Rep 2020 Aug 14;69(32):1049-1057 [FREE Full text] [doi: 10.15585/mmwr.mm6932a1] [Medline: 32790653]
9. Ben-Zeev D. The Digital Mental Health Genie Is Out of the Bottle. Psychiatr Serv 2020 Dec 01;71(12):1212-1213. [doi: 10.1176/appi.ps.202000306] [Medline: 32576123]

10. Dores AR, Geraldo A, Carvalho IP, Barbosa F. The Use of New Digital Information and Communication Technologies in Psychological Counseling during the COVID-19 Pandemic. Int J Environ Res Public Health 2020 Oct 21;17(20):7663 [FREE Full text] [doi: 10.3390/ijerph17207663] [Medline: 33096650]
11. Chew AMK, Ong R, Lei H, Rajendram M, Verma SK, Fung DSS, et al. Digital Health Solutions for Mental Health Disorders During COVID-19. Front Psychiatry 2020 Sep 9;11:582007 [FREE Full text] [doi: 10.3389/fpsyt.2020.582007] [Medline: 33033487]
12. Wind TR, Rijkeboer M, Andersson G, Riper H. The COVID-19 pandemic: The 'black swan' for mental health care and a turning point for e-health. Internet Interv 2020 Apr;20:100317 [FREE Full text] [doi: 10.1016/j.invent.2020.100317] [Medline: 32289019]
13. Torous J, Jän Myrick K, Rauseo-Ricupero N, Firth J. Digital Mental Health and COVID-19: Using Technology Today to Accelerate the Curve on Access and Quality Tomorrow. JMIR Ment Health 2020 Mar 26;7(3):e18848 [FREE Full text] [doi: 10.2196/18848] [Medline: 32213476]
14. Muñoz RF. Using evidence-based internet interventions to reduce health disparities worldwide. J Med Internet Res 2010 Dec 17;12(5):e60 [FREE Full text] [doi: 10.2196/jmir.1463] [Medline: 21169162]
15. Inkster B. Early Warning Signs of a Mental Health Tsunami: A Coordinated Response to Gather Initial Data Insights From Multiple Digital Services Providers. Front Digit Health 2021 Feb 10;2:578902. [doi: 10.3389/fdgth.2020.578902]
16. Mobile app usage booms during COVID-19, says App Annie. ZDNet. URL: https://www.zdnet.com/article/mobile-app-usage-booms-during-covid-19-says-app-annie/ [accessed 2021-04-08]
17. Muoio D. Mental wellness app use, adoption hasn't spiked during COVID-19. Mobihealthnews. URL: https://www.mobihealthnews.com/news/mental-wellness-app-use-adoption-hasnt-spiked-during-COVID-19 [accessed 2021-04-08]
18. Chandler J, Rosenzweig C, Moss AJ, Robinson J, Litman L. Online panels in social science research: Expanding sampling methods beyond Mechanical Turk. Behav Res Methods 2019 Oct 11;51(5):2022-2038 [FREE Full text] [doi: 10.3758/s13428-019-01273-7] [Medline: 31512174]
19. Chandler J, Shapiro D. Conducting Clinical Research Using Crowdsourced Convenience Samples. Annu Rev Clin Psychol 2016 Mar 28;12(1):53-81. [doi: 10.1146/annurev-clinpsy-021815-093623] [Medline: 26772208]
20. Robinson J, Rosenzweig C, Moss AJ, Litman L. Tapped out or barely tapped? Recommendations for how to harness the vast and largely unused potential of the Mechanical Turk participant pool. PLoS One 2019 Dec 16;14(12):e0226394 [FREE Full text] [doi: 10.1371/journal.pone.0226394] [Medline: 31841534]
21. Buhrmester M, Kwang T, Gosling SD. Amazon's Mechanical Turk: A New Source of Inexpensive, Yet High-Quality, Data? Perspect Psychol Sci 2011 Jan 03;6(1):3-5. [doi: 10.1177/1745691610393980] [Medline: 26162106]
22. Kroenke K, Spitzer RL, Williams JBW. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
23. Manea L, Gilbody S, McMillan D. Optimal cut-off score for diagnosing depression with the Patient Health Questionnaire (PHQ-9): a meta-analysis. CMAJ 2012 Feb 21;184(3):E191-E196 [FREE Full text] [doi: 10.1503/cmaj.110829] [Medline: 22184363]
24. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
25. Dodson J, Saint PN, Thang J, Eikey EV. Depression Management as Lifestyle Management: Exploring Existing Practices and Perceptions Among College Students. In: iConference 2020: Sustainable Digital Communities. 2020 Presented at: iConference: International Conference on Information; March 23-26, 2020; Boras, Sweden p. 237-255. [doi: 10.1007/978-3-030-43687-2_19]
26. United States Census Bureau. American Community Survey 5-Year Estimates, Table B01003. 2018. URL: https://tinyurl.com/272cnrba [accessed 2020-11-04]
27. Abdi H. The Bonferonni-Šidák Corrections for Multiple Comparisons. In: Salkind N, editor. Encyclopedia of Measurement and Statistics. Thousands Oaks, CA, USA: Sage; 2007:103-107.
28. StataCorp. Stata: Software for Statistics and Data Science. URL: https://www.stata.com/ [accessed 2021-04-08]
29. Holman EA, Thompson RR, Garfin DR, Silver RC. The unfolding COVID-19 pandemic: A probability-based, nationally representative study of mental health in the United States. Sci Adv 2020 Oct 18;6(42):eabd5390 [FREE Full text] [doi: 10.1126/sciadv.abd5390] [Medline: 32948511]
30. COVID-19 Digital Mental Health Resources. NYC Well. URL: https://nycwell.cityofnewyork.us/en/COVID-19-digital-mental-health-resources/ [accessed 2021-04-08]
31. Mordecai D, Histon T, Neuwirth E, Heisler WS, Kraft A, Bang Y, et al. How Kaiser Permanente Created a Mental Health and Wellness Digital Ecosystem. NEJM Catalyst 2021 Jan;2(1):1. [doi: 10.1056/CAT.20.0295]
32. Eschler J, Burgess E, Reddy M, Mohr DC. Emergent Self-Regulation Practices in Technology and Social Media Use of Individuals Living with Depression. In: CHI '20: Proceedings of the CHI Conference on Human Factors in Computing Systems. 2020 Presented at: CHI '20: CHI Conference on Human Factors in Computing Systems; April, 2020; Honolulu, HI, USA p. 1-13. [doi: 10.1145/3313831.3376773]

33. O'Reilly M, Dogra N, Hughes J, Reilly P, George R, Whiteman N. Potential of social media in promoting mental health in adolescents. Health Promot Int 2019 Oct 01;34(5):981-991 [FREE Full text] [doi: 10.1093/heapro/day056] [Medline: 30060043]
34. De Choudhury M, De S. Mental Health Discourse on reddit: Self-Disclosure, Social Support, and Anonymity. In: Proceedings of the International AAAI Conference on Web and Social Media. 2014 Presented at: Eighth International AAAI Conference on Weblogs and Social Media; June 1-4, 2014; Ann Arbor, MI, USA URL: https://ojs.aaai.org/index.php/ICWSM/article/view/14526
35. Burgess ER, Zhang R, Ernala SK, Feuston JL, De Choudhury M, Czerwinski M, et al. Technology ecosystems. Interactions 2021 Jan;28(1):66-71. [doi: 10.1145/3434564]
36. Gilbert D. Patient Commentary: Online screening for depression-old (paternalistic) wine in new (digital) bottles. BMJ 2017 Sep 13;358:j4207. [doi: 10.1136/bmj.j4207] [Medline: 28903927]
37. Lau N, O'Daffer A, Colt S, Yi-Frazier JP, Palermo TM, McCauley E, et al. Android and iPhone Mobile Apps for Psychosocial Wellness and Stress Management: Systematic Search in App Stores and Literature Review. JMIR mHealth uHealth 2020 May 22;8(5):e17798 [FREE Full text] [doi: 10.2196/17798] [Medline: 32357125]
38. O'Loughlin K, Neary M, Adkins EC, Schueller SM. Reviewing the data security and privacy policies of mobile apps for depression. Internet Interv 2019 Mar;15:110-115 [FREE Full text] [doi: 10.1016/j.invent.2018.12.001] [Medline: 30792962]
39. Huckvale K, Torous J, Larsen ME. Assessment of the Data Sharing and Privacy Practices of Smartphone Apps for Depression and Smoking Cessation. JAMA Netw Open 2019 Apr 05;2(4):e192542 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.2542] [Medline: 31002321]
40. Neary M, Schueller SM. State of the Field of Mental Health Apps. Cogn Behav Pract 2018 Nov;25(4):531-537 [FREE Full text] [doi: 10.1016/j.cbpra.2018.01.002] [Medline: 33100810]
41. Leigh S, Ouyang J, Mimnagh C. Effective? Engaging? Secure? Applying the ORCHA-24 framework to evaluate apps for chronic insomnia disorder. Evid Based Ment Health 2017 Nov 25;20(4):e20-e20. [doi: 10.1136/eb-2017-102751] [Medline: 28947676]
42. Torous JB, Chan SR, Gipson SYT, Kim JW, Nguyen T, Luo J, et al. A Hierarchical Framework for Evaluation and Informed Decision Making Regarding Smartphone Apps for Clinical Care. Psychiatr Serv 2018 May 01;69(5):498-500. [doi: 10.1176/appi.ps.201700423] [Medline: 29446337]
43. National Advisory Mental Health Council (NAMHC) Workgroup. Opportunities and Challenges of Developing Information Technologies on Behavioral and Social Science Clinical Research. URL: https://www.nimh.nih.gov/about/advisory-boards-and-groups/namhc/reports/opportunities-and-challenges-of-developing-information-technologies-on-behavioral-and-social-science-clinical-research.shtml [accessed 2020-01-06]
44. Wykes T, Schueller S. Why Reviewing Apps Is Not Enough: Transparency for Trust (T4T) Principles of Responsible Health App Marketplaces. J Med Internet Res 2019 May 02;21(5):e12390 [FREE Full text] [doi: 10.2196/12390] [Medline: 31045497]
45. Cheung JH, Burns DK, Sinclair RR, Sliter M. Amazon Mechanical Turk in Organizational Psychology: An Evaluation and Practical Recommendations. J Bus Psychol 2016 Jun 29;32(4):347-361. [doi: 10.1007/s10869-016-9458-5]
46. Bol N, Helberger N, Weert JCM. Differences in mobile health app use: A source of new digital inequalities? The Information Society 2018 Apr 26;34(3):183-193. [doi: 10.1080/01972243.2018.1438550]
47. Schueller SM, Hunter JF, Figueroa C, Aguilera A. Use of Digital Mental Health for Marginalized and Underserved Populations. Curr Treat Options Psych 2019 Jul 5;6(3):243-255. [doi: 10.1007/s40501-019-00181-z]
48. Ahorsu DK, Lin C, Imani V, Saffari M, Griffiths MD, Pakpour AH. The Fear of COVID-19 Scale: Development and Initial Validation. Int J Ment Health Addiction 2020 Mar 27:1. [doi: 10.1007/s11469-020-00270-8]
49. Lee SA. Coronavirus Anxiety Scale: A brief mental health screener for COVID-19 related anxiety. Death Stud 2020 Apr 16;44(7):393-401. [doi: 10.1080/07481187.2020.1748481] [Medline: 32299304]
50. Leenders NYJM, Sherman WM, Nagaraja HN. Comparisons of four methods of estimating physical activity in adult women. Med Sci Sports Exerc 2000 Jul;32(7):1320-1326. [doi: 10.1097/00005768-200007000-00021] [Medline: 10912900]
51. Taylor S, Landry CA, Paluszek MM, Fergus TA, McKay D, Asmundson GJ. Development and initial validation of the COVID Stress Scales. J Anxiety Disord 2020 May;72:102232 [FREE Full text] [doi: 10.1016/j.janxdis.2020.102232] [Medline: 32408047]
52. Chmielewski M, Kucker SC. An MTurk Crisis? Shifts in Data Quality and the Impact on Study Results. Social Psychological and Personality Science 2019 Oct 10;11(4):464-473. [doi: 10.1177/1948550619875149]
53. Király O, Potenza MN, Stein DJ, King DL, Hodgins DC, Saunders JB, et al. Preventing problematic internet use during the COVID-19 pandemic: Consensus guidance. Compr Psychiatry 2020 Jul;100:152180 [FREE Full text] [doi: 10.1016/j.comppsych.2020.152180] [Medline: 32422427]
54. Olatunji O, Ayandele O, Ashirudeen D, Olaniru O. "Infodemic" in a pandemic: COVID-19 conspiracy theories in an African country. Soc Health Behav 2020;3(4):152-157. [doi: 10.4103/SHB.SHB_43_20]
55. Lin M, Cheng Y. Policy actions to alleviate psychosocial impacts of COVID-19 pandemic: Experiences from Taiwan. Soc Health Behav 2020;3(2):72. [doi: 10.4103/shb.shb_18_20]

56. Rieger M. To wear or not to wear? Factors influencing wearing face masks in Germany during the COVID-19 pandemic. Soc Health Behav 2020;3(2):50. [doi: 10.4103/shb.shb_23_20]
57. Machimbarrena JM, Calvete E, Fernández-González L, Álvarez-Bardón A, Álvarez-Fernández L, González-Cabrera J. Internet Risks: An Overview of Victimization in Cyberbullying, Cyber Dating Abuse, Sexting, Online Grooming and Problematic Internet Use. Int J Environ Res Public Health 2018 Nov 05;15(11):2471 [FREE Full text] [doi: 10.3390/ijerph15112471] [Medline: 30400659]
58. Spada MM. An overview of problematic Internet use. Addictive Behaviors 2014 Jan;39(1):3-6. [doi: 10.1016/j.addbeh.2013.09.007]
59. Ettman CK, Abdalla SM, Cohen GH, Sampson L, Vivier PM, Galea S. Prevalence of Depression Symptoms in US Adults Before and During the COVID-19 Pandemic. JAMA Netw Open 2020 Sep 01;3(9):e2019686 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.19686] [Medline: 32876685]
60. Bueno-Notivol J, Gracia-García P, Olaya B, Lasheras I, López-Antón R, Santabárbara J. Prevalence of depression during the COVID-19 outbreak: A meta-analysis of community-based studies. Int J Clin Health Psychol 2021 Jan;21(1):100196 [FREE Full text] [doi: 10.1016/j.ijchp.2020.07.007] [Medline: 32904715]
61. Gallagher MW, Zvolensky MJ, Long LJ, Rogers AH, Garey L. The Impact of Covid-19 Experiences and Associated Stress on Anxiety, Depression, and Functional Impairment in American Adults. Cognit Ther Res 2020 Aug 29;44(6):1-9 [FREE Full text] [doi: 10.1007/s10608-020-10143-y] [Medline: 32904454]
62. Ammerman BA, Burke TA, Jacobucci R, McClure K. Preliminary investigation of the association between COVID-19 and suicidal thoughts and behaviors in the U.S. J Psychiatr Res 2021 Feb;134:32-38. [doi: 10.1016/j.jpsychires.2020.12.037] [Medline: 33360222]

Abbreviations

GAD-7: Generalized Anxiety Disorder-7
MTurk: Amazon Mechanical Turk
OR: odds ratio
PHQ-9: Patient Health Questionnaire-9

Edited by G Eysenbach; submitted 06.01.21; peer-reviewed by CY Lin, J Schleider; comments to author 28.01.21; revised version received 18.02.21; accepted 03.04.21; published 16.04.21.

Please cite as:
Sorkin DH, Janio EA, Eikey EV, Schneider M, Davis K, Schueller SM, Stadnick NA, Zheng K, Neary M, Safani D, Mukamel DB
Rise in Use of Digital Mental Health Tools and Technologies in the United States During the COVID-19 Pandemic: Survey Study
J Med Internet Res 2021;23(4):e26994
URL: https://www.jmir.org/2021/4/e26994
doi:10.2196/26994
PMID:33822737

Original Paper

Prediction and Feature Importance Analysis for Severity of COVID-19 in South Korea Using Artificial Intelligence: Model Development and Validation

Heewon Chung[1*], MSc; Hoon Ko[1*], MSc; Wu Seong Kang[2*], MD, PhD; Kyung Won Kim[3], MD, PhD; Hooseok Lee[1], MSc; Chul Park[4], MD; Hyun-Ok Song[5], PhD; Tae-Young Choi[6], PhD; Jae Ho Seo[7], PhD; Jinseok Lee[1], PhD

[1]Department of Artificial Intelligence, The Catholic University of Korea, Bucheon, Republic of Korea
[2]Department of Trauma Surgery, Jeju Regional Trauma Center, Cheju Halla General Hospital, Jeju, Republic of Korea
[3]Radiology and Research Institute of Radiology, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Republic of Korea
[4]Department of Internal Medicine, Wonkwang University School of Medicine, Iksan, Republic of Korea
[5]Department of Infection Biology, Wonkwang University School of Medicine, Iksan, Republic of Korea
[6]Department of Pathology, Wonkwang University School of Medicine, Iksan, Republic of Korea
[7]Department of Biochemistry, Wonkwang University School of Medicine, Iksan, Republic of Korea
[*]these authors contributed equally

Corresponding Author:
Jinseok Lee, PhD
Department of Artificial Intelligence
The Catholic University of Korea
43 Jibong-ro
Bucheon, 14662
Republic of Korea
Phone: 82 2 2164 5523
Email: gonasago@catholic.ac.kr

Abstract

Background: The number of deaths from COVID-19 continues to surge worldwide. In particular, if a patient's condition is sufficiently severe to require invasive ventilation, it is more likely to lead to death than to recovery.

Objective: The goal of our study was to analyze the factors related to COVID-19 severity in patients and to develop an artificial intelligence (AI) model to predict the severity of COVID-19 at an early stage.

Methods: We developed an AI model that predicts severity based on data from 5601 COVID-19 patients from all national and regional hospitals across South Korea as of April 2020. The clinical severity of COVID-19 was divided into two categories: low and high severity. The condition of patients in the low-severity group corresponded to no limit of activity, oxygen support with nasal prong or facial mask, and noninvasive ventilation. The condition of patients in the high-severity group corresponded to invasive ventilation, multi-organ failure with extracorporeal membrane oxygenation required, and death. For the AI model input, we used 37 variables from the medical records, including basic patient information, a physical index, initial examination findings, clinical findings, comorbid diseases, and general blood test results at an early stage. Feature importance analysis was performed with AdaBoost, random forest, and eXtreme Gradient Boosting (XGBoost); the AI model for predicting COVID-19 severity among patients was developed with a 5-layer deep neural network (DNN) with the 20 most important features, which were selected based on ranked feature importance analysis of 37 features from the comprehensive data set. The selection procedure was performed using sensitivity, specificity, accuracy, balanced accuracy, and area under the curve (AUC).

Results: We found that age was the most important factor for predicting disease severity, followed by lymphocyte level, platelet count, and shortness of breath or dyspnea. Our proposed 5-layer DNN with the 20 most important features provided high sensitivity (90.2%), specificity (90.4%), accuracy (90.4%), balanced accuracy (90.3%), and AUC (0.96).

Conclusions: Our proposed AI model with the selected features was able to predict the severity of COVID-19 accurately. We also made a web application so that anyone can access the model. We believe that sharing the AI model with the public will be helpful in validating and improving its performance.

(J Med Internet Res 2021;23(4):e27060) doi:10.2196/27060

KEYWORDS

COVID-19; artificial intelligence; blood samples; mortality prediction

Introduction

The COVID-19 pandemic has had a major impact on health care systems globally. Since early 2020, COVID-19 has spread rapidly around the world, exceeding 100 million cases and 2 million deaths [1]. In the COVID-19 pandemic situation, the most important issue in the management of COVID-19 patients is to triage patients at high risk of mortality and provide tailored treatment, so that medical costs and mortality rates can be reduced.

Several models have been proposed to predict the severity or mortality of COVID-19 patients using artificial intelligence (AI) techniques. The majority of them have been developed based on limited information or variables, such as medical images [2-7], blood and/or urine information [8,9], clinical characteristics [10-12], individual-level epidemiological data sets [13], and electronic health records (ie, demographics, laboratory results, medical history, and vital signs) during hospitalization [14]. However, most of them were developed based on relatively small samples from limited data sources, which makes their generalization problematic. More specifically, the numbers of patients used for training of some models were 375 [15], 443 [16], 548 [17], and 663 [18].

To overcome the generalization issue, we aimed to develop an AI prediction model based on confirmed nationwide patient data obtained from the South Korean government, which included 5601 patients from more than 100 hospitals. In this model, we used comprehensive data sets composed of 37 factors, including basic demographic information, vital signs, physical examination results, clinical symptoms and severity, comorbid diseases, and general blood test results. To the best of our knowledge, this is the first attempt to develop an AI model to predict the severity of COVID-19 based on a nationwide cohort and comprehensive data set in South Korea.

Methods

Data Sets

This study was approved by the Korea Disease Control and Prevention Agency (KDCA) in South Korea. Informed consent was waived. The KDCA has been managing comprehensive data from COVID-19–confirmed patients in Korea obtained from approximately 100 hospitals. The KDCA discloses this data to few selected researchers during a specific study period. Thus, we investigated this data between September 15 and October 5, 2020, under the approval of the KDCA.

Table 1 describes the KDCA data set. The basic patient information includes the patient's ID, age, gender, outcome, quarantine period, pregnancy status, and pregnancy week. The physical index includes body mass index. The initial examination findings include systolic and diastolic blood pressure, heart rate average, and body temperature at the hospital admission stage. The clinical findings include the status of fever, cough, sputum production, sore throat, rhinorrhea, myalgia, malaise, dyspnea, headache, confusion, nausea, and diarrhea. The current or previous comorbid diseases include diabetes mellitus, hypertension, heart failure, chronic heart disease, chronic obstructive pulmonary disease, chronic kidney disease, cancer, chronic liver disease, rheumatism or autoimmune disease, and dementia. The clinical severity has two categories: low and high severity. The conditions of patients in the low-severity group correspond to no limit of activity, oxygen support with nasal prong or facial mask, and noninvasive ventilation. The conditions of patients in the high-severity group correspond to invasive ventilation, multi-organ failure with extracorporeal membrane oxygenation required, and death. The general blood test results include levels of hemoglobin, hematocrit, lymphocytes, platelets, and white blood cells.

Out of 5628 COVID-19 patient records, the clinical severity information was missing in 27 patient records, so we excluded them from our study. Thus, we used 5601 patient data records to develop the AI prediction model for clinical severity. For each patient data record, we used 37 variables as model inputs; these variables are summarized in Table 1 without ID, outcome, quarantine period, and clinical severity. As the model output, we used clinical severity, which is a binary component composed of low and high severity.

Table 2 summarizes the clinical features from the high-severity group (271/5601, 4.8%) and the low-severity group (5330/5601, 95.2%). Notably, in the high-severity group, 241 out of 271 patients were deceased (88.9%), while no patients died in the low-severity group.

Table 1. Description of COVID-19 patient data.

Item category and data	Type	Description
Basic patient information		
ID	Number	Anonymous
Age (years)	9 categories	0-9 (0), 10-19 (1), 20-29 (2), 30-39 (3), 40-49 (4), 50-59 (5), 60-69 (6), 70-79 (7), ≥80 (8)
Gender	2 categories	Male (0), female (1)
Outcome	2 categories	Survived (0), deceased (1)
Quarantine period	Continuous	Days (0 if confirmed after death)
Pregnancy	2 categories	No (0), yes (1)
Pregnancy week	Number	Weeks (0 if not pregnant)
Physical index: BMI (kg/m^2)	5 categories	<18.5 (0), 18.5-22.9 (1), 23.0-24.9 (2), 25.0-29.9 (3), ≥30 (4)
Initial examination findings		
Systolic blood pressure	5 categories	<120 (0), 120-129 (1), 130-139 (2), 140-159 (3), ≥160 (4)
Diastolic blood pressure	4 categories	<80 (0), 80-89 (1), 90-99 (2), ≥100 (3)
Heart rate	Number	Heart rate
Temperature	Number	Temperature
Clinical findings		
Fever	2 categories	No (0), yes if higher than 37.5 °C (1)
Cough	2 categories	No (0), yes (1)
Sputum production	2 categories	No (0), yes (1)
Sore throat	2 categories	No (0), yes (1)
Runny nose or rhinorrhea	2 categories	No (0), yes (1)
Muscle aches or myalgia	2 categories	No (0), yes (1)
Fatigue or malaise	2 categories	No (0), yes (1)
Shortness of breath or dyspnea	2 categories	No (0), yes (1)
Headache	2 categories	No (0), yes (1)
Altered consciousness or confusion	2 categories	No (0), yes (1)
Vomiting or nausea	2 categories	No (0), yes (1)
Diarrhea	2 categories	No (0), yes (1)
Current or previous comorbid diseases		
Diabetes mellitus	2 categories	No (0), yes (1)
Hypertension	2 categories	No (0), yes (1)
Heart failure	2 categories	No (0), yes (1)
Chronic cardiac disease	2 categories	No (0), yes (1)
Asthma	2 categories	No (0), yes (1)
Chronic obstructive pulmonary disease	2 categories	No (0), yes (1)
Chronic kidney disease	2 categories	No (0), yes (1)
Cancer	2 categories	No (0), yes (1)
Chronic liver disease	2 categories	No (0), yes (1)
Rheumatism or autoimmune diseases	2 categories	No (0), yes (1)
Dementia	2 categories	No (0), yes (1)

Item category and data	Type	Description
Clinical severity	2 categories	Low severity, including no limit of activity, oxygen support required with nasal prong or facial mask, and noninvasive ventilation (0); high severity, including invasive ventilation, multi-organ failure, extracorporeal membrane oxygenation, and death (1)
General blood test results		
Hemoglobin	Number	g/dL
Hematocrit	Number	%
Lymphocytes	Number	%
Platelets	Number	10^9/L
White blood cells	Number	10^9/L

Table 2. Statistical summary of clinical features from the low-severity group and high-severity group (N=5601).

Participant data	Low-severity group (n=5330)	High-severity group (n=271)	*P* value
Basic patient information			
Age category[a], mean (SD)	4.26 (1.92)	7.05 (1.08)	<.001
Gender, n (%)			<.001
Male	2166 (40.6)	144 (53.1)	
Female	3164 (59.4)	127 (46.9)	
Pregnancy status (yes), n (%)	19 (0.4)	0 (0)	.33
Pregnancy week, mean (SD)	16.50 (10.01)	N/A[b]	N/A
Physical index: BMI category[c], mean (SD)	1.79 (1.02)	1.84 (1.13)	.54
Initial examination findings, mean (SD)			
Systolic blood pressure category[d]	1.75 (1.31)	1.98 (1.46)	.008
Diastolic blood pressure category[e]	1.00 (0.97)	0.90 (1.00)	.11
Heart rate (beats per minute)	85.66 (14.79)	89.05 (19.64)	<.001
Temperature (°C)	36.94 (0.54)	37.11 (0.80)	<.001
Clinical findings (low-severity group n=5326), n (%)			
Fever	1197 (22.5)	105 (38.7)	<.001
Cough	2239 (42.0)	92 (33.9)	.008
Sputum production	1532 (28.8)	79 (29.2)	.89
Sore throat	858 (16.1)	14 (5.2)	<.001
Runny nose or rhinorrhea	609 (11.4)	8 (3.0)	<.001
Muscle aches or myalgia	894 (16.8)	26 (9.6)	.002
Fatigue or malaise	215 (4.0)	18 (6.6)	.04
Shortness of breath or dyspnea	531 (10.0)	134 (49.4)	<.001
Headache	946 (17.8)	17 (6.3)	<.001
Altered consciousness or confusion	9 (0.2)	26 (9.6)	<.001
Vomiting or nausea	226 (4.2)	18 (6.6)	.06
Diarrhea	496 (9.3)	20 (7.4)	.28
Current or previous comorbid diseases, n (%)			
Diabetes mellitus	582/5327 (10.9)	106 (39.1)	<.001
Hypertension	1034/5327 (19.4)	164 (60.5)	<.001
Heart failure	39/5327 (0.7)	20 (7.4)	<.001
Chronic cardiac disease	150/5311 (2.8)	29 (10.7)	<.001
Asthma	115/5327 (2.2)	13 (4.8)	.005
Chronic obstructive pulmonary disease	31/5327 (0.6)	9 (3.3)	<.001
Chronic kidney disease	37/5327 (0.7)	18 (6.6)	<.001
Cancer	123/5326 (2.3)	22 (8.1)	<.001
Chronic liver disease	76/5004 (1.5)	7 (2.6)	.17
Rheumatism or autoimmune diseases	35/4998 (0.7)	3 (1.1)	.44
Dementia	148/5001 (3.0)	76 (28.0)	<.001
General blood test results, mean (SD)			
Hemoglobin (g/dL)	13.37 (1.69)	11.89 (2.23)	<.001
Hematocrit (%)	39.51 (4.72)	35.28 (6.56)	<.001

Participant data	Low-severity group (n=5330)	High-severity group (n=271)	*P* value
Lymphocytes (%)	30.08 (11.12)	15.08 (10.69)	<.001
Platelets (10^9/L)	239.96 (81.57)	188.51 (87.38)	<.001
White blood cells (10^9/L)	6.00 (2.55)	7.99 (5.10)	<.001

[a]Age categories were as follows (years): 0-9 (0), 10-19 (1), 20-29 (2), 30-39 (3), 40-49 (4), 50-59 (5), 60-69 (6), 70-79 (7), ≥80 (8).

[b]N/A: not applicable; there were no pregnant participants in the high-severity group.

[c]BMI categories were as follows (kg/m^2): <18.5 (0), 18.5-22.9 (1), 23.0-24.9 (2), 25.0-29.9 (3), ≥30 (4).

[d]Systolic blood pressure categories were as follows (mm Hg): <120 (0), 120-129 (1), 130-139 (2), 140-159 (3), ≥160 (4).

[e]Diastolic blood pressure categories were as follows (mm Hg): <80 (0), 80-89 (1), 90-99 (2), ≥100 (3).

Imputation and Standardization

In the data set, some features were missing (Table S1 in Multimedia Appendix 1). To handle the missing data, we calculated the mean value from the training data set for each feature and replaced the missing data with the mean value in both the training and testing data sets. We then performed standardization of the data set, which is a common requirement for machine learning algorithms. The standardization changes the data distribution of each feature with a mean of zero and standard deviation of one:

where *mean*(*train*) and *SD*(*train*) are the mean and standard deviation values, respectively, for each feature from the training data set. We applied the standardization to both the training and testing data sets.

Data Split

For the feature importance analysis and the AI prediction model development, we performed a grid search with a 5-fold cross-validation and 10-time repetition. For that, we divided the 5601 records into training (4480/5601, 80.0%) and testing (1121/5601, 20.0%) data sets in a stratified fashion (Table 3). We used 4480 records as the training data set (4260/4480, 95.1% low severity and 220/4480, 4.9% high severity) and 1121 records as the testing data set (1070/1121, 95.5% low severity and 51/1121, 4.5% high severity). The testing data set was isolated and used only for evaluating the performance of the proposed model.

The training data set (n=4480) was randomly shuffled and partitioned into 5 equal folds in a stratified manner: each fold included 433 low-severity records and 15 high-severity records. Of the 5 folds, a single fold was retained as the validation data set for testing the model, and the remaining 4 folds were used as the training data. We repeated the process 10 times, with each of the 10 folds used exactly once as the validation data. Here, since the number of low-severity records was much higher than the number of high-severity records, we up-sampled the high-severity data by randomly copying the data to prevent the model's bias toward the low-severity data by balancing the amount of data in the two groups.

Table 3. Summary of training and testing data sets.

Data set	Records, n (%)	
	Low-severity group	High-severity group
Training (n=4480)	4260 (95.1)	220 (4.9)
Testing (n=1121)	1070 (95.5)	51 (4.5)
Total (N=5601)	5330 (95.2)	271 (4.8)

Feature Selection

In order to select important features that influence clinical severity, we first investigated the contribution of each of the 37 input variables on severity via feature importance analysis using AdaBoost [19,20], random forest [21], and eXtreme Gradient Boosting (XGBoost) [22] algorithms. After analyzing the feature importance values from each classifier algorithm, we normalized and averaged the values to calculate the combination feature importance values.

By repeating the 5-fold cross-validation 10 times, we found the best hyperparameters. For AdaBoost, we set the hyperparameters as follows: the number of tree estimators was set to 50 and the learning rate was set to 0.4. For random forest, we set the number of tree estimators to 100, the maximum depth to 4, and the maximum features to 5. For XGBoost, we set the maximum depth to 2, the learning rate to 0.2, the number of tree estimators to 100, the value of the regularization parameter α to 1.0, the fraction of observations to 0.9, and the fraction of columns to 0.9.

The 10-time repeated 5-fold cross-validation provided 50 sets of feature importance values for each classifier (ie, AdaBoost, random forest, and XGBoost). We then averaged the 50 sets of importance values and normalized them so that the importance values from each classifier were in the range from 0 to 1. Finally, we averaged the importance values for the final ranked feature importance value. Moreover, we determined the optimal number of top features to incorporate into the AI prediction model based on the cross-validation results.

AI Prediction Model Development

To develop the final AI model for severity prediction, we used a deep neural network (DNN). In the DNN approach, we investigated up to 5 hidden layers and each layer depth (ie, node) up to the previous layer depth (ie, node). For the input layer, we first ranked the features according to their importance and increased the number of top features used in the input layer from 1 to 37. For the fully connected (FC) layers as hidden layers, we applied dropouts by changing the dropout rate from 0 to 0.5 with 0.1 increments. The last FC layer was fed into a sigmoid layer, which is an output layer providing the probabilities for the patient severity. We trained the models with the Adam optimizer and binary cross-entropy cost function with a learning rate of 0.0001 and batch size of 64. We implemented the models using R, version 4.0.2 (The R Foundation), with TensorFlow, version 1.13.1, for DNN; scikit-learn, version 0.22.1, for machine learning algorithms; and xgboost, version 0.6.4, for the XGBoost algorithm.

For each set of top features, we found the best cross-validation accuracy using the metrics of area under the curve (AUC) and balanced accuracy:

Given the cross-validation accuracy analysis, we finally modeled with the 5-layer DNN using the top 20 features. The 5-layer DNN comprised an input layer, 3 FC layers as hidden layers, and an output layer. The input layer was fed into a series of 3 FC layers consisting of 20, 16, and 8 nodes, respectively. In the first 2 FC layers, we used a dropout rate of 0.5. Then, the last FC layer was fed into a sigmoid layer.

Performance Evaluation

We evaluated the prediction performance of our proposed 5-layer DNN model with the isolated testing data set (n=1121). To compare the prediction performance of the DNN model with those of other external AI models, we separately trained the following models: logistic regression, decision tree, random forest, support vector machine, XGBoost, AdaBoost, GradBoost, and HistBoost. We evaluated the prediction performance of these AI models as single models as well as ensemble models.

Results

Feature Selection

Figure 1 shows the results of the ranked feature importance analysis from AdaBoost, random forest, XGBoost, and their combination. Results from AdaBoost indicate that platelet count had the highest importance value, followed by lymphocyte level, age, and body mass index (Figure 1, a). Results from random forest indicate that age had the highest importance value, followed by lymphocyte level, shortness of breath or dyspnea, and platelet count (Figure 1, b). Results from XGBoost indicate that platelet count had the highest importance value, followed by age, lymphocyte level, and temperature (Figure 1, c). By averaging the values obtained from the three models, age had the highest importance value, followed by lymphocyte level, platelet count, and shortness of breath or dyspnea (Figure 1, d). On the other hand, cancer, fatigue or malaise, chronic obstructive pulmonary disease, sputum production, chronic cardiac disease, heart failure, asthma, rheumatism or autoimmune diseases, pregnancy, and pregnancy week rarely contributed to the predictive model. The normalized feature importance values from AdaBoost, random forest, and XGBoost, as well as the combined ranked feature importance values with those averages, are summarized in Table S2 in Multimedia Appendix 1.

We investigated the cross-validation performance with the metrics of AUC and balanced accuracy (Figure 2). The results show that both AUC and balanced accuracy reached the highest values when the top 20 features from the combination of AdaBoost, random forest, and XGBoost were used for the input layer. Therefore, we incorporated the top 20 features into the AI prediction model, which yielded a sensitivity of 88%, specificity of 90%, accuracy of 90%, balanced accuracy of 89%, and AUC of 0.96 (Table 4).

Figure 1. Results of normalized feature importance analysis from (a) AdaBoost, (b) random forest, and (c) eXtreme Gradient Boosting (XGBoost) as well as (d) the combined average ranked feature importance. ACC: altered consciousness/confusion; BMI: body mass index; CCD: chronic cardiac disease; CKD: chronic kidney disease; CLD: chronic liver disease; COPD: chronic obstructive pulmonary disease; DBP: diastolic blood pressure; DEMEN: dementia; DIARR: diarrhea; DM: diabetes mellitus; FM: fatigue/malaise; HCT: hematocrit; HEADA: headache; HF: heart failure; HGB: hemoglobin; HR: heart rate; HTN: hypertension; LYMPHO: lymphocyte; MAM: muscle aches/myalgia; PLT: platelets; Preg: pregnancy; PregWk: pregnancy weeks; RDAD: rheumatism/autoimmune disease; RNR: runny nose/rhinorrhea; SBP: systolic blood pressure; SOB: shortness of breath/dyspnea; SPUTUM: sputum production; ST: sore throat; Temp: temperature; VN: vomiting/nausea; WBC: white blood cells.

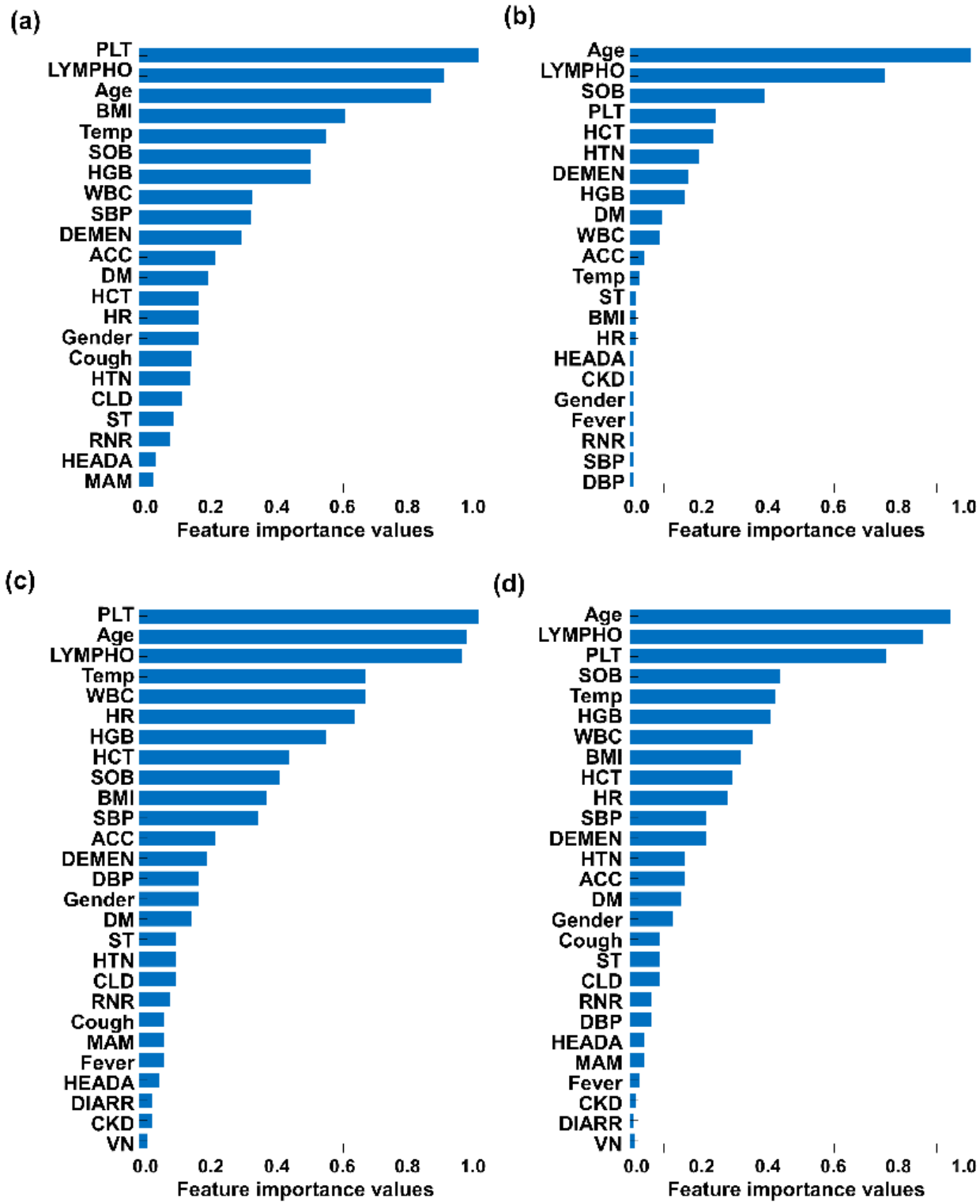


Figure 2. The influence of feature importance values on cross-validation accuracy. AUC: area under the curve.

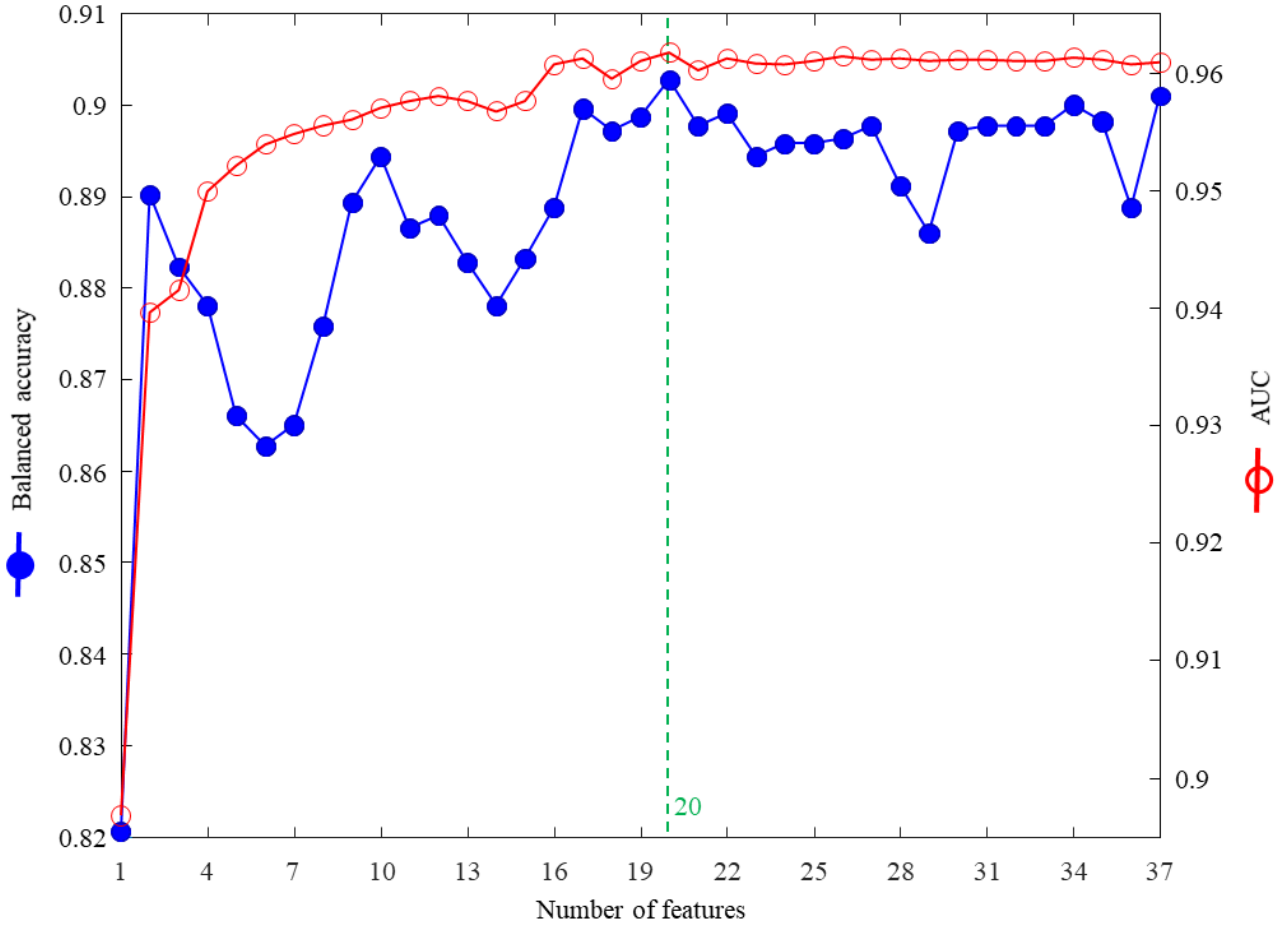


Table 4. Cross-validation results.

Model	Cross-validation measures (n=448), mean (SD)				
	Sensitivity	Specificity	Accuracy	Balanced accuracy	Area under the curve
5-layer deep neural network	0.88 (0.06)	0.90 (0.02)	0.90 (0.02)	0.89 (0.04)	0.96 (0.01)

Performance of the AI Prediction Model

With the isolated testing data set (n=1121), our proposed 5-layer DNN showed a sensitivity of 90.20%, specificity of 90.37%, accuracy of 90.37%, balanced accuracy of 90.28%, and AUC of 0.96. Table 5 shows the prediction performances on the testing data set. First, we compared the accuracy metrics when the synthetic minority oversampling technique was applied, and we found that both balanced accuracy and AUC were slightly lower. Second, we compared the accuracy metrics when principal component analysis (PCA)–based feature reduction was applied with eight dimensions, and we found that both balanced accuracy and AUC were also slightly lower. Table 5 [19-28] also shows the prediction performances of various AI models; it can be seen that our proposed 5-layer DNN method provided higher accuracy, balanced accuracy, and AUC values than the other external AI models (ie, logistic regression, decision tree, random forest, support vector machine, XGBoost, AdaBoost, GradBoost, and HistBoost).

Furthermore, we investigated the prediction performance of ensemble AI models (ie, combination of AI models); none of the ensemble AI models outperformed our proposed 5-layer DNN model (Table 6).

Table 5. Testing data results and comparison with other machine learning algorithms.

Model	TN[a]	FP[b]	FN[c]	TP[d]	Sen[e]	Spe[f]	Acc[g]	BA[h]	AUC[i]
5-layer DNN[j]: copying	967	103	5	46	0.9020	0.9037	0.9037	0.9028	0.9617
5-layer DNN: SMOTE[k] [23]	984	86	8	43	0.8431	0.9196	0.9161	0.8814	0.9555
5-layer DNN with PCA[l] (8 features)	922	148	5	46	0.9020	0.8617	0.8635	0.8818	0.9549
Linear regression [24]	983	87	7	44	0.8627	0.9187	0.9161	0.8907	0.9563
Decision tree [25]	915	155	5	46	0.9020	0.8551	0.8573	0.8786	0.9252
Random forest [21]	955	115	5	46	0.9020	0.8925	0.8930	0.8972	0.9590
Support vector machine [26]	955	115	5	46	0.9020	0.8925	0.8930	0.8972	0.9588
XGBoost[m] [22]	945	125	6	45	0.8824	0.8832	0.8831	0.8828	0.9558
AdaBoost [19,20]	937	133	5	46	0.9020	0.8757	0.8769	0.8888	0.9586
GradBoost [27]	936	134	6	45	0.8824	0.8748	0.8751	0.8786	0.9525
HistBoost [28]	959	111	7	44	0.8627	0.8963	0.8947	0.8795	0.9535

[a]TN: true negative.
[b]FP: false positive.
[c]FN: false negative.
[d]TP: true positive.
[e]Sen: sensitivity.
[f]Spe: specificity.
[g]Acc: accuracy.
[h]BA: balanced accuracy.
[i]AUC: area under the curve.
[j]DNN: deep neural network.
[k]SMOTE: synthetic minority oversampling technique.
[l]PCA: principal component analysis.
[m]XGBoost: eXtreme Gradient Boosting.

Table 6. Test result comparison with ensemble approaches.

Model	TN[a]	FP[b]	FN[c]	TP[d]	Sen[e]	Spe[f]	Acc[g]	BA[h]	AUC[i]
5-layer deep neural network (DNN) (proposed)	967	103	5	46	0.9020	0.9037	0.9037	0.9028	0.9617
DNN + linear regression (LR)	976	94	6	45	0.8824	0.9121	0.9108	0.8973	0.9589
DNN + random forest (RF)	967	103	5	46	0.9020	0.9037	0.9037	0.9028	0.9572
DNN + AdaBoost	965	105	5	46	0.9020	0.9019	0.9019	0.9019	0.9607
DNN + eXtreme Gradient Boosting (XGBoost)	963	107	6	45	0.8824	0.9000	0.8992	0.8912	0.9490
DNN + support vector machine (SVM)	962	108	5	46	0.9020	0.8991	0.8992	0.9005	0.9563
RF + AdaBoost	954	116	5	46	0.9020	0.8916	0.8921	0.8968	0.9515
DNN + RF + AdaBoost	967	103	5	46	0.9020	0.9037	0.9037	0.9028	0.9579
DNN + RF + SVM	962	108	5	46	0.9020	0.8991	0.8992	0.9005	0.9556
DNN + RF + LR	963	107	5	46	0.9020	0.9000	0.9001	0.9010	0.9585
DNN + RF + AdaBoost + XGBoost	944	126	5	46	0.9020	0.8822	0.8831	0.8921	0.9571
DNN + RF + AdaBoost + SVM	959	111	5	46	0.9020	0.8963	0.8965	0.8991	0.9562
DNN + RF + AdaBoost + XGBoost + SVM	978	92	6	45	0.8824	0.9140	0.9126	0.8982	0.9572

[a]TN: true negative.

[b]FP: false positive.

[c]FN: false negative.

[d]TP: true positive.

[e]Sen: sensitivity.

[f]Spe: specificity.

[g]Acc: accuracy.

[h]BA: balanced accuracy.

[i]AUC: area under the curve.

Discussion

Principal Findings

Our proposed AI model, the 5-layer DNN using the selected top 20 features, was able to predict the severity of COVID-19 patients at the hospital admission stage with excellent prediction performance: 90.2% sensitivity, 90.4% specificity, and 90.4% accuracy. The model has several unique characteristics. First, it was developed based on nationwide confirmed COVID-19 patient data obtained from the KDCA. In South Korea, all confirmed cases must be reported to the KDCA; thus, the KDCA data are very accurate and updated on a daily basis [4]. The Korean government designated more than 100 general hospitals, including 20 tertiary hospitals, as specialized infection control hospitals equipped with isolation and negative pressure rooms. These designated hospitals should report important clinical information about COVID-19 patients to the KDCA, especially for patients who are admitted to hospitals or show severe conditions. When we were allowed to access the KDCA data sets in September 2020, there were data from 5601 patients with comprehensive clinical information that we could use to develop an AI prediction model. This is the largest cohort with a sufficient amount of data to develop reliable and generalizable AI prediction models.

Second, our AI prediction model development started with feature importance analysis of the 37 features in the comprehensive data set. Of these, 20 were selected based on ranked feature importance analysis results in order to develop an accurate AI prediction model. The cross-validation demonstrated that the AI prediction model showed higher accuracy using the selected 20 features compared to using all 37 features. In addition, the selected 20 features (ie, age, lymphocyte level, platelet count, shortness of breath or dyspnea, temperature, hemoglobin level, white blood cell count, body mass index, hematocrit level, heart rate, systolic blood pressure, dementia, hypertension, altered consciousness or confusion, diabetes, gender, cough, sore throat, chronic liver disease, and runny nose or rhinorrhea) can be easily acquired from patient history, basic physical examinations, and routine laboratory tests. Thus, our AI prediction model can be easily incorporated into routine clinical practice. Furthermore, we observed that PCA-based feature selection also provided as good of a performance as did the feature importance analysis. In particular, we expect that many researchers will be able to flexibly diversify the model for predicting the severity of COVID-19 patients, in that a similar level of accuracy could be obtained with only eight features.

In terms of our feature selection process, we combined AdaBoost, random forest, and XGBoost machine learning algorithms to rank the important features. The AdaBoost algorithm is part of the family of boosting algorithms and sequentially growing decision trees as weak learners [19]. It is

well known that it rarely overfits in low-noise data sets [20]. The random forest algorithm is based on a bagging approach, which is based on the aggregation of a set of weak learners [21]. XGBoost is a recently introduced algorithm with optimized gradient boosting [22]. In low-dimensional or highly separable data, all of the classifiers generally provide reasonably good performance. However, they may provide different performances depending on various factors, such as feature dimension, data separability, data balancing, and feature correlation. That is the reason we have combined the three algorithm results.

We named our proposed AI prediction model KOVIDnet, indicating the deep learning algorithm for Korean COVID-19 patients. Owing to its high accuracy and generalizability in Korea, we expect that KOVIDnet will be able to provide treatment priority guidance at the time of admission regarding who should be treated intensively. Although most patients with COVID-19 showed mild and self-limiting illness, some patients became severely and critically ill, showing the rapid progression to acute respiratory failure, sepsis, septic shock, multi-organ failure, and eventual death [29-32]. The mortality rate of severe cases is about 20 times higher than that of mild cases [30,33]. This indicates that early identification of patients at risk of mortality is important for the management of COVID-19 patients.

Limitations and Future Work

In our earlier study, we grouped the patients into eight subgroups. Subgroup 1 patients had no limits to their activity. Subgroup 2 patients had limits to their activity, but did not need oxygen. Subgroup 3 patients needed oxygen with a nasal prong. Subgroup 4 patients needed oxygen with a facial mask. Subgroup 5 patients needed noninvasive ventilation. Subgroup 6 patients needed invasive ventilation. Subgroup 7 patients had multi-organ failure or underwent extracorporeal membrane oxygenation. Subgroup 8 patients died. For the multiclass classification, we also trained the model through the same procedure as mentioned above, but the accuracy when using the testing data was not satisfactory (Table S3 in Multimedia Appendix 1). It may be because there was no distinct difference in features according to each subgroup or because the number of training data values was insufficient. In addition, the data were extremely imbalanced: the imbalance ratio was 405 (Table S4 in Multimedia Appendix 1). After analyzing the results from the eight-subgroup multiclass classification problem, we considered the binary classification problem, where the low-severity group included subgroups 1 to 4 and the high-severity group included subgroups 5 to 8. Not only was the binary classification problem the most realistic in training the predictive model based on our current data, but it was also useful to convey clinically important implications. We believe that we can extend our model to the multiclass classification problem based on more extensive data.

Our study also has additional limitations. First, our proposed AI prediction model was validated with an isolated test data set (n=1121), which was a data set that was split from the entire data set. It may be necessary to validate our AI model with external data sets, such as prospectively collected data. To validate and update KOVIDnet, we made a web application [34] so that anyone can access the model. We believe that sharing the AI model with the public will be helpful in validating and improving its performance. Second, our data did not include patients of other races, such as Caucasian or Middle East Asian. In the near future, we have a plan to apply our AI model to various data sets, including data from patients of other races. To realize this goal, we will establish a real-time training framework that can train our model using prospectively collected data from all over the world. We believe that we can improve KOVIDnet for better generalization based on the extended data.

Conclusions

In conclusion, we developed our AI model with 20 selected features based on a large nationwide data set, and it was able to predict the severity of COVID-19 accurately. We believe that our model can help health care providers to effectively treat COVID-19 patients at an early stage and ultimately reduce deaths.

Acknowledgments

We acknowledge all the health care workers involved in the diagnosis and treatment of COVID-19 patients in South Korea. We also thank the KDCA, the National Medical Center, and the health information managers at the hospitals for their efforts in collecting the medical records. This study was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF), funded by the Ministry of Science, ICT, and Future Planning (grants 202012B04 and NRF-2020R1A2C1014829), and by the Korea Health Industry Development Institute (grant HI18C1216).

Authors' Contributions

HC and HK carried out the machine learning and deep learning simulation for hyperparameter search and modeling. CP performed data validation to be applied to COVID-19 patients. KWK, H-OS, T-YC, and JHS validated and confirmed the simulations and helped to draft the manuscript. HL developed and maintained the web application. JL and WSK conceived of the study, participated in the study’s design and coordination, and wrote the initial manuscript. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Supplementary tables.
[DOCX File , 54 KB - jmir_v23i4e27060_app1.docx]

References

1. Honein MA, Christie A, Rose DA, Brooks JT, Meaney-Delman D, Cohn A, CDC COVID-19 Response Team. Summary of guidance for public health strategies to address high levels of community transmission of SARS-CoV-2 and related deaths, December 2020. MMWR Morb Mortal Wkly Rep 2020 Dec 11;69(49):1860-1867 [FREE Full text] [doi: 10.15585/mmwr.mm6949e2] [Medline: 33301434]
2. Zhu J, Shen B, Abbasi A, Hoshmand-Kochi M, Li H, Duong TQ. Deep transfer learning artificial intelligence accurately stages COVID-19 lung disease severity on portable chest radiographs. PLoS One 2020;15(7):e0236621 [FREE Full text] [doi: 10.1371/journal.pone.0236621] [Medline: 32722697]
3. Shan F, Gao Y, Wang J, Shi W, Shi N, Han M, et al. Abnormal lung quantification in chest CT images of COVID-19 patients with deep learning and its application to severity prediction. Med Phys 2020 Nov 22:1-13 [FREE Full text] [doi: 10.1002/mp.14609] [Medline: 33225476]
4. Ko H, Chung H, Kang WS, Kim KW, Shin Y, Kang SJ, et al. COVID-19 pneumonia diagnosis using a simple 2D deep learning framework with a single chest CT image: Model development and validation. J Med Internet Res 2020 Jun 29;22(6):e19569 [FREE Full text] [doi: 10.2196/19569] [Medline: 32568730]
5. Rajaraman S, Siegelman J, Alderson PO, Folio LS, Folio LR, Antani SK. Iteratively pruned deep learning ensembles for COVID-19 detection in chest x-rays. IEEE Access 2020;8:115041-115050 [FREE Full text] [doi: 10.1109/access.2020.3003810] [Medline: 32742893]
6. Das D, Santosh KC, Pal U. Truncated inception net: COVID-19 outbreak screening using chest x-rays. Phys Eng Sci Med 2020 Sep;43(3):915-925 [FREE Full text] [doi: 10.1007/s13246-020-00888-x] [Medline: 32588200]
7. Brunese L, Martinelli F, Mercaldo F, Santone A. Machine learning for coronavirus COVID-19 detection from chest x-rays. Procedia Comput Sci 2020;176:2212-2221 [FREE Full text] [doi: 10.1016/j.procs.2020.09.258] [Medline: 33042308]
8. Yao H, Zhang N, Zhang R, Duan M, Xie T, Pan J, et al. Severity detection for the coronavirus disease 2019 (COVID-19) patients using a machine learning model based on the blood and urine tests. Front Cell Dev Biol 2020;8:683 [FREE Full text] [doi: 10.3389/fcell.2020.00683] [Medline: 32850809]
9. Brinati D, Campagner A, Ferrari D, Locatelli M, Banfi G, Cabitza F. Detection of COVID-19 infection from routine blood exams with machine learning: A feasibility study. J Med Syst 2020 Jul 01;44(8):135 [FREE Full text] [doi: 10.1007/s10916-020-01597-4] [Medline: 32607737]
10. Liang W, Yao J, Chen A, Lv Q, Zanin M, Liu J, et al. Early triage of critically ill COVID-19 patients using deep learning. Nat Commun 2020 Jul 15;11(1):3543 [FREE Full text] [doi: 10.1038/s41467-020-17280-8] [Medline: 32669540]
11. Izquierdo JL, Ancochea J, Savana COVID-19 Research Group, Soriano JB. Clinical characteristics and prognostic factors for intensive care unit admission of patients with COVID-19: Retrospective study using machine learning and natural language processing. J Med Internet Res 2020 Oct 28;22(10):e21801 [FREE Full text] [doi: 10.2196/21801] [Medline: 33090964]
12. An C, Lim H, Kim D, Chang JH, Choi YJ, Kim SW. Machine learning prediction for mortality of patients diagnosed with COVID-19: A nationwide Korean cohort study. Sci Rep 2020 Oct 30;10(1):18716 [FREE Full text] [doi: 10.1038/s41598-020-75767-2] [Medline: 33127965]
13. Li Y, Horowitz MA, Liu J, Chew A, Lan H, Liu Q, et al. Individual-level fatality prediction of COVID-19 patients using AI methods. Front Public Health 2020;8:587937 [FREE Full text] [doi: 10.3389/fpubh.2020.587937] [Medline: 33102426]
14. Vaid A, Somani S, Russak AJ, De Freitas JK, Chaudhry FF, Paranjpe I, et al. Machine learning to predict mortality and critical events in a cohort of patients with COVID-19 in New York City: Model development and validation. J Med Internet Res 2020 Nov 06;22(11):e24018 [FREE Full text] [doi: 10.2196/24018] [Medline: 33027032]
15. Yan L, Zhang H, Goncalves J, Xiao Y, Wang M, Guo Y, et al. An interpretable mortality prediction model for COVID-19 patients. Nat Mach Intell 2020 May 14;2(5):283-288. [doi: 10.1038/s42256-020-0180-7]
16. Shang W, Dong J, Ren Y, Tian M, Li W, Hu J, et al. The value of clinical parameters in predicting the severity of COVID-19. J Med Virol 2020 Oct;92(10):2188-2192 [FREE Full text] [doi: 10.1002/jmv.26031] [Medline: 32436996]
17. Li X, Xu S, Yu M, Wang K, Tao Y, Zhou Y, et al. Risk factors for severity and mortality in adult COVID-19 inpatients in Wuhan. J Allergy Clin Immunol 2020 Jul;146(1):110-118 [FREE Full text] [doi: 10.1016/j.jaci.2020.04.006] [Medline: 32294485]
18. Zhang J, Wang X, Jia X, Li J, Hu K, Chen G, et al. Risk factors for disease severity, unimprovement, and mortality in COVID-19 patients in Wuhan, China. Clin Microbiol Infect 2020 Jun;26(6):767-772 [FREE Full text] [doi: 10.1016/j.cmi.2020.04.012] [Medline: 32304745]
19. Freund Y, Schapire R. Game theory, on-line prediction and boosting. In: Proceedings of the 9th Annual Conference on Computational Learning Theory.: Association for Computing Machinery; 1996 Presented at: 9th Annual Conference on Computational Learning Theory; June 28-July 1, 1996; Desenzano del Garda, Italy p. 325-332. [doi: 10.1145/238061.238163]
20. Rätsch G, Onoda T, Müller KR. Soft margins for AdaBoost. Mach Learn 2001;42:287-320. [doi: 10.1023/A:1007618119488]
21. Breiman L. Random forests. Mach Learn 2001;45:5-32. [doi: 10.1023/A:1010933404324]

22. Chen T, Guestrin C. XGBoost: A scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.: Association for Computing Machinery; 2016 Presented at: 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; August 13-17, 2016; San Francisco, CA p. 785-794. [doi: 10.1145/2939672.2939785]
23. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: Synthetic minority over-sampling technique. J Artif Intell Res 2002 Jun 01;16:321-357. [doi: 10.1613/jair.953]
24. Cox DR. The regression analysis of binary sequences. J R Stat Soc Series B Stat Methodol 2018 Dec 05;20(2):215-232. [doi: 10.1111/j.2517-6161.1958.tb00292.x]
25. Kamiński B, Jakubczyk M, Szufel P. A framework for sensitivity analysis of decision trees. Cent Eur J Oper Res 2018;26(1):135-159 [FREE Full text] [doi: 10.1007/s10100-017-0479-6] [Medline: 29375266]
26. Cortes C, Vapnik V. Support-vector networks. Mach Learn 1995 Sep;20(3):273-297. [doi: 10.1007/bf00994018]
27. Natekin A, Knoll A. Gradient boosting machines, a tutorial. Front Neurorobot 2013;7:21. [doi: 10.3389/fnbot.2013.00021] [Medline: 24409142]
28. Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, et al. LightGBM: A highly efficient gradient boosting decision tree. In: Proceedings of the 31st Annual Conference on Neural Information Processing Systems (NIPS). 2017 Presented at: 31st Annual Conference on Neural Information Processing Systems (NIPS); December 4-9, 2017; Long Beach, CA p. 3146-3154 URL: https://papers.nips.cc/paper/2017/file/6449f44a102fde848669bdd9eb6b76fa-Paper.pdf
29. Epidemiology Working Group for NCIP Epidemic Response, Chinese Center for Disease Control and Prevention. The epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (COVID-19) in China [Article in Chinese]. Zhonghua Liu Xing Bing Xue Za Zhi 2020 Feb 10;41(2):145-151. [doi: 10.3760/cma.j.issn.0254-6450.2020.02.003] [Medline: 32064853]
30. Chen N, Zhou M, Dong X, Qu J, Gong F, Han Y, et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: A descriptive study. Lancet 2020 Feb 15;395(10223):507-513 [FREE Full text] [doi: 10.1016/S0140-6736(20)30211-7] [Medline: 32007143]
31. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet 2020 Feb 15;395(10223):497-506 [FREE Full text] [doi: 10.1016/S0140-6736(20)30183-5] [Medline: 31986264]
32. Wu Z, McGoogan JM. Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China: Summary of a report of 72 314 cases from the Chinese Center for Disease Control and Prevention. JAMA 2020 Apr 07;323(13):1239-1242. [doi: 10.1001/jama.2020.2648] [Medline: 32091533]
33. Shahriarirad R, Khodamoradi Z, Erfani A, Hosseinpour H, Ranjbar K, Emami Y, et al. Epidemiological and clinical features of 2019 novel coronavirus diseases (COVID-19) in the South of Iran. BMC Infect Dis 2020 Jun 18;20(1):427 [FREE Full text] [doi: 10.1186/s12879-020-05128-x] [Medline: 32552751]
34. KOVIDnet. URL: http://kcovidnet.site/ [accessed 2021-04-02]

Abbreviations

AI: artificial intelligence
AUC: area under curve
DNN: deep neural network
FC: fully connected
KDCA: Korea Disease Control and Prevention Agency
NRF: National Research Foundation of Korea
PCA: principal component analysis
XGBoost: eXtreme Gradient Boosting

Edited by R Kukafka; submitted 09.01.21; peer-reviewed by R Jaimies, C Jeong, J Lee, D Cha; comments to author 31.01.21; revised version received 18.02.21; accepted 24.03.21; published 19.04.21.

Please cite as:
Chung H, Ko H, Kang WS, Kim KW, Lee H, Park C, Song HO, Choi TY, Seo JH, Lee J
Prediction and Feature Importance Analysis for Severity of COVID-19 in South Korea Using Artificial Intelligence: Model Development and Validation
J Med Internet Res 2021;23(4):e27060
URL: https://www.jmir.org/2021/4/e27060
doi:10.2196/27060
PMID:33764883

Original Paper

Measuring Stress in Health Professionals Over the Phone Using Automatic Speech Analysis During the COVID-19 Pandemic: Observational Pilot Study

Alexandra König[1], PhD; Kevin Riviere[2], MD; Nicklas Linz[3], MSc; Hali Lindsay[4], MSc; Julia Elbaum[2], MD; Roxane Fabre[2], MSc; Alexandre Derreumaux[5], MSc; Philippe Robert[5], MD, PhD

[1]Stars Team, Institut national de recherche en informatique et en automatique, Valbonne, France

[2]Département de Santé Publique, Centre Hospitalier Universitaire de Nice, Université Côte d'Azur, Nice, France

[3]ki elements, Saarbrücken, Germany

[4]German Research Center for Artificial Intelligence (DFKI), Saarbrücken, Germany

[5]CoBteK (Cognition-Behaviour-Technology) Lab, La Fédération de Recherche Interventions en Santé, Université Côte d'Azur, Nice, France

Corresponding Author:
Alexandra König, PhD
Stars Team
Institut national de recherche en informatique et en automatique
2004 Route des Lucioles, 06902
Sophia Antipolis
Valbonne, 06200
France
Phone: 33 +33652021156
Email: alexandra.konig@inria.fr

Abstract

Background: During the COVID-19 pandemic, health professionals have been directly confronted with the suffering of patients and their families. By making them main actors in the management of this health crisis, they have been exposed to various psychosocial risks (stress, trauma, fatigue, etc). Paradoxically, stress-related symptoms are often underreported in this vulnerable population but are potentially detectable through passive monitoring of changes in speech behavior.

Objective: This study aims to investigate the use of rapid and remote measures of stress levels in health professionals working during the COVID-19 outbreak. This was done through the analysis of participants' speech behavior during a short phone call conversation and, in particular, via positive, negative, and neutral storytelling tasks.

Methods: Speech samples from 89 health care professionals were collected over the phone during positive, negative, and neutral storytelling tasks; various voice features were extracted and compared with classical stress measures via standard questionnaires. Additionally, a regression analysis was performed.

Results: Certain speech characteristics correlated with stress levels in both genders; mainly, spectral (ie, formant) features, such as the mel-frequency cepstral coefficient, and prosodic characteristics, such as the fundamental frequency, appeared to be sensitive to stress. Overall, for both male and female participants, using vocal features from the positive tasks for regression yielded the most accurate prediction results of stress scores (mean absolute error 5.31).

Conclusions: Automatic speech analysis could help with early detection of subtle signs of stress in vulnerable populations over the phone. By combining the use of this technology with timely intervention strategies, it could contribute to the prevention of burnout and the development of comorbidities, such as depression or anxiety.

(J Med Internet Res 2021;23(4):e24191) doi:10.2196/24191

KEYWORDS

stress detection; speech; voice analysis; COVID-19; phone monitoring; computer linguistics

Introduction

In December 2019 in the Chinese city of Wuhan, a new coronavirus pneumonia, COVID-19, emerged. The pathogen involved is SARS-CoV-2. Here, we will refer to the pathology as COVID-19. COVID-19 has spread very rapidly in China but also in many other countries [1]. On March 11, 2020, the World Health Organization declared that the COVID-19 outbreak had become a pandemic [2].

According to previous studies on SARS or Ebola epidemics, the onset of a sudden and immediately fatal disease could put extraordinary pressure on health care professionals [3]. Increased workloads, physical exhaustion, inadequate personal equipment, nosocomial transmission, and the need to make ethically difficult decisions about rationing care can have dramatic effects on their physical and mental well-being. Their resilience may be further compromised by isolation and loss of social support, risk or loss of friends and relatives, and radical, often worrying changes in working methods. Health care workers are, therefore, particularly vulnerable to mental health problems, including fear, anxiety, depression, and insomnia [4,5]. Initial results estimate that 23% and 22% of health care workers experienced depression and anxiety, respectively, during the COVID-19 pandemic [6].

Paradoxically, health care workers do not tend to seek professional help, and stress-related symptoms are often not immediately reported: “burnout, stress, and anxiety will have to wait.” Most of the time there will not even be a demand for care. Early implicit stress detection is of great importance in this population and would allow for timely intervention strategies in order to prevent escalation and complete occupational burnout.

To measure stress in clinical practice, various scales and questionnaires are available, such as the Perceived Stress Scale (PSS) [7], the Stressful Life Event Questionnaire [8], the Stress Overload Scale [9], and the Trier Inventory for Chronic Stress [10]. However, the present health crisis pushed research teams to investigate the use of new technological tools in this specific population. One possible avenue is the use of automatic speech analysis allowing extraction of voice features during standard consultation or over a simple phone call.

Psychological stress induces multiple effects on the body, including increased muscle tension, increased breathing rate, and changes in salivation rate, which may, in turn, affect vocal production [11,12]. Under psychological stress, voice pitch (ie, the acoustic correlate of fundamental frequency [F0]) usually increases, as it is inversely related to the rate of vocal fold vibration, which stretches under stress and becomes tenser together with an increase in subglottal pressure and vocal intensity [13,14]. Indeed, an increase in voice pitch is the most commonly reported finding in studies examining speech under stress. However, stress can also affect other voice parameters, such as an increase in speech prosody [11,13]. In depression, the analysis of speech characteristics has recently attracted considerable research attention [15-17]. Studies revealed that patients show flattened affect, reduced speech variability, monotonicity in pitch and loudness, increased pause duration, and reduced speech rate [18-20]. A recent study investigated the use of speech parameters extracted from audio recordings to differentiate patients suffering from posttraumatic stress disorder from healthy controls [21].

Thus, the detection of subtle events in the voice may offer a window into assessing the impact of stress in situations where circumstances make it difficult to monitor stress directly but need to be addressed urgently [22].

In this work, we aim to investigate the use of a rapid and remote measure of stress levels in health professionals working during the COVID-19 outbreak, utilizing the automatic analysis of their speech behavior during a short phone call conversation.

Firstly, speech samples of health care professionals were collected over the phone during the COVID-19 pandemic, and various voice features were extracted and compared with classical stress measures. Secondly, based on the extracted features, scores from the completed stress scale that were obtained by participants were predicted. The purpose of this pilot study was to assess whether this technological method could be of interest to support early screening of subtle signs of stress.

Methods

Participants

Health care professionals were recruited through outreach telephone calls. They worked during the COVID-19 outbreak in the local university hospital center of Nice, France, in either private practices or as independent workers in the Provence-Alpes-Côte d’Azur region. They could occupy any function in these structures. The only criterion for noninclusion was the subjects’ refusal to participate in the study. Inclusion of participants was carried out from May 5 to June 7, 2020.

The study was approved by the Ethical Board for noninterventional studies of the University Côte d’Azur, France (approval 2020-58). Participants were given all the information about the study prior to the call so they could give informed consent. For those interested, the option for a follow-up call with a clinician was provided.

Procedure

The telephone calls were made by psychiatrists (n=3) or psychologists (n=1) belonging to the Cognition Behavior Technology research team and the memory clinic of the University Côte d’Azur. Calls lasted about 15 minutes and were composed of the following:

1. An informative part explaining the reasons for the call and its structure and how the study is conducted. The participant’s consent was requested to continue and to proceed with a recording of his or her voice.
2. The Motivation Stress Affect (MSA) questionnaire. The MSA questionnaire is a self-administered questionnaire composed of 11 questions that must be answered by “yes” or “no.” The first five questions assess motivation [23], the next two questions assess depression, and the last four questions assess stress [24].

3. Three open standardized questions: neutral, positive, and negative storytelling. In order to capture natural speech, but within a limited time frame, the participant was asked to talk about something emotionally neutral (ie, describe where he or she is), to talk about a negative event in his or her life, and, finally, to talk about a positive event in his or her life. Each answer should have lasted about 1 minute and was recorded in a secure and encrypted way. It was not specified whether the event had to be experienced during COVID-19; thus, it was open to the participant to recall whatever event first came to mind. These free-speech tasks were used in previous studies [18,25] and allowed for a greater range of induced emotional effects, potentially sensitive to signs of stress and depression. The comparison of speech features between neutral and emotionally loaded questions may give insight into the affective state of participants.
4. The PSS. This scale [7] is a hetero-questionnaire composed of 10 questions to be answered by "never," "almost never," "sometimes," "quite often," or "often."
5. An open listening part aimed at exploring certain points in greater depth in order to refine the clinical needs.
6. Decision and advice. Following the above steps, the psychiatrist or psychologist offered or did not offer psychological follow-up depending on whether he or she considered that the patient was at risk of developing or had a mood or anxiety disorder. He or she may also have offered advice on intervention strategies (eg, relaxation, yoga, physical activity, and national call platform for psychological support for caregivers).

Materials

To perform the phone calls for this study, the phone version of the DELTA application [26] was used. The DELTA solution allows for the use of a dedicated interface in the form of an iOS app to make phone calls and locally record these calls on the internal memory of an iPad. The phone calls were made directly with the iPad and through its internal microphone.

These recordings were then automatically transmitted—the iPad had to be connected to the internet—to the DELTA application programming interface (API) for analysis of acoustic and semantic parameters. Once the analysis was complete, the results were displayed directly on the DELTA interface. The recordings were made locally on the phone, the connection between the interface and the DELTA API was secure and encrypted, and the recordings were destroyed from the DELTA servers once the analysis was complete and the results sent to the experimenter.

Analysis

Audio features were extracted directly and automatically from the recorded audio signals of the three open standardized questions (see item #3 in the Procedure section). Characteristics were extracted from four main areas:

1. Prosodic characteristics, on long-term variations in perceived stress and speech rhythm. Prosodic features also measure alterations in personal speech style (eg, perceived pitch and speech intonation).
2. Formant characteristics represent the dominant components of the speech spectrum and convey information about the acoustic resonance of the vocal tract and its use. These markers are often indicative of articulatory coordination problems in motor speech control disorders.
3. Source characteristics that are related to the source of voice production, the airflow through the glottal speech production system. These features make operational irregularities in the movement of the vocal fold (eg, voice quality measurements).
4. Temporal characteristics include measures of the proportion of speech (eg, duration of pauses and duration of speech segments), speech segment connectivity, and overall speech rate.

Features were extracted using Python 3.7 (Python Software Foundation) [27] and free and publicly available packages. For the temporal features, the My-Voice Analysis [28] package was used. This package was built off of the speech analysis research tool praat [29]. Temporal features were actualized as the speech rate, syllable count, rate of articulation, speaking duration, total duration, and ratio of speaking to nonspeaking. This package was also used to extract prosodic features, namely the F0 values: mean, standard deviation, minimum, maximum, and upper and lower quartiles. The F0 value is the representation of what is known as the pitch.

Formant features were calculated using the Python Speech Features library [30]. To characterize this aspect of speech, the original sound recording was refit according to a series of transformations commonly used for speech recognition that yield a better representation of the sound called the mel-frequency cepstrum (MFC). From this new representation of the sound form, the first 14 coefficients of the MFC were extracted. The MFC values were extracted given that they describe the spectral shape of the audio file, generally with diminishing returns in terms of how informative they are, which is why we only considered the first 14 coefficients. If we were to select a greater number of MFC values, it would result in a potentially needlessly more complex machine learning model using less informative features.

From each of these waves, the mean, variance, skewness, and kurtosis were calculated for the energy (static coefficient), velocity (first differential), and acceleration (second differential).

The Librosa package [31] was used to calculate the mean, maximum, minimum, and standard deviation of the root mean square value, centroid, bandwidth, flatness, zero-crossing rate, loudness, and flux of the spectrogram, or the visualization of the recording.

The source characteristics were extracted using the Signal_Analysis package, version 0.1.26, to extract the micromovements of the sound wave: harmonics-to-noise ratio (HNR), jitter, shimmer, and glottal pulses. Jitter and shimmer are two features of vocal signals that describe the frequency variation from cycle to cycle of the sound wave and the waveform amplitude, respectively [32,33]. While jitter rises with the growing lack of control of vocal cord vibration, higher shimmer is coupled with increased breathiness. HNR is the ratio between periodic components and nonperiodic components that

constitute a voiced speech segment [34]. These components correspond to the vibration from vocal cords and glottal noise, respectively.

Speech features vary naturally between males and females due to differences in the length of the vocal tract. These differences have been leveraged in gender classification through speech analysis based on pitch and formant frequencies [35], HNR [36], linear predictive components, and mel-frequency cepstral coefficients (MFCCs) [37]. Previous work found differences in speech depending on gender in the effects of depression and the effectiveness of classifiers for its detection [38]. This is why this study considers males and females separately.

Statistical Analysis

The data collected were described using mean and standard deviation for quantitative variables, and frequency and percentage for qualitative variables. Demographic characteristics, such as age and gender, were compared between different groups of caregivers using a chi-square test for qualitative variables (eg, gender) and an analysis of variance performed for quantitative variables (eg, age). Similarly, the data measured for voice and scores were compared between different groups of caregivers. The normality of the collected data was tested using a Shapiro test. In order to test the relationship between the different voice measures and the measured scores, Spearman correlations were used. In addition, to test the link between the voice measures and the therapist's decision, Student *t* tests or Wilcoxon-Mann-Whitney tests were performed. A *P* value of less than .05 was considered significant. The analyses were performed using the free statistical software RStudio 4.0.0 [39]. Further, regression analyses were performed with the extracted vocal features to determine the error rate for predicting the participants' stress scores.

Results

Participants

In total, 89 French-speaking health professionals, aged between 20 and 74 years, accepted the outreach phone calls and their speech samples were recorded and analyzed. Their demographic characteristics are presented in Table 1.

The mean age of the participants was 40.53 years (SD 14.19). The mean stress score on the PSS was 22.43 (SD 7.16) and on the MSA questionnaire was 2.92 (SD 2.09). The majority of the participants scored below 26 on the PSS but above 0 on the MSA questionnaire. Results on the PSS and on the MSA stress scale were proportional. We found that 27% (24/89) of the recorded health professionals experienced intense stress, and 28% (25/89) experienced occasional stress. Only 16% (14/89) of the participants requested a follow-up. The stress level was gender dependent, with females reporting higher stress levels. For males, stress levels tended to drop with age. Figure 1 shows a distribution of the total stress scores across genders. The total stress scores in the female group are more dispersed than in the male group and are generally higher. A total of 14 out of 88 (16%) participants (11/57, 19% of all females; 3/31, 10% of all males) asked for a follow-up call. Their mean PSS score (mean 31.78, SD 7.40) and mean MSA scale score (mean 5.57, SD 1.34) were significantly higher than for those who did not ask for a follow-up, whose mean PSS score was 20.60 (SD 5.63) and mean MSA scale score was 2.38 (SD 1.8).

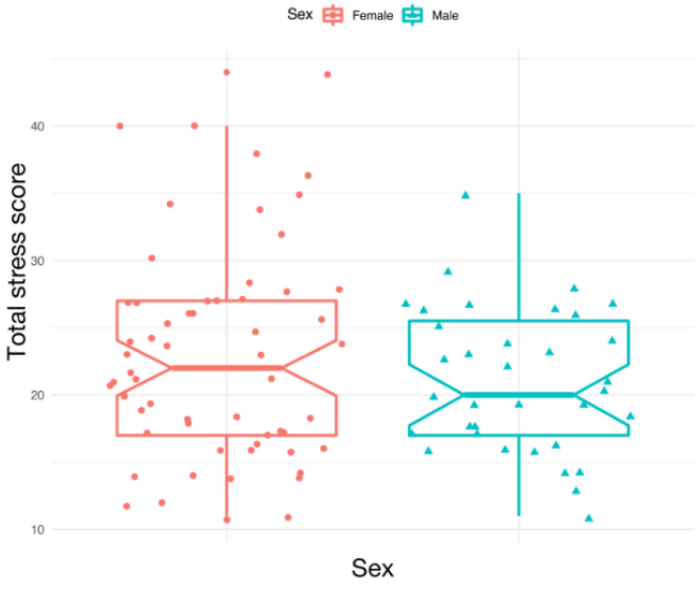
Table 1. Descriptive statistics for participant characteristics (N=89).

Characteristic	Participants, n (%)			*P* value[a]
	Total (N=89)	Male (n=31)	Female (n=58)	
Gender	89 (100)	31 (35)	58 (65)	N/A[b]
Education (years) (n=81)				.03
<12	19 (23)	1/28 (4)	18/53 (34)	
≥12	62 (77)	27/28 (96)	35/53 (66)	
Timing of call				.03
During lockdown	34 (38)	7 (23)	27 (47)	
After lockdown	55 (62)	24 (77)	31 (53)	
Perceived Stress Scale score				.47
Knows how to manage stress (<21)	40 (45)	16 (52)	24 (41)	
Generally knows how to cope with stress (21-26)	25 (28)	9 (29)	16 (28)	
Life is a constant threat (>26)	24 (27)	6 (19)	18 (31)	
Motivation Stress Affect (MSA) scale score				.99
0	23 (26)	8 (26)	15 (26)	
>0	66 (74)	23 (74)	43 (74)	
MSA motivation scale score				.47
0	30 (34)	12 (39)	18 (31)	
>0	59 (66)	19 (61)	40 (69)	
MSA depression scale score				.32
0	57 (64)	22 (71)	35 (60)	
>0	32 (36)	9 (29)	23 (40)	
Follow-up request (n=88)				.36
No	74 (84)	28 (90)	46/57 (81)	
Yes	14 (16)	3 (10)	11/57 (19)	

[a]Chi-square test or Fisher exact test.

[b]N/A: not applicable; the *P* value was not calculated for gender.

Figure 1. Stress score distribution across genders.



Correlations

First, vocal and nonvocal features were analyzed in relation to the stress level. The data set was quite small and, therefore, rather than training a classifier, we performed correlation analysis between the features computed for each speech task and the reported stress level. Further, only extracted speech features were considered; a priori, nonmeaningful features, like ID, were removed.

We performed a selection of the top k features based on their descriptive power for the target variable *total stress score*. Vocal features might be gender dependent. Therefore, we performed a selection of top features for male and female data sets separately. We used Spearman correlation, since we had both ordinal and continuous features: the target *total stress score* is ordinal. Since Spearman correlation uses only the ranks of the variables and not their raw values, we could omit the normalization step. We considered absolute values of the correlation coefficient for feature scoring. Results are presented in Table 2.

The main speech parameters correlating with stress levels in both genders were spectral (ie, formant) features, namely the MFCCs. These features characterize the spectrum of speech, which is the frequency distribution of the speech signal at a specific time. MFCCs were derived by computing a spectrum of the log-magnitude mel-spectrum of the audio segment. The lower coefficients represent the vocal tract filter and the higher coefficients represent periodic vocal fold sources [18]. Moreover, in males' prosodic characteristics, such as the F0, and in females with the positive storytelling, pitch ranges were associated with stress levels.

Table 2. Correlation between stress levels and speech features.

Top 10 features for each data set	Task	Spearman correlation
Female data set		
MFCC[a]3 acceleration skewness	Positive story	0.49
MFCC2 mean	Neutral story	0.44
Pitch range	Positive story	0.44
MFCC3 acceleration skewness	Negative story	0.43
MFCC2 mean	Positive story	0.44
MFCC5 acceleration kurtosis	Negative story	–0.42
MFCC2 mean	Negative story	0.43
MFCC5 velocity kurtosis	Negative story	–0.40
MFCC3 acceleration skewness	Neutral story	0.39
MFCC5 velocity kurtosis	Negative story	0.39
Male data set		
Upper quartile F0[b]	Neutral story	–0.54
Pronunciation posteriori probability score percentage	Positive story	–0.50
Energy acceleration mean	Positive story	0.52
Mean F0	Neutral story	–0.51
MFCC9 kurtosis	Positive story	0.41
MFCC9 variance	Positive story	–0.44
Upper quartile F0	Negative story	–0.47
MFCC4 acceleration mean	Positive story	–0.40
Upper quartile F0	Positive story	–0.47
MFCC12 acceleration skewness	Neutral story	–0.42

[a]MFCC: mel-frequency cepstral coefficient; the numbers following MFCC are part of the feature names presenting their location on a spectrum.

[b]F0: fundamental frequency.

For female participants, correlation analyses between negative, positive, and neutral features and the target feature *total stress score* were performed. Among the top 5 features, we have MFCC acceleration skewness, which correlates with the stress level by 0.45 and 0.37 in the positive and neutral tasks, respectively. The other features among top 5 features are task specific. Thus, for each task there is a different set of features associated with stress level.

For male participants, the selection was performed analogously. The top features are task specific as well, and they differ from the features for the female data set. In this sample, we obtained more negatively correlating features than for the female data

set; this meant that features, for instance, related to F0 of low value (mean F0 in the neutral story with –0.51, upper quartile F0 in the negative and positive story with –0.47, and upper quartile F0 in the neutral story with –0.54) are associated with high stress scores. In general, low values represent a smaller pitch range.

Regression

Stress scores were regressed against measurements for positive, neutral, and negative tasks. Similarly, the regression for tasks of different sentiments was performed for groups of female and male participants to allow for possible impacts of gender on stress levels. For the regressors, we used linear, support vector machine (SVM), and random forest regressors to predict the stress scores.

The first regression approximated the stress score by estimating coefficients for each feature in the training data, where greater coefficients indicate a greater influence over the predicted value. Linear regression models are fast, highly interpretable, and commonly used for prediction of stress scores from audio features and speech analysis, according to previous studies [40-42]. The random forest regressor created a number of decision trees that were constructed based on random sampling from the training data; each tree then attempted to determine the best way to predict the scores given the data it received. Each decision tree outputted a predicted value and the mode value was selected. Decision tree methods have shown high accuracy with good interpretability in similar studies where vocal and linguistic features were employed for detection of emotions, social signals, and mental health problems [43-45]. The SVM regressor took each set of features and projected them as a vector onto a space and attempted to find the optimal way to separate the data. The stress score was then based on the distance from that separator. Stress modeling with inputs from physiological sensors or audio sources using SVM has also been previously reported to give high model performance [46-48]. In recent studies, both SVM and random forest provided notably high prediction and classification strength for stress detection using various speech features [49-51].

The caret package from R, version 3.4.2 (The R Foundation), was used for data training and validation. A 10-fold cross-validation was performed and performance was evaluated using the mean absolute error (MAE): the average of the absolute difference between the predicted and actual values from our models for all participants. The score ranges from 0 to infinity, where a score closer to 0 indicates a better-fitting model.

The prediction of total stress scores using all or a subset of tasks among male or female subjects was carried out using various baseline regression models, whose performances were evaluated by the plots in Figure 2, where the MAE values are presented on the y-axis. Overall, the prediction strength in males was better than in females for all sentiments, as shown by a trend of lower errors (lowest MAE for males was 3.84; lowest MAE for females was 5.56). It is notable that stress score regression models based on negative tasks in males and neutral tasks in females performed relatively poorly compared to other tasks. For both male and female participants, using positive tasks for regression yielded equivalent or better results than using all tasks, suggesting that a subset of tasks could be employed for accurate and less time-consuming prediction of stress scores. An overview of the lowest scores for each testing scenario is presented in Table 3.

All regression models outperformed their respective baseline MAE values (4.46 and 6.35 in males and females, respectively). Linear models and the SVM regressor were the most precise for the prediction of total stress scores in general.

Figure 2. Performances of different computerized regression models in predicting stress levels based on vocal features. Boosted: boosted linear model; ElasticNet: mix of L1 and L2 regularized linear regression; MAE: mean absolute error; Poly: support vector machine with polynomial basis function kernel; Quantile: quantile regression forest; Radial: support vector machine with radial basis function kernel; SVM: support vector machine.

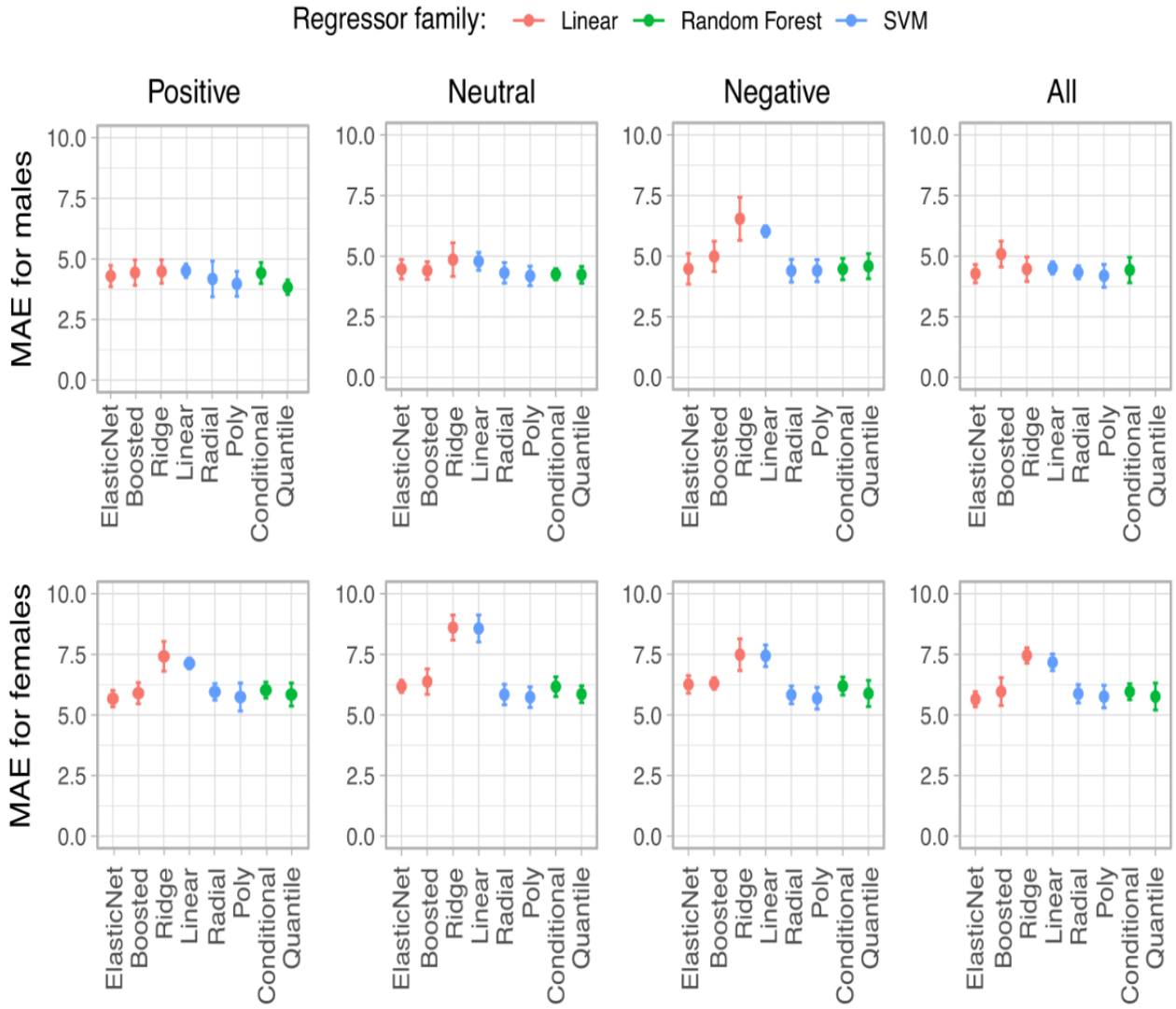


Table 3. The lowest scores for each testing scenario.

Participant group	Positive tasks		Neutral tasks		Negative tasks	
	MAE[a] (SD)	Model	MAE (SD)	Model	MAE (SD)	Model
All	5.31 (0.25)	ElasticNet[b]	5.25 (0.28)	QuantileRF[c]	5.34 (0.35)	PolySVM[d]
Male	3.84 (0.43)	QuantileRF	4.40 (0.37)	BoostedLM[e]	4.37 (0.43)	PolySVM
Female	5.56 (0.41)	ElasticNet	5.84 (0.42)	RadialSVM[f]	5.68 (0.45)	PolySVM

[a]MAE: mean absolute error.

[b]ElasticNet: mix of L1 and L2 regularized linear regression.

[c]QuantileRF: quantile regression forest.

[d]PolySVM: support vector machine with polynomial basis function kernel.

[e]BoostedLM: boosted linear model.

[f]RadialSVM: support vector machine with radial basis function kernel.

Discussion

Principal Findings

The purpose of this study was to investigate the potential of using automatic speech analysis for the detection of stress in health care professionals during the current COVID-19 pandemic. This would potentially lead to earlier and timely prevention among this high-risk population. Firstly, speech samples were collected over the phone, and various voice features were extracted and compared with classical stress measures. Secondly, based on the extracted features, scores obtained by participants on the completed stress scale were predicted.

The main outcome of this study was the demonstration of this approach's feasibility under the given context, as all participants were cooperative and appreciated the initiative of rapidly applying this existing technology to this specific use case. Moreover, from phone call recordings, a number of vocal correlates of stress have been identified, namely in the area of spectral features (ie, MFCC) as well as prosodic features such as F0, which seem to be the most commonly reported features in well-controlled trials [11]. Stress scores could be predicted based on speech features with relatively small errors.

Spectral features characterize the speech spectrum; the frequency distribution of the speech signal at a specific time indicates information in some high-dimensional representation [18]. The features capture information regarding changes in muscle tension and control and have consistently been observed to change with a speaker's mental state. A few depression studies reported a relative shift in energy with increasing depression severity [52,53].

Another result we obtained was that most identified vocal features were task dependent as well as gender dependent. Interestingly, in the female group, MFCC features seemed to be associated with stress levels during all tasks, meaning that it did not matter what participants were talking about; as long as sufficient speech was captured, meaningful information could be extracted and subtle signs of stress level could be detected. On the other hand, in the male data set, the upper quartile F0 appeared as a task-independent feature sensitive to stress levels. Overall, in the male data set, we observed more features with a negative correlation than we did for the female data set.

Voice production can be divided into three processes: breathing, phonation, and resonance stress [54]. For the second process, phonation, the vocal folds must close and open again to create vibration. The frequency rate of these pulses determines the F0 of the vocal source contributing to the perceived pitch of the sound.

Previous research showed that increased muscle tension tends to be caused by stress [55,56], resulting in a tensing of the vocal folds, which, in turn, most likely causes a raising of F0. A recent review on voice analysis in stress [22] stated that the parameter F0 has been considered as a "universal stress indicator," whereas increased levels of F0 might be linked with acute bottom-up processes of sympathetic arousal. Similar studies of analysis of phone call recordings during situational stress situations revealed an increase in F0 and intensity with presumed levels of stress [55,57,58]. Our findings seem consistent with the majority of acoustic studies, pointing to F0 as one important marker of stress levels.

However, most correlations we found were with resonance (ie, formant) parameters, which are involved in the quality of sound shaping and vowel and consonant pronunciation and are produced by the muscle activity involved in the shaping of the resonant cavities of the vocal tract system [59]. These parameters are less documented in regard to stress. The MFCC, in particular, can be indicative of breathiness in the voice [60]. Interestingly, one study found a circadian pattern in MFCCs due to sleep deprivation. For this, voice perturbations were compared with classical sleep measures [61] and correlations were found between fatigue scores and MFCCs. This might eventually explain our results, as most participants also reported signs of fatigue during the interviews.

Another study examined speech in students under exam stress and a few days later; in this case, heart rate was measured to control for the actual stress levels. Under stress, students' heart rates increased, F0 and F0 SD increased, first formant (F1) and second formant (F2) frequencies increased, and MFCCs decreased in relation to baseline levels [62].

It can be hypothesized that given our recorded population who reported relatively mild to moderate levels of stress, rather subtle changes in voice parameters were found and, therefore, weaker correlations were observed. However, it is important to underline that changes in features that we found to be sensitive to stress levels were gender dependent but not necessarily task dependent. They were most likely too small to be detectable by the human ear but were captured by the automatic speech analysis. We assume that by applying this technology to regular check-up calls with people experiencing high stress levels, such as health care professionals, very early signs of stress can be detected in their voices, allowing for timely preventive strategies.

Regression models using vocal features performed relatively well in predicting stress scores, namely in the positive story task for both genders (MAE of 5.31). It shows that the technology could capture indicative patterns from even a short amount of time, possibly even from one task, to recognize tendencies of stress levels in a fragile but healthy population; this represents a promising rapid tool for prediction of stress scores.

Strengths of This Study

This study is a first step into the early identification of stress in an at-risk population, such as caregivers, who do not directly express their psychological suffering. We can imagine extending this technique to other fragile populations for early screening of stress, such as teenagers who are victims of school harassment or women who are victims of abuse, where timely management could potentially prevent the development of comorbidities, such as depression and anxiety. Moreover, patient populations who have difficulty expressing their problems, such as those with autism spectrum disorder or dementia, could benefit from this technology.

Generally, remote psychological counseling is controversial. Nevertheless, it is becoming necessary due to current economic, social, and health constraints, but has been received by professionals and patients with mixed feelings. Indeed, the nonverbal part of communication is lost and the dynamics of interaction are not the same. However, contrary to these preconceived ideas, we have noticed during this work that it is easier for certain participants to open up and speak about personal issues during these interviews in a liberating manner, similar to a confessional. Not being in the physical presence of the listener may facilitate personal expression, with less fear of being judged. This aspect is very interesting during a screening because it considerably accelerates the process of detection and diagnosing of psychological symptoms.

Weaknesses of the Study

This project has been rapidly implemented, initially with an approach of qualitative and quantitative data analysis, that should contribute to the early and timely assistance of health professionals during the COVID-19 pandemic. The staff members available to participate in the study were limited. Patient selection was done on a voluntary basis. It is conceivable that the population studied were more concerned about their state of psychological suffering and, therefore, potentially had a selection bias.

Although the voice recordings were made in the middle of the interview without this time being precisely stated, it is possible that some patients may have suspected this, which could have been anxiety provoking and skewed our results. Recording throughout the interview for parameters not affected by the tasks would provide more data and more robust results.

Finally, the obtained correlations can be considered as rather moderate, which makes it difficult to draw any strong conclusions. A larger data set, ideally of a longitudinal nature, with more precise characterization of the speakers is needed in order to verify whether the correlating features represent real markers of stress.

Future Perspective

For future work, we propose to perform this analysis on a larger data set and to build a prediction model. In case of an insufficient number of observations per stress level, the number of stress levels can be reduced by binning. Binning can also be carried out on characteristic values.

Further studies with acoustic measurements and stress questionnaires at regular time intervals would allow for the analysis of the kinetics of the markers and a better perception of their sensitivity and specificity. In addition, adding clinical measurements of psychiatric symptoms, such as the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition [63], would make it possible to perceive whether one of the markers is predictive of an anxiety or depression disorder. The use of the tool could be combined with the delivery of preventive strategies, such as physical exercises, adaptation of diet, psychotherapy, meditation, or the use of symptomatic treatments, and it could be employed at the same time for the evaluation of the obtained effects. However, in order to produce a real-world application of this technology, larger validation studies have to be performed to demonstrate clinical meaningfulness by comparing its performance to standardized measurement tools.

Acknowledgments

This research was supported by a grant from the Fondation GSF (Groupe Services France) Jean Louis Noisiez and the Association Innovation Alzheimer. Thanks to all of the professional health care workers who agreed to participate in this study.

Authors' Contributions

AK, KR, JE, and PR designed and conducted the study. AD contributed as technical support to this study. NL, HL, and RF analyzed the data. KR, AK, NL, HL, and PR drafted the manuscript. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest

NL is an employee and shareholder of ki elements UG.

References

1. Pneumonia of unknown cause – China. World Health Organization. 2020 Jan 05. URL: https://www.who.int/csr/don/05-january-2020-pneumonia-of-unkown-cause-china/en/ [accessed 2020-01-05]
2. WHO announces COVID-19 outbreak a pandemic. World Health Organization. 2020 Mar 12. URL: https://www.euro.who.int/en/health-topics/health-emergencies/coronavirus-covid-19/news/news/2020/3/who-announces-covid-19-outbreak-a-pandemic [accessed 2020-03-12]
3. Liu X, Kakade M, Fuller CJ, Fan B, Fang Y, Kong J, et al. Depression after exposure to stressful events: Lessons learned from the severe acute respiratory syndrome epidemic. Compr Psychiatry 2012 Jan;53(1):15-23 [FREE Full text] [doi: 10.1016/j.comppsych.2011.02.003] [Medline: 21489421]
4. Lung F, Lu Y, Chang Y, Shu B. Mental symptoms in different health professionals during the SARS attack: A follow-up study. Psychiatr Q 2009 Jun;80(2):107-116. [doi: 10.1007/s11126-009-9095-5] [Medline: 19247834]

5. Wu P, Fang Y, Guan Z, Fan B, Kong J, Yao Z, et al. The psychological impact of the SARS epidemic on hospital employees in China: Exposure, risk perception, and altruistic acceptance of risk. Can J Psychiatry 2009 May;54(5):302-311 [FREE Full text] [doi: 10.1177/070674370905400504] [Medline: 19497162]
6. Pappa S, Ntella V, Giannakas T, Giannakoulis VG, Papoutsi E, Katsaounou P. Prevalence of depression, anxiety, and insomnia among healthcare workers during the COVID-19 pandemic: A systematic review and meta-analysis. Brain Behav Immun 2020 Aug;88:901-907 [FREE Full text] [doi: 10.1016/j.bbi.2020.05.026] [Medline: 32437915]
7. Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. J Health Soc Behav 1983 Dec;24(4):385-396. [Medline: 6668417]
8. Roohafza H, Ramezani M, Sadeghi M, Shahnam M, Zolfagari B, Sarafzadegan N. Development and validation of the stressful life event questionnaire. Int J Public Health 2011 Aug;56(4):441-448. [doi: 10.1007/s00038-011-0232-1] [Medline: 21327856]
9. Amirkhan J. Stress overload: A new approach to the assessment of stress. Am J Community Psychol 2012 Mar;49(1-2):55-71. [doi: 10.1007/s10464-011-9438-x] [Medline: 21538152]
10. Petrowski K, Paul S, Albani C, Brähler E. Factor structure and psychometric properties of the trier inventory for chronic stress (TICS) in a representative German sample. BMC Med Res Methodol 2012 Apr 01;12:42 [FREE Full text] [doi: 10.1186/1471-2288-12-42] [Medline: 22463771]
11. Giddens CL, Barron KW, Byrd-Craven J, Clark KF, Winter AS. Vocal indices of stress: A review. J Voice 2013 May;27(3):390.e21-390.399. [doi: 10.1016/j.jvoice.2012.12.010] [Medline: 23462686]
12. Pisanski K, Nowak J, Sorokowski P. Individual differences in cortisol stress response predict increases in voice pitch during exam stress. Physiol Behav 2016 Sep 01;163:234-238. [doi: 10.1016/j.physbeh.2016.05.018] [Medline: 27188981]
13. Kirchhübel C, Howard DM, Stedmon AW. Acoustic correlates of speech when under stress: Research, methods and future directions. Int J Speech Lang Law 2011 Sep 13;18(1):75-98. [doi: 10.1558/ijsll.v18i1.75]
14. Hollien H. Vocal fold dynamics for frequency change. J Voice 2014 Jul;28(4):395-405. [doi: 10.1016/j.jvoice.2013.12.005] [Medline: 24726331]
15. Sobin C, Sackeim HA. Psychomotor symptoms of depression. Am J Psychiatry 1997 Jan;154(1):4-17. [doi: 10.1176/ajp.154.1.4] [Medline: 8988952]
16. Schrijvers D, Hulstijn W, Sabbe BG. Psychomotor symptoms in depression: A diagnostic, pathophysiological and therapeutic tool. J Affect Disord 2008 Jul;109(1-2):1-20. [doi: 10.1016/j.jad.2007.10.019] [Medline: 18082896]
17. Bylsma LM, Morris BH, Rottenberg J. A meta-analysis of emotional reactivity in major depressive disorder. Clin Psychol Rev 2008 Apr;28(4):676-691. [doi: 10.1016/j.cpr.2007.10.001] [Medline: 18006196]
18. Cummins N, Scherer S, Krajewski J, Schnieder S, Epps J, Quatieri TF. A review of depression and suicide risk assessment using speech analysis. Speech Commun 2015 Jul;71:10-49. [doi: 10.1016/j.specom.2015.03.004]
19. Nilsonne A. Speech characteristics as indicators of depressive illness. Acta Psychiatr Scand 1988 Mar;77(3):253-263. [doi: 10.1111/j.1600-0447.1988.tb05118.x] [Medline: 3394527]
20. Leff J, Abberton E. Voice pitch measurements in schizophrenia and depression. Psychol Med 1981 Nov;11(4):849-852. [doi: 10.1017/s0033291700041349] [Medline: 7323240]
21. Marmar CR, Brown AD, Qian M, Laska E, Siegel C, Li M, et al. Speech-based markers for posttraumatic stress disorder in US veterans. Depress Anxiety 2019 Jul;36(7):607-616 [FREE Full text] [doi: 10.1002/da.22890] [Medline: 31006959]
22. Van Puyvelde M, Neyt X, McGlone F, Pattyn N. Voice stress analysis: A new framework for voice and effort in human performance. Front Psychol 2018;9:1994 [FREE Full text] [doi: 10.3389/fpsyg.2018.01994] [Medline: 30515113]
23. Robert P, Lanctôt KL, Agüera-Ortiz L, Aalten P, Bremond F, Defrancesco M, et al. Is it time to revise the diagnostic criteria for apathy in brain disorders? The 2018 international consensus group. Eur Psychiatry 2018 Oct;54:71-76 [FREE Full text] [doi: 10.1016/j.eurpsy.2018.07.008] [Medline: 30125783]
24. Yesavage JA, Brink T, Rose TL, Lum O, Huang V, Adey M, et al. Development and validation of a geriatric depression screening scale: A preliminary report. J Psychiatr Res 1982 Jan;17(1):37-49. [doi: 10.1016/0022-3956(82)90033-4]
25. König A, Linz N, Zeghari R, Klinge X, Tröger J, Alexandersson J, et al. Detecting apathy in older adults with cognitive disorders using automatic speech analysis. J Alzheimers Dis 2019 Jun 18;69(4):1183-1193. [doi: 10.3233/jad-181033]
26. ki:elements. URL: https://ki-elements.de [accessed 2020-11-01]
27. Van Rossum G, Drake FL. Python 3 Reference Manual. Scotts Valley, CA: CreateSpace; 2009:1-242.
28. Shahab S. My-Voice Analysis. GitHub. 2020. URL: https://github.com/Shahabks/my-voice-analysis [accessed 2020-11-01]
29. Boersma P. PRAAT, a system for doing phonetics by computer. Glot International. Vol 5, No. 9/10. 2001. URL: https://www.researchgate.net/profile/Paul-Boersma-2/publication/208032992_PRAAT_a_system_for_doing_phonetics_by_computer/links/53d77c1c0cf29f55cfb4d26b/PRAAT-a-system-for-doing-phonetics-by-computer.pdf [accessed 2021-04-07]
30. Lyons J. python_speech_features. GitHub. 2020 Jan 14. URL: https://github.com/jameslyons/python_speech_features [accessed 2020-11-01]
31. librosa 0.8.0. Python Package Index. 2020 Jul 21. URL: https://pypi.org/project/librosa/ [accessed 2020-11-01]
32. Kreiman J, Gerratt BR. Perception of aperiodicity in pathological voice. J Acoust Soc Am 2005 Apr;117(4 Pt 1):2201-2211. [doi: 10.1121/1.1858351] [Medline: 15898661]

33. Michaelis D, Fröhlich M, Strube HW, Kruse E, Story B, Titze IR. Some simulations concerning jitter and shimmer measurement. In: Proceedings of the 3rd International Workshop on Advances in Quantitative Laryngoscopy, Voice and Speech Research. 1998 Jun Presented at: 3rd International Workshop on Advances in Quantitative Laryngoscopy, Voice and Speech Research; 1998; Aachen, Germany p. 744-754.
34. Murphy PJ, Akande OO. Cepstrum-based estimation of the harmonics-to-noise ratio for synthesized and human voice signals. In: Proceedings of the International Conference on Nonlinear Analyses and Algorithms for Speech Processing. 2005 Apr Presented at: International Conference on Nonlinear Analyses and Algorithms for Speech Processing; April 19-22, 2005; Barcelona, Spain p. 150-160. [doi: 10.1007/11613107_13]
35. Childers D, Wu K. Gender recognition from speech. Part II: Fine analysis. J Acoust Soc Am 1991 Oct;90(4 Pt 1):1841-1856. [doi: 10.1121/1.401664] [Medline: 1755877]
36. Heffernan K. Evidence from HNR that/s/is a social marker of gender. Toronto Working Papers in Linguistics 2004;23:71-84 [FREE Full text]
37. Wu K, Childers DG. Gender recognition from speech. Part I: Coarse analysis. J Acoust Soc Am 1991 Oct;90(4 Pt 1):1828-1840. [doi: 10.1121/1.401663] [Medline: 1960278]
38. Low LA, Maddage MC, Lech M, Sheeber LB, Allen NB. Detection of clinical depression in adolescents' speech during family interactions. IEEE Trans Biomed Eng 2011 Mar;58(3):574-586. [doi: 10.1109/tbme.2010.2091640]
39. The Comprehensive R Archive Network. URL: https://cran.r-project.org/ [accessed 2020-11-01]
40. Gillespie S, Moore E, Laures-Gore J, Farina M, Russell S, Logan YL. Detecting stress and depression in adults with aphasia through speech analysis. In: Proceedings of the 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). 2017 Presented at: 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP); March 5-9, 2017; New Orleans, LA p. 5140-5144. [doi: 10.1109/icassp.2017.7953136]
41. van den Broek EL, van der Sluis F, Dijkstra T. Telling the story and re-living the past: How speech analysis can reveal emotions in post-traumatic stress disorder (PTSD) patients. In: Sensing Emotions. Dordrecht, the Netherlands: Springer; 2010:153-180.
42. Muaremi A, Arnrich B, Tröster G. Towards measuring stress with smartphones and wearable devices during workday and sleep. Bionanoscience 2013;3:172-183 [FREE Full text] [doi: 10.1007/s12668-013-0089-2] [Medline: 25530929]
43. Hasan M, Rundensteiner E, Agu E. Emotex: Detecting emotions in twitter messages. In: Proceedings of the 2014 ASE Bigdata/Socialcom/Cybersecurity Conference. 2014 Presented at: 2014 ASE Bigdata/Socialcom/Cybersecurity Conference; May 27-31, 2014; San Francisco, CA p. 1-10 URL: https://web.cs.wpi.edu/~emmanuel/publications/PDFs/C30.pdf
44. Gosztolya G, Busa-Fekete R, Tóth L. Detecting autism, emotions and social signals using AdaBoost. In: Proceedings of the 14th Annual Conference of the International Speech Communication Association. 2013 Presented at: 14th Annual Conference of the International Speech Communication Association; August 25-29, 2013; Lyon, France p. 220-224 URL: https://www.isca-speech.org/archive/archive_papers/interspeech_2013/i13_0220.pdf
45. Howes C, Purver M, McCabe R. Linguistic indicators of severity and progress in online text-based therapy for depression. In: Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality. 2014 Presented at: Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality; June 27, 2014; Baltimore, MA p. 7-16 URL: https://www.aclweb.org/anthology/W14-3202.pdf [doi: 10.3115/v1/w14-3202]
46. Al-Shargie F, Tang TB, Badruddin N, Kiguchi M. Towards multilevel mental stress assessment using SVM with ECOC: An EEG approach. Med Biol Eng Comput 2018 Jan;56(1):125-136. [doi: 10.1007/s11517-017-1733-8] [Medline: 29043535]
47. Rabaoui A, Davy M, Rossignol S, Ellouze N. Using one-class SVMs and wavelets for audio surveillance. IEEE Trans Inf Forensics Secur 2008 Dec;3(4):763-775. [doi: 10.1109/tifs.2008.2008216]
48. Chang C, Chang C, Zheng J, Chung P. Physiological emotion analysis using support vector regression. Neurocomputing 2013 Dec;122:79-87. [doi: 10.1016/j.neucom.2013.02.041]
49. Soury M, Devillers L. Stress detection from audio on multiple window analysis size in a public speaking task. In: Proceedings of the 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction. 2013 Presented at: 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction; September 2-5, 2013; Geneva, Switzerland p. 529-533. [doi: 10.1109/acii.2013.93]
50. Sysoev M, Kos A, Pogačnik M. Noninvasive stress recognition considering the current activity. Pers Ubiquitous Comput 2015 Aug 26;19(7):1045-1052 [FREE Full text] [doi: 10.1007/s00779-015-0885-5]
51. Gjoreski M, Gjoreski H, Lutrek M, Gams M. Automatic detection of perceived stress in campus students using smartphones. In: Proceedings of the 2015 International Conference on Intelligent Environments. 2015 Presented at: 2015 International Conference on Intelligent Environments; July 15-17, 2015; Prague, Czech Republic p. 132-135. [doi: 10.1109/IE.2015.27]
52. Cummins N, Epps J, Ambikairajah E. Spectro-temporal analysis of speech affected by depression and psychomotor retardation. In: Proceedings of the 2013 IEEE International Conference on Acoustics, Speech and Signal Processing. 2013 Presented at: 2013 IEEE International Conference on Acoustics, Speech and Signal Processing; May 26-31, 2013; Vancouver, BC p. 7542-7546. [doi: 10.1109/ICASSP.2013.6639129]
53. Cummins N, Epps J, Sethu V, Breakspear M, Goecke R. Modeling spectral variability for the classification of depressed speech. In: Proceedings of the 14th Annual Conference of the International Speech Communication Association. 2013

Presented at: 14th Annual Conference of the International Speech Communication Association; August 25-29, 2013; Lyon, France p. 857-861 URL: https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.654.9966&rep=rep1&type=pdf
54. Kreiman J, Sidtis D. Foundations of Voice Studies: An Interdisciplinary Approach to Voice Production and Perception. Hoboken, NJ: John Wiley & Sons; 2011.
55. Streeter L, Macdonald NH, Apple W, Krauss RM, Galotti KM. Acoustic and perceptual indicators of emotional stress. J Acoust Soc Am 1983 Apr;73(4):1354-1360. [doi: 10.1121/1.389239] [Medline: 6853847]
56. Scherer KR, Grandjean D, Johnstone T, Klasmeyer G, Bänziger T. Acoustic correlates of task load and stress. In: Proceedings of the 7th International Conference on Spoken Language Processing. 2002 Presented at: 7th International Conference on Spoken Language Processing; September 16-20, 2002; Denver, CO p. 1-4 URL: https://www.isca-speech.org/archive/archive_papers/icslp_2002/i02_2017.pdf
57. Ruiz R, Absil E, Harmegnies B, Legros C, Poch D. Time- and spectrum-related variabilities in stressed speech under laboratory and real conditions. Speech Commun 1996 Nov;20(1-2):111-129. [doi: 10.1016/s0167-6393(96)00048-9]
58. Jessen M. Einfluss von Stress auf Sprache und Stimme: Unter besondere Berücksichtigung polizeidienstlicher Anforderungen. Idstein, Germany: Schulz-Kirchner Verlag; 2006:1-218.
59. Gopalan K, Wenndt S, Cupples EJ. An analysis of speech under stress using certain modulation features. In: Proceedings of the 25th Annual Conference of the IEEE Industrial Electronics Society. 1999 Presented at: 25th Annual Conference of the IEEE Industrial Electronics Society; November 29-December 3, 1999; San Jose, CA p. 1193-1197. [doi: 10.1109/iecon.1999.819381]
60. Hillenbrand J, Houde RA. Acoustic correlates of breathy vocal quality: Dysphonic voices and continuous speech. J Speech Hear Res 1996 Apr;39(2):311-321. [doi: 10.1044/jshr.3902.311] [Medline: 8729919]
61. Greeley HP, Friets E, Wilson JP, Raghavan S, Picone J, Berg J. Detecting fatigue from voice using speech recognition. In: Proceedings of the IEEE International Symposium on Signal Processing and Information Technology. 2006 Presented at: IEEE International Symposium on Signal Processing and Information Technology; August 27-30, 2006; Vancouver, BC p. 567-571. [doi: 10.1109/ISSPIT.2006.270865]
62. Sigmund M. Introducing the database ExamStress for speech under stress. In: Proceedings of the 7th Nordic Signal Processing Symposium. 2006 Presented at: 7th Nordic Signal Processing Symposium; June 7-9, 2006; Reykjavik, Iceland p. 290-293. [doi: 10.1109/NORSIG.2006.275258]
63. American Psychiatric Association. Diagnostic And Statistical Manual Of Mental Disorders, Fifth Edition. Arlington, VA: American Psychiatric Association Publishing; 2013.

Abbreviations

API: application programming interface
F0: fundamental frequency
F1: first formant
F2: second formant
GSF: Groupe Services France
HNR: harmonics-to-noise ratio
MAE: mean absolute error
MFC: mel-frequency cepstrum
MFCC: mel-frequency cepstral coefficient
MSA: Motivation Stress Affect
PSS: Perceived Stress Scale
SVM: support vector machine

Edited by G Eysenbach; submitted 28.09.20; peer-reviewed by A Khaleghi, S D'Arcy, E van der Velde; comments to author 24.10.20; revised version received 13.11.20; accepted 17.03.21; published 19.04.21.

Please cite as:
König A, Riviere K, Linz N, Lindsay H, Elbaum J, Fabre R, Derreumaux A, Robert P
Measuring Stress in Health Professionals Over the Phone Using Automatic Speech Analysis During the COVID-19 Pandemic: Observational Pilot Study
J Med Internet Res 2021;23(4):e24191
URL: https://www.jmir.org/2021/4/e24191
doi:10.2196/24191
PMID:33739930

Original Paper

The Uncounted Casualties of a Hidden COVID-19 Epidemic in China: Cross-sectional Study on Deaths Related to Overwork

Zhicheng Wang[1,2], PhD; Leesa Lin[3], PhD; Yan Guo[4], MPH; Huayi Xiong[5], BA; Kun Tang[1], DPhil

[1]Vanke School of Public Health, Tsinghua University, Beijing, China

[2]School of Medicine, Tsinghua University, Beijing, China

[3]Department of Infectious Disease Epidemiology, London School of Hygiene & Tropical Medicine, London, United Kingdom

[4]Department of Health Policy and Management, Peking University, Beijing, China

[5]School of Health Humanities, Peking University, Beijing, China

Corresponding Author:
Kun Tang, DPhil
Vanke School of Public Health
Tsinghua University
No. 30, Shuangqing Road, Haidian District
Beijing
China
Phone: 86 13671129425
Email: tangk@mail.tsinghua.edu.cn

Abstract

Background: During the COVID-19 response, nonclinical essential workers usually worked overtime and experienced significant work stress, which subsequently increased their risk of mortality due to cardiovascular diseases, stroke, and pre-existing conditions. Deaths on duty, including deaths due to overwork, during the COVID-19 response were usually reported on web-based platforms for public recognition and solidarity. Although no official statistics are available for these casualties, a list of on-duty deaths has been made publicly available on the web by crowdsourcing.

Objective: This study aims to understand the trends and characteristics of deaths related to overwork among the frontline nonclinical essential workers participating in nonpharmaceutical interventions during the first wave of COVID-19 in China.

Methods: Based on a web-based crowdsourced list of deaths on duty during the first wave of the COVID-19 response in China, we manually verified all overwork-related death records against the full-text web reports from credible sources. After excluding deaths caused by COVID-19 infection and accidents, a total of 340 deaths related to overwork among nonclinical essential workers were attributed to combatting the COVID-19 crisis. We coded the key characteristics of the deceased workers, including sex, age at death, location, causes of death, date of incidence, date of death, containment duties, working area, and occupation. The temporal and spatial correlations between deaths from overwork and COVID-19 cases in China were also examined using Pearson correlation coefficient.

Results: From January 20 to April 26, 2020, at least 340 nonclinical frontline workers in China were reported to have died as a result of overwork while combatting COVID-19. The weekly overwork mortality was positively correlated with weekly COVID-19 cases (r=0.79, P<.001). Two-thirds of deceased workers (230/340, 67.6%) were under 55 years old, and two major causes of deaths related to overwork were cardiovascular diseases (138/340, 40.6%) and cerebrovascular diseases (73/340, 21.5%). Outside of Hubei province, there were almost 2.5 times as many deaths caused by COVID-19–related overwork (308/340, 90.6%) than by COVID-19 itself (n=120).

Conclusions: The high number of deaths related to overwork among nonclinical essential workers at the frontline of the COVID-19 epidemic is alarming. Policies for occupational health protection against work hazards should therefore be prioritized and enforced.

(J Med Internet Res 2021;23(4):e23311) doi:10.2196/23311

KEYWORDS

nonpharmaceutical interventions; on-duty deaths; COVID-19; overwork death; crowdsourced data; intervention; mortality; casualty; cross-sectional; overwork; stress

Introduction

The first wave of the COVID-19 epidemic in China was brought under control within 3 months—from mid-January 2020 (when human-to-human transmission of COVID-19 was confirmed) to the end of April 2020 [1]. The effective containment of the initial wave of COVID-19 in China was credited to not only the frontline medical response but also swift, massive, and aggressive nonpharmaceutical interventions (NPIs) [2]. Implementation of such interventions was time-sensitive and labor-intensive, demanding the continuous efforts of staff from all sectors to meet community health care and logistical needs, such as setting up checkpoints for temperature screening, conducting travel history inquiries to screen for suspected cases, and protecting incarcerated people in prisons and detention houses. However, although COVID-19 cases and related deaths among medical professionals have been well acknowledged [3], the health of nonclinical essential workers engaged in NPIs to contain the spread of COVID-19 should not be overlooked.

In China, these frontline workers have suffered from high levels of psychological and physical stress, and they have worked long hours without sufficient rest for weeks, even months, due to the rapid acceleration of the epidemic and understaffing [4]. It has long been recognized that "overwork" can kill [5], as prolonged working hours and heightened psychological stress increase the risk of coronary heart disease and stroke by inducing increased catecholamine secretion, eventually leading to increased mortality risk [6-8]. However, the hidden casualties of these deaths from overwork have not been studied. Deaths on duty during the COVID-19 response were usually reported by web-based news platforms for reasons of public recognition and solidarity, but there are no official statistics concerning these casualties. Fortunately, a list of on-duty deaths in the combat against the COVID-19 crisis has been made publicly available on the social media platform Weibo [9] by crowdsourcing and has been recognized by the Chinese public. This study aims to conduct a comprehensive search of web-based news reports to describe the trends and characteristics of deaths from overwork in the fight against the COVID-19 epidemic in China.

Methods

Based on a widely recognized list of on-duty deaths in the fight against COVID-19 in China, from January 20 to April 26, 2020, a total of 496 deaths on duty were included in this study, with no duplication [9]. All records of deaths in the line of duty were verified with the corresponding full-text, reports on the web from credible news platforms such as Xinhua News Agency, People's Daily, Sohu.com, and Sina.com, as well as official government websites (detailed sources are presented in Multimedia Appendix 1). If the on-duty death reports explicitly mentioned *overwork* and the deaths were not attributed to COVID-19 or accidents, then those deaths would be identified as *deaths related to overworks*. After excluding the deaths caused by COVID-19 or accidents, we found that 340 of the 496 (68.5%) deaths related to overwork among nonclinical essential workers were attributed to combatting the COVID-19 crisis.

We coded the key characteristics of the 340 deceased workers, including sex, age at death, location, causes of death, date of incidence, date of death, containment duties, working area, and occupation (see Table S1 in Multimedia Appendix 1). The date of incidence refers to the date when the frontline worker's health condition deteriorated suddenly due to overwork, and they could no longer perform their duties, whereas the date of death refers to the date when the overwork-related death occurred. If a frontline worker died from overwork immediately or was found dead, the date of incidence and the date of death would be the same. However, if the frontline worker's health condition deteriorated suddenly due to overwork, but they died only after several days of rescue and treatment at the hospital, then the date of incidence and date of death would be different. This study only used the date of incidence to analyze the trends and correlation of deaths from overwork and COVID-19 incidence, as it reflected the timely physical and psychological stress caused by the severity of the epidemic.

In addition, the number of daily deaths from overwork and daily COVID-19 incidences were aggregated into weekly COVID-19 incidence to reflect the time trends of deaths from overwork and COVID-19 incidence [10]. The temporal correlation between weekly overwork mortality and COVID-19 incidence and the spatial correlation between provincial deaths from overwork and COVID-19 case counts were also examined using Pearson correlation coefficient (r); the correlation was considered statistically significant at $P<.05$. As a sensitivity analysis, we analyzed cases in which the death occurred within 2 days of the date of incidence to examine the correlation. This study was approved by the institutional review board of Research Center for Public Health, School of Medicine, Tsinghua University, Beijing, China (THUSM/PHREC 2020400-009).

Results

The earliest deaths from overwork occurred on January 24, 2020, and during the 14 weeks from January 20 to April 26, 2020, at least 340 nonclinical frontline workers in China were reported to have died due to overwork in the fight against the COVID-19 epidemic. Both the number of COVID-19 cases (n=30,396) and the number of deaths from overwork (n=53) reached the peak in the fourth week after January 20 (Figure 1). Weekly overwork mortality was positively correlated with weekly COVID-19 cases (r=0.79, $P<.001$). A total of 86.5% (294/340) of cases died within 2 days of the incidence. As indicated by the sensitivity analysis, if only considering the deaths within 2 days of incidence, the temporal correlation between weekly overwork mortality and weekly COVID-19 cases would increase to 0.81 ($P<.001$).

Among those individuals who died due to overwork (Table 1), the mean age at death was 49.50 (SD 9.13) years, with two-thirds (230/340, 67.6%) of them under 55 years old. Two of the major specified underlying causes of death were cardiovascular diseases (138/340, 40.6%), including myocardial infarction and sudden cardiac arrest, and cerebrovascular

diseases (73/340, 21.5%), including stroke. A majority of the deaths from overwork were among men (324/340, 95.3%) in China; this reflected that the epidemic containment responsibilities were carried out by village leaders (111/340, 32.6%), police (96/340, 28.2%), and civil servants (58/340, 17.1%), which are traditionally male-dominated occupations in China. Most deceased workers were involved in community mobilization and support (217/340, 63.8%), including community closure and access control, temperature screening, travel history collection, and delivery of daily necessities, followed by public security (77/340, 22.6%). Apart from working in rural (156/340, 45.9%) and urban (132/340, 38.8%) settings, 15.3% (52/340) of the deceased workers used to work at traffic checkpoints on expressways and in prisons and detention houses.

In addition, 32 of the 340 (9.4%) deaths from overwork occurred in Hubei province, which reported the highest prevalence of COVID-19 in China. Among the remaining 308 (90.6%) deaths reported outside Hubei province, provincial deaths related to overwork were positively correlated with the number of COVID-19 cases (r=0.55, P=.002). If only considering the deaths that occurred within 2 days of incidence, the correlation coefficient remains nearly the same (r=0.55, P=.002).

Figure 1. Trends of weekly overwork-related deaths and COVID-19 incidence between January 20 and April 26, 2020, in China.

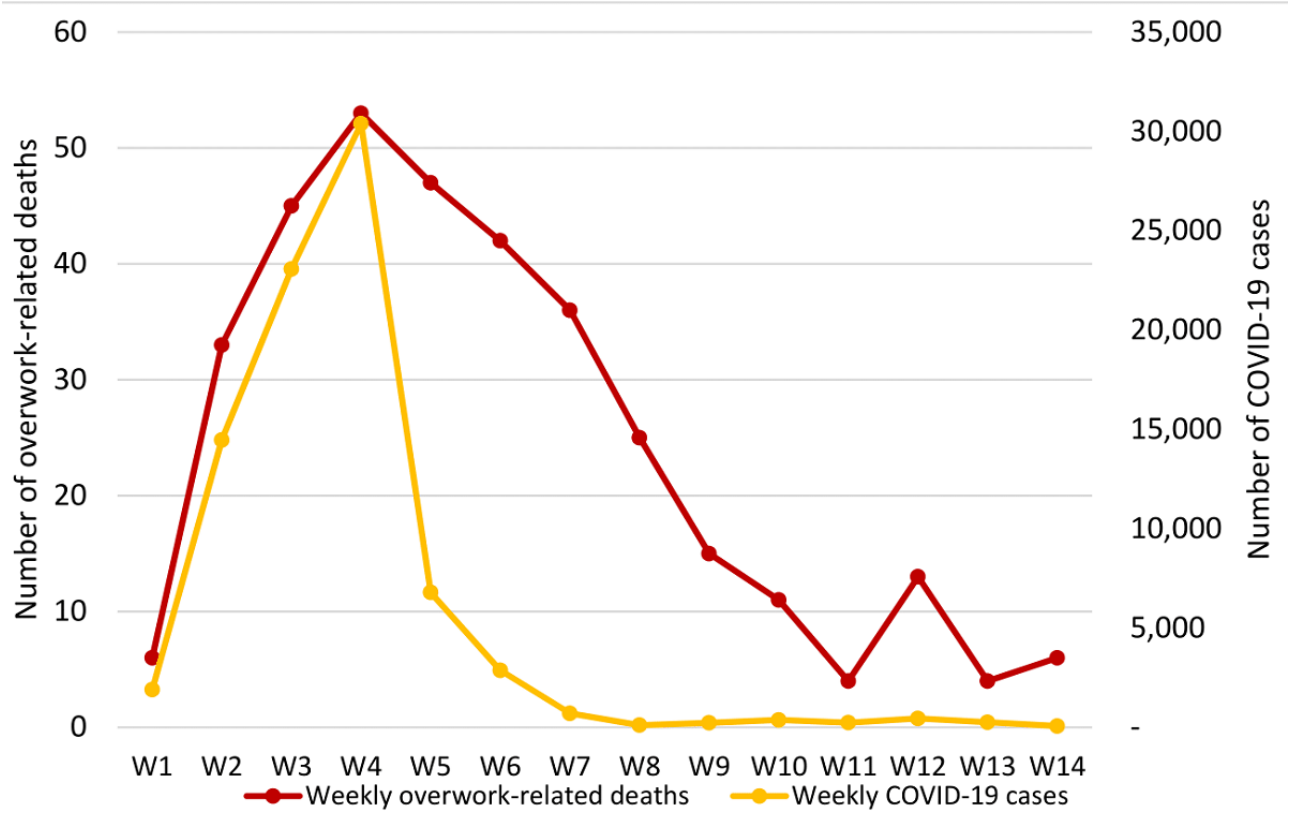


Table 1. Characteristics of documented frontline workers who died due to overwork during the emergency response to COVID-19 in China between January 20 and April 26, 2020 (N=340).

Characteristic	Value, n (%)
Sex	
Female	16 (4.7)
Male	324 (95.3)
Age range (years)	
25-34	21 (6.2)
35-44	69 (20.3)
45-54	140 (41.2)
55-64	92 (27.1)
65-71	15 (4.4)
Unspecified	3 (0.9)
Working area	
Rural	156 (45.9)
Urban	132 (38.8)
Other[a]	52 (15.3)
Underlying cause of death	
Cardiovascular diseases	138 (40.6)
Cerebrovascular diseases	73 (21.5)
Other specific causes[b]	27 (7.9)
Unspecified	102 (30)
Epidemic containment duty	
Community mobilization and support	217 (63.8)
Public security	77 (22.6)
Traffic checkpoint and control	30 (8.5)
Other[c]	17 (5)
Occupation	
Village leaders	111 (32.6)
Police	96 (28.2)
Civil servant	58 (17.1)
Volunteer	32 (9.4)
Other public sectors[d]	26 (7.1)
Corporate employee	17 (5)
Location	
Hubei province	32 (9.4)
Outside Hubei province	308 (90.6)
Difference between date of incidence and data of death (days)	
0	244 (71.8)
1	36 (10.6)
2	14 (4.1)
>2	46 (13.5)

[a]Other working areas include prisons, detention houses, and traffic checkpoints at the expressway.

[b]Other causes of death include acute hepatic failure, acute pancreatitis, and pre-existing conditions.

[c]Other epidemic containment duties include logistics, electricity or telecommunication maintenance, and health communication.

[d]Other public sectors include schools and institutions funded by the government.

Discussion

Principal Findings

This study utilized the manually verified crowdsourced data available on the internet to illustrate the trends and characteristics of deaths from overwork, which were casualties of the effort to contain the COVID-19 epidemic in China. The 340 deaths from overwork reveal a hidden "epidemic within the epidemic" that has not been documented thus far and has further sounded the alarm of fatigue and occupational burnout among nonclinical frontline workers. The severity of the epidemic presented a considerable health burden to those who were involved in the battle against COVID-19, and it explains why the peak of deaths from overwork was synchronized with the peak of the COVID-19 pandemic. Notably, outside Hubei province, there were over 2.5 times as many deaths caused by COVID-19–related overwork (308/340, 90.6%) than deaths caused by the disease itself (n=120) [10], and these provincial deaths related to overwork were correlated with COVID-19 case counts, signaling the high intensity of NPIs to curb the COVID-19 epidemic.

This epidemic embodies the health threats that all nonclinical essential workers faced. First, nonclinical essential workers had mental distress due to the fear of contracting COVID-19. Unlike ordinary residents under home quarantine, nonclinical essential workers had to work outside to perform their duties, especially screening for suspected COVID-19 cases. During the early stages of the epidemic in China, all personal protective equipment (PPE) was prioritized for clinical staff [11]; thus, nonclinical workers faced a shortage of PPE, which may have heightened their fears of COVID-19 infection. Second, most of the containment strategies were implemented by public sector employees, and they were accountable by the law and regulation for containing the epidemic. Thus, they experienced considerable work stress and workload due to understaffing, especially in provinces other than Hubei. The Chinese central government dispatched many clinical and nonclinical workers from other provinces to support Hubei province [12], whereas for other provinces, they had to manage to contain the transmission of COVID-19 largely on their own. Nonclinical essential workers worked overtime at the peak of the epidemic, which may have increased the risk of death due to cardiovascular and cerebrovascular diseases [6,7]. Third, due to understaffing, many workers with pre-existing conditions joined the workforce in the intensive battle against COVID-19. Those workers did not have enough time to rest and recover, and their pre-existing poor health conditions may have deteriorated and contributed to the observed deaths from overwork.

The high number of deaths from overwork among essential workers serving at the frontline during the COVID-19 epidemic is alarming. Although protection for health care workers in the line of duty against COVID-19 has been a global focus [11,13], policies for occupational health protection against work hazards such as adequate training, sufficient PPE, and proper working shifts, should be prioritized and enforced. Counseling services should also be provided to mitigate the psychosocial impacts. The possibility of a second wave of COVID-19 cannot be ruled out [14], so NPIs will likely remain in place until there is a breakthrough in treatment or a vaccination becomes widely available.

Limitations

Our study also has some limitations. First, the deaths related to overwork were identified based on news reports on the web, not from the hospital health records; hence, some misclassification of data may exist. We verified the full-text reports of all deaths from overwork by using credible and official sources to minimize this misclassification. Second, the length and details of the individual death reports varied and, as such, some characteristics could not be retrieved and coded. Finally, the number of deaths from overwork was likely to be underestimated, as not every incident was reported on the internet.

Conclusions

This cross-sectional study based on web-based crowdsourced data emphasizes that deaths related to overwork among nonclinical essential workers should be acknowledged. Corresponding policies for occupational health protection against deaths due to overwork should be implemented. Nonclinical essential workers are the unsung heroes, and their safety and well-being should be prioritized as they constitute society's frontline of defense against COVID-19.

Acknowledgments

We would like to thank Jianing Ren, Xinran Qi, Jie He, Yu Wang, Chunyu Liu, Siyu Zou, Chunxiao Peng, Luoqi Yuan, Jiashu Shen, Zhao Jin, Yiqian Xin, and Fan Yang for their diligent work on the full-text verification of reports on deaths related to overwork. This study has been funded by National Nature Science Foundation of China (72074130) and Spring Breeze Foundation of Tsinghua University (20203080035).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Table S1. Deidentified individual characteristics of documented frontline workers who died due to overwork during the emergency response to COVID-19 between January 20 and April 26, 2020, in China.
[XLSX File (Microsoft Excel File), 38 KB - jmir_v23i4e23311_app1.xlsx]

References

1. Fighting Covid-19: China in Action. The State Council Information Office – The People's Republic of China. 2020 Jun 07. URL: http://english.scio.gov.cn/whitepapers/2020-06/07/content_76135269.htm [accessed 2020-07-01]
2. Lai S, Ruktanonchai NW, Zhou L, Prosper O, Luo W, Floyd JR, et al. Effect of non-pharmaceutical interventions to contain COVID-19 in China. Nature 2020 Sep;585(7825):410-413 [FREE Full text] [doi: 10.1038/s41586-020-2293-x] [Medline: 32365354]
3. Zhan M, Qin Y, Xue X, Zhu S. Death from Covid-19 of 23 health care workers in China. N Engl J Med 2020 Jun 04;382(23):2267-2268. [doi: 10.1056/nejmc2005696]
4. Sun S. Heroes, just like this–portraits of heroes who sacrificed their lives in the fight against the epidemic Webpage in Chinese. Xinhua News Agency. URL: http://www.xinhuanet.com/politics/2020-04/03/c_1125812075.htm [accessed 2020-05-15]
5. Michie S, Cockcroft A. Overwork can kill. BMJ 1996 Apr 13;312(7036):921-922 [FREE Full text] [doi: 10.1136/bmj.312.7036.921] [Medline: 8616291]
6. Kivimäki M, Jokela M, Nyberg ST, Singh-Manoux A, Fransson EI, Alfredsson L, et al. Long working hours and risk of coronary heart disease and stroke: a systematic review and meta-analysis of published and unpublished data for 603 838 individuals. The Lancet 2015 Oct;386(10005):1739-1746. [doi: 10.1016/s0140-6736(15)60295-1]
7. Backé EM, Seidler A, Latza U, Rossnagel K, Schumann B. The role of psychosocial stress at work for the development of cardiovascular diseases: a systematic review. Int Arch Occup Environ Health 2012 Jan;85(1):67-79 [FREE Full text] [doi: 10.1007/s00420-011-0643-6] [Medline: 21584721]
8. Ke DS. Overwork, stroke, and karoshi-death from overwork. Acta Neurol Taiwan 2012 Jun;21(2):54-59 [FREE Full text] [Medline: 22879113]
9. Pay tribute! 496 people have given their lives to fight the epidemic. Webpage in Chinese. Weibo. 2020 Apr 29. URL: https://weibo.com/1887344341/IFBwOa5Ew [accessed 2020-05-05]
10. Daily briefing on novel coronavirus cases in China. National Health Commission of the People's Republic of China. 2020. URL: http://en.nhc.gov.cn/DailyBriefing.html [accessed 2020-05-15]
11. COVID-19: protecting health-care workers. The Lancet 2020 Mar;395(10228):922. [doi: 10.1016/s0140-6736(20)30644-9]
12. Zhang Z, Yao W, Wang Y, Long C, Fu X. Wuhan and Hubei COVID-19 mortality analysis reveals the critical role of timely supply of medical resources. J Infect 2020 Jul;81(1):147-178. [doi: 10.1016/j.jinf.2020.03.018] [Medline: 32209384]
13. Sim MR. The COVID-19 pandemic: major risks to healthcare and other workers on the front line. Occup Environ Med 2020 May;77(5):281-282. [doi: 10.1136/oemed-2020-106567] [Medline: 32238444]
14. Xu S, Li Y. Beware of the second wave of COVID-19. The Lancet 2020 Apr;395(10233):1321-1322. [doi: 10.1016/s0140-6736(20)30845-x]

Abbreviations

NPIs: nonpharmaceutical interventions
PPE: personal protective equipment

Edited by G Eysenbach; submitted 07.08.20; peer-reviewed by S Lalmuanawma; comments to author 12.02.21; revised version received 13.03.21; accepted 03.04.21; published 20.04.21.

Please cite as:
Wang Z, Lin L, Guo Y, Xiong H, Tang K
The Uncounted Casualties of a Hidden COVID-19 Epidemic in China: Cross-sectional Study on Deaths Related to Overwork
J Med Internet Res 2021;23(4):e23311
URL: https://www.jmir.org/2021/4/e23311
doi:10.2196/23311
PMID:33822735

Original Paper

Novel Predictors of COVID-19 Protective Behaviors Among US Adults: Cross-sectional Survey

Ken Resnicow[1,2], PhD; Elizabeth Bacon[1], MPH; Penny Yang[1], BSc; Sarah Hawley[1,3,4], PhD; M Lee Van Horn[5], PhD; Lawrence An[1,3], MD

[1]Center for Health Communications Research, Rogel Cancer Center, University of Michigan, Ann Arbor, MI, United States

[2]Department of Health Behavior & Health Education, School of Public Health, University of Michigan, Ann Arbor, MI, United States

[3]Division of General Medicine, School of Medicine, University of Michigan, Ann Arbor, MI, United States

[4]VA Center for Clinical Management Research, VA Ann Arbor Healthcare System, Ann Arbor, MI, United States

[5]Department of Psychology, University of New Mexico, Albuquerque, NM, United States

Corresponding Author:
Ken Resnicow, PhD
Department of Health Behavior & Health Education
School of Public Health
University of Michigan
109 Observatory Street
Room 3867 SPH I
Ann Arbor, MI, 48109
United States
Phone: 1 734 764 9494
Email: kresnic@umich.edu

Abstract

Background: A central component of the public health strategy to control the COVID-19 pandemic involves encouraging mask wearing and social distancing to protect individuals from acquiring and transmitting the virus.

Objective: This study aims to understand the psychological factors that drive adoption or rejection of these protective behaviors, which can inform public health interventions to control the pandemic.

Methods: We conducted an online survey of a representative sample of 1074 US adults and assessed three novel potential predictors of COVID-19 behaviors: trait reactance, COVID-19 conspiracy beliefs, and COVID-19 apocalypse beliefs. Key outcomes (dependent variables) included an index of COVID-19 protective behaviors, the number of trips taken from the home, and COVID-19 knowledge.

Results: In bivariate analyses, all three predictors were significantly correlated in the hypothesized direction with the three COVID-19 outcomes. Specifically, each predictor was negatively (P<.01) correlated with the COVID-19 protective behaviors index and COVID-19 knowledge score, and positively correlated with trips taken from home per week (more of which was considered higher risk). COVID-19 protective behaviors and COVID-19 knowledge were significantly lower in the top median compared to the bottom median for all three predictors. In general, these findings remained significant after adjusting for all novel predictors plus age, gender, income, education, race, political party, and religiosity. Self-identified Republicans (vs other political affiliations) reported the highest values for each of the novel predictors.

Conclusions: This study can inform the development of health communication interventions to encourage the adoption of COVID-19 protective behaviors. Interestingly, we found that higher scores of all three novel predictors were associated with lower COVID-19 knowledge, suggesting that lack of an accurate understanding of the virus may be driving some of these attitudes; although, it is also possible that these attributes may interfere with one's willingness or ability to seek and absorb accurate health information. These individuals may be particularly immune to accepting new information and yielding their beliefs. Health communication professionals may apply lessons learned from countering similar beliefs around climate change and vaccine hesitancy. Messages designed for individuals prone to reactance may be more effective if they minimize controlling language and emphasize the individual's independence in adopting these behavioral recommendations. Messaging for those who possess conspiracy beliefs should similarly not assume that providing evidence contrary to these beliefs will alone alter behavior. Other communication techniques such as *rolling with resistance*, a strategy used in motivational interviewing, may be helpful. Messaging

for those with apocalyptic beliefs may require using religious leaders as the message source and using scripture that would support the adoption of COVID-19 protection behaviors.

(J Med Internet Res 2021;23(4):e23488) doi:10.2196/23488

KEYWORDS

COVID-19; protective behavior; psychological predictors; reactance; conspiracy beliefs; public health; health communication; communication; protection; behavior; psychology

Introduction

A central component of the public health strategy to control the spread of the COVID-19 pandemic and the associated morbidity and mortality is to encourage behaviors that protect individuals from acquiring and transmitting the virus. Key protective behaviors that have been recommended by US and global health organizations such as the World Health Organization and the Centers for Disease Control and Prevention include consistent wearing of a facial mask, social distancing, handwashing, and avoiding large gatherings [1]. Until a vaccine or more effective treatments become widely available, behavior change will remain the core of the public health strategy.

Understanding individual-level attributes that are associated with adoption of these behaviors is critical to controlling the spread of COVID-19. To date, most of the research on COVID-19 protective behaviors has focused on demographic variables such as gender and race as well as social cognitive variables such as perceived risk and knowledge regarding the virus [2-5]. Less attention has been given to personality factors and constructs beyond the traditional models of health behavior (eg, perceived risk). Understanding the psychological traits that drive adoption of these protective behaviors can inform social marketing campaigns and behavior change interventions.

To address these gaps in understanding about what drives individual protective behavior choices, we identified three novel predictors based on both theoretical and empirical considerations. Psychological *reactance* theory, originally proposed by Brehm and Brehm [6] posits that when an individual's sense of behavioral freedom is threatened, the individual is motivated to restore the perceived loss of freedom by psychologically and behaviorally rejecting the behavior, even if the behavior may be in their best interest. *Conspiracy beliefs* can be defined as unsubstantiated, implausible assertions that malevolent and hidden forces control our social institutions, and these nefarious forces secretly plot major events such as 9/11; covering up alien visitations; and, in the case of this study, the creation and spread of COVID-19. Often these beliefs reject other simpler explanations that are more probable and plausible [7]. Conspiracy beliefs have previously been found to be associated with lower adoption of protective behaviors such as vaccination and condom use [7-11]. With regard to COVID-19, a few studies have found a negative association between conspiracy beliefs, both measured as a global trait and specific to COVID-19, and positive attitudes toward and adoption of COVID-19 protective behaviors [4,10,12-14]. Finally, we were interested in the potential impact of *apocalyptic beliefs*. Apocalypticism is the generally religious belief that the end of the world is imminent [15], and civilization will soon come to a tumultuous end due to some catastrophic global event such as war, famine, or disease and more recently global warming and the COVID-19 pandemic. These beliefs often include some sense of divine punishment for immorality or disobedience and spare the *righteous* who obey God's rules. For some Christians, these beliefs include the rapture, when both living and dead believers will ascend in to heaven to meet Jesus Christ at the Second Coming. Our underlying assumption is that individuals who believe in the apocalypse will be less likely to adhere to public health recommendations around COVID-19, in part because they welcome the end of days and the Second Coming of Christ. Although we could find no studies reporting the association between apocalyptic beliefs and COVID-19 protective behaviors, conceptual and empirical work has shown that such beliefs may impact behaviors related to climate change and violence [16,17], and we suspected it might play a role in the adoption of COVID-19 protective behaviors.

We conducted a national online survey and report here the association between these three potential novel predictors—(1) trait reactance, (2) COVID-19 conspiracy beliefs, and (3) COVID-19 apocalypse beliefs—and COVID-19 protective behaviors.

Methods

Sampling

Surveys were completed online using a sample provided by Dynata [18] between May 19-22, 2020. Dynata's research panel comprises an opt-in list of over 60 million individuals worldwide. For this study, we requested a nationally representative sample of 1000 US adults 18 years or older. Quotas were used to approximate national rates for age, gender, race, income, and region for the overall US population. Our survey was conducted as open enrollment, whereby eligible panel members who log in to the Dynata website were offered a chance to partake in this survey. Surveys were completed using the Qualtrics online platform. Participants received a modest payment from Dynata for completing their survey. Dynata incentives vary based on individual preferences and include cash, frequent traveler or customer loyalty points, or a donation to a charity. The reward value is based on the amount of effort required and the population surveyed. Regardless of the type of incentive, the value is the same for every respondent in a given study. In this study, the value was US $1.00. The full survey assessed a range of individual and household characteristics and attitudes and behaviors related to the COVID-19 pandemic as well as demographics. Other than screener items, no survey items were required to progress (ie, no strict validation was used). After excluding implausible

values (<10 minutes and >2 hours), the mean minutes to complete the survey was 25.3 (range 10.1-117.1) minutes.

A total of 2272 individuals clicked on the invitation link, 187 did not complete the age screener or consent, and 609 were ineligible or refused consent. This yielded 1476 surveys from age-eligible consenting individuals. To ensure the quality of the respondent data, we excluded 402 of the 1476 surveys based on two criteria. First, we excluded 375 surveys from individuals who completed the full survey in less than 10 minutes. We considered 10 minutes the minimum time required to complete a valid survey. Second, we excluded 27 surveys for individuals who answered all items within a 16-item block of items assessing attitudes toward the pandemic with an identical response. This is the equivalent of clicking down an entire column (eg, all strongly agree or disagree) for all items. Because some of the 16 items in this section were worded in the positive direction (eg, social distancing has slowed the spread of COVID-19) and others in the negative direction (eg, social distancing is not really doing much good), we considered these *response set* patterns contradictory and therefore an indication that the validity of that survey was suspect. After applying these exclusions, 1074 surveys remained for the present analyses.

Measures

Primary Predictor Measures (Independent Variables)

Trait Reactance

We selected 5 items from the widely used Hong Reactance scale [19-23]. The scale measures trait reactance rather than reactance specific to COVID-19 recommendations. Each item was answered along a five-point continuum ranging from strongly disagree to strongly agree. Internal consistency in our sample was 0.87. The five items, averaged to create a mean scale, were:

1. I become angry when my freedom of choice is restricted.
2. Regulations trigger a sense of resistance in me.
3. When something is prohibited, I usually think, “That’s exactly what I am going to do.”
4. It disappoints me to see others submitting to society’s standards and rules.
5. Advice and recommendations usually induce me to do just the opposite.

COVID-19 Conspiracy Beliefs

We developed a brief, three-item scale based on prior studies of COVID-19 and other health issues [7,14]. The scale is intended to measure conspiracy beliefs regarding COVID-19 rather than a generalized conspiracy trait or worldview [14]. Each item was answered along a five-point continuum ranging from definitely false to definitely true. Internal consistency in our sample was 0.74. The three items, averaged to create a mean scale, were:

1. The real truth about COVID-19 is being kept from the public.
2. People in power are using COVID-19 as an excuse to monitor and control the public.
3. The media is making COVID-19 seem more dangerous than it really is.

COVID-19 Apocalypse Beliefs

We developed a new brief scale informed by theological definitions and prior related work [15,17]. Each item was answered along a five-point continuum ranging from strongly disagree to strongly agree. Internal consistency in our sample was 0.92. The three items, averaged to create a mean scale, were:

1. The COVID-19 pandemic is a sign that the apocalypse is coming.
2. The COVID-19 pandemic is a sign that Jesus will soon be returning.
3. The COVID-19 pandemic is a sign that the rapture is coming.

Outcomes Measures (Dependent Variables)

Adoption of Positive COVID-19 Protection Behaviors

We examined the frequency of five self-reported behaviors over the past week, all of which are recommended for reducing the risk of transmitting or acquiring COVID-19 [1]. For each item the responses were: rarely or never (coded 1), some of the time (coded 2), most of the time (coded 3), almost all of the time (coded 4), and all of the time (coded 5). The values of 1-5 for each item were summed to form an index score with a range of 5-25. The alpha value for the five behaviors was .84 in this sample.

1. Staying home as much as possible.
2. Wearing a mask or face covering when I go out of the house.
3. Staying at least 6 feet (about 3 steps) away from people I do not live with.
4. Avoiding gatherings or groups of other people.
5. Keeping my hands clean.

COVID-19 Knowledge

We created a seven-item scale, with each item answered definitely false to definitely true. A response was coded as correct by answering definitely or probably false for items 4, 5, 6, and 7, and definitely or probably true for items 1, 2, and 3. Correct scores were summed, yielding a total score from 0 to 7. Internal consistency for the seven items was 0.77.

1. A vaccine is not yet available for COVID-19.
2. COVID-19 can be easily spread from one person to another.
3. Many thousands of people have died from COVID-19.
4. Most people already have immunity to COVID-19.
5. Symptoms of COVID-19 are always visible.
6. There are effective treatments for COVID-19 that can cure most people.
7. Having COVID-19 is about as dangerous as having the flu.

Trips Leaving the Home

We assumed that a higher number of trips from home indicated higher risk behavior. We queried leaving the home in the past week across various types of trips. For each, we asked, “In the last seven days, how many times did you go out of your home for each of the following reasons?” Responses ranged from none to five or more times. The reasons included going to work; the grocery store or market; to get takeout from a restaurant or

fast-food location; eat at a restaurant or fast-food location; the drug store or pharmacy; seek health care; check on or help care for a vulnerable person; visit friends, family, or neighbors; take a child or minor to day care or some activity, exercise, or some other outdoor activity; and attend a gathering of 10 or more. Items were summed to create a trips leaving home index with an observed range of 0-39.

Demographic Variables (Covariates)

Gender was initially assessed with five categories: male, female, transgender (identify as male), transgender (identify as female), and other. Transgender and other were collapsed.

Political party was assessed with four categories: Republican, Democrat, Independent, and something else.

Race and ethnicity were coded as White, Black, Hispanic, multiracial, and other, which included American Indian, Asian, and other.

Income was initially assessed with 9 strata that, for ease of presentation, were collapsed into three categories: less than US $30,000; US $30,000 to US $74,999; and US $75,000 and greater.

Education was initially assessed with 10 strata that were collapsed into four categories for ease of presentation: none through high school or General Educational Development, postsecondary (trade school, some college, or associate's degree), bachelor's degree, and advanced degree (master's degree, doctoral degree, or professional degree).

Religiosity was measured with a single item: "How religious are you?" Responses ranged from not at all religious, which was coded as 1, to very religious, coded as 7.

Analyses

We first present sample demographic frequencies and means for key continuous independent and dependent variables. Next, bivariate correlations between the three novel predictors and the correlation of the predictors with the three COVID-19 outcomes are presented. For ease of presentation, a dichotomous variable was created using the median split for each of the three predictors. Using the median split for the three novel predictors, we next presented means for the three COVID-19 outcomes (which are all continuous), first unadjusted, with only the novel predictor in the model, then using a general linear model, adjusted for the other novel predictors along with age, gender, income, education, political party, and religiosity. Income, education, political party, and race, all categorical variables, were all dummy coded prior to entry into the multivariate model. All analyses were performed using SPSS version 25 (IBM Corp) [24]. This survey was approved by the University of Michigan's Institutional Review Board.

Hypotheses

The first hypothesis was that individuals that score higher on trait reactance will report fewer COVID-19 protective behaviors, higher daily excursions from their home, and lower COVID-19 knowledge.

The second hypothesis was that individuals that score higher on conspiracy beliefs regarding COVID-19 will report fewer COVID-19 protective behaviors, higher daily excursions from their home, and lower COVID-19 knowledge.

The third hypothesis was that individuals that score higher on COVID-19 apocalypse beliefs will report fewer COVID-19 protective behaviors, higher daily excursions from their home, and lower COVID-19 knowledge.

Results

Sample Description

The 1074 sample was 55% (n=573) female, 70% (n=723) White, 8% (n=84) Black, 9% (n=95) Hispanic, and 6% (n=65) multiracial. About 22% (n=225) of the sample had high school or lower education, and 47% (n=482) had at least a bachelor's degree. Income distribution was about even across the three strata. With regard to political party, 29% (n=297) identified as Republican, 38% (n=395) as Democrat, 27% (n=283) as Independent, and 6% (n=61) as other. The mean for trait reactance, COVID-19 conspiracy beliefs, and COVID-19 apocalypse beliefs were 2.4, 2.9, and 2.2, respectively. Assuming a mean value of 4 or higher (corresponding to a response of agree for reactance and apocalypse items or probably true for conspiracy items) indicates a high presence of the attribute. The prevalence was 9.8% (102/1041) for apocalypse beliefs, 20.3% (214/1052) for conspiracy beliefs, and 6.9% (72/1041) for trait reactance (Table 1).

Table 1. Sample description (N=1074).

Variable	Participants
Gender, n (%)	
Male	459 (44.3)
Female	573 (55.4)
Other	3 (0.1)
Race/ethnicity, n (%)	
White	723 (69.9)
Black	84 (8.1)
Hispanic	95 (9.2)
Multiracial	65 (6.3)
Other	67 (6.5)
Age (years), n (%)	
18-35	304 (29.5)
36-50	263 (25.6)
51-65	277 (26.9)
>65	185 (18.0)
Education, n (%)	
None through high school/GED[a]	225 (21.8)
Postsecondary (trade school/some college/associate's degree)	326 (31.6)
Bachelor's degree	310 (30.0)
Advanced degree (master's/doctoral/professional degree)	172 (16.7)
Income (US $), n (%)	
<30,000	291 (28.1)
30,000-74,999	397 (38.4)
≥75,000	346 (33.5)
Political party, n (%)	
Republican	297 (28.7)
Democrat	395 (38.1)
Independent	283 (27.3)
Something else	61 (5.9)
Variable means	
COVID-19 protective behaviors index	
Mean (SD)	20.1 (4.6)
Range	5-25
Trips leaving home per week	
Mean (SD)	6.7 (6.7)
Range	0-39
Knowledge score	
Mean (SD)	26.8 (5.3)
Range	11-35
Religiosity	
Mean (SD)	3.9 (2.1)
Range	1-7

Variable	Participants
Trait reactance	
Mean (SD)	2.4 (1.0)
Range	1-5
COVID-19 conspiracy beliefs	
Mean (SD)	2.9 (1.1)
Range	1-5
COVID-19 apocalypse beliefs	
Mean (SD)	2.2 (1.1)
Range	1-5

[a]GED: General Educational Development.

Correlations

All three predictors were significantly (P<.01) correlated in the hypothesized direction with the three COVID-19 outcomes. Specifically, each predictor was negatively correlated with the COVID-19 protective behaviors index (range –0.10 to –0.39) and COVID-19 knowledge (range –0.42 to –0. 57). All three predictors were positively correlated with trips from home per week (range 0.27-0.31; see Table 2). The three predictors were all positively correlated (P<.001). Specifically, apocalypse beliefs were correlated 0.31 and 0.33 with conspiracy beliefs and reactance, respectively. Conspiracy beliefs and reactance were correlated 0.51 (data not shown).

Table 2. Pearson correlations of novel predictors and COVID-19 outcomes (N=1074).

Variable	COVID-19 protective behaviors	Trips leaving home per week	COVID-19 knowledge score
Trait reactance	–0.38[a]	0.31	–0.54
COVID-19 conspiracy beliefs	–0.32	0.21	–0.57
COVID-19 apocalypse beliefs	–0.09[b]	0.27	–0.42

[a]All correlations were significant with a P of <.001, unless otherwise indicated.

[b]P=.002

Bivariate Means

Using the median split for the three predictors, the differences between the top and bottom half of participants were statistically significant across all three variables for each of the three COVID-19 outcomes (ie, protective behaviors index, trips from home, and COVID-19 knowledge). Specifically, the mean for the COVID-19 protective behaviors index and the COVID-19 knowledge scale were significantly lower in the top median compared to the bottom median for all three predictors. The mean trips from home was significantly higher in the top median compared to the bottom median for all three predictors. This model does not adjust for covariates or other novel predictors (see the unadjusted mean columns in Table 3).

In analyses accounting for all novel predictors simultaneously, plus age, gender, income, education, race, political party, and religiosity, the adjusted means remained significantly different for the top and bottom median for all outcomes except for apocalypse beliefs and the COVID-19 protective behaviors index. Thus overall, these findings were generally consistent with a priori hypotheses (see the adjusted mean columns in Table 3).

Table 3 also presents results by political affiliation. In adjusted analyses, respondents identifying as Democrat had the highest mean for the protective behaviors index and COVID-19 knowledge, whereas for trips from home, Republicans had the highest number. For all three predictors, values were highest among those identifying as Republican and lowest among those identifying as Democrat (data not shown).

Finally, we previously reported the means, using the median split of each of the three novel predictors, for each of the three outcomes. We also examined the outcome means using a threshold of 4 or higher compared to those scoring less than 4, for each of the predictors. For individuals with an average score of 4 or higher on the reactance scale, the means were 17.5, 11.2, and 22.0 for the protective behaviors index, trips from home, and knowledge score, respectively. For individuals with an average score of 4 or higher on the COVID-19 conspiracy scale, the means were 18.0, 7.8, and 23.2 for the protective behaviors index, trips from home, and knowledge score, respectively. For individuals with an average score of 4 or higher on the COVID-19 apocalypse scale, the means were 20.4, 11.0, and 23.5 for the protective behaviors index, trips from home, and knowledge score, respectively. Thus, in general, the pattern of results using the threshold of 4 or higher for the three predictors was similar to results using the median split (data not shown).

Table 3. Bivariate and adjusted means of COVID-19 protective behaviors, leaving home episodes, and COVID-19 knowledge by novel predictors (N=1074).

Variable	COVID-19 protective behaviors				Trips leaving home per week				COVID-19 knowledge score			
	Bivariate unadjusted, mean (SD)	*P* value	Adjusted[a], mean (SE)	*P* value	Bivariate unadjusted, mean (SD)	*P* value	Adjusted, mean (SE)	*P* value	Bivariate unadjusted, mean (SD)	*P* value	Adjusted, mean (SE)	*P* value
Trait reactance		<.001		<.001		<.001		<.001		<.001		<.001
Low[b]	21.4 (3.8)		20.9 (0.18)		5.2 (5.2)		5.9 (0.25)		28.9 (4.7)		27.8 (0.17)	
High[c]	18.2 (5.0)		18.9 (0.22)		8.5 (7.9)		7.3 (0.30)		23.9 (4.8)		25.4 (0.20)	
Conspiracy beliefs		<.001		<.001		<.001		.02		<.001		<.001
Low	21.5 (3.8)		20.9 (0.21)		5.1 (5.4)		6.0 (0.28)		29.8 (4.7)		28.7 (0.19)	
High	18.9 (4.9)		19.4 (0.19)		7.9 (7.4)		6.9 (0.26)		24.2 (4.4)		25.2 (0.18)	
Apocalypse beliefs		<.001		.89		<.001		.004		<.001		<.001
Low	20.6 (4.4)		20.1 (0.19)		5.1 (5.0)		5.9 (0.26)		28.8 (4.8)		27.6 (0.18)	
High	19.5 (4.8)		20.1 (0.21)		8.4 (7.8)		7.1 (0.29)		24.5 (5.0)		25.9 (0.20)	
Political party		<.001		<.001		.19		.01		<.001		<.001
Republican	19.3 (4.9)		19.5 (0.26)		7.1 (6.8)		7.2 (0.35)		25.4 (5.1)		26.0 (0.24)	
Democrat	21.3 (3.8)		21.0 (0.22)		6.5 (6.9)		6.4 (0.30)		28.1 (5.5)		27.4 (0.21)	
Independent	19.6 (4.9)		19.8 (0.25)		6.4 (6.5)		6.1 (0.34)		26.8 (5.1)		27.0 (0.23)	
Something else	19.0 (5.2)		19.2 (0.54)		5.3 (4.1)		4.7 (0.74)		25.6 (4.4)		26.6 (0.50)	

[a]Adjusted model includes all novel predictors plus gender, race, income, education, political party, age, and religiosity.

[b]Low indicates the bottom half of the median split.

[c]High indicates the upper half of the median split.

Discussion

Primary Findings

Our findings indicate that three psychological factors—trait reactance, COVID-19 conspiracy beliefs, and COVID-19 apocalypse beliefs—were associated with key COVID-19 outcomes, all in the hypothesized direction. In unadjusted analyses, individuals scoring higher on trait reactance, COVID-19 conspiracy beliefs, and COVID-19 apocalypse beliefs reported lower protective behaviors and lower COVID-19 knowledge. With the exception of apocalypse beliefs and the protective behaviors index, all of these bivariate associations remained significant after adjustment for age, gender, race, income, education, religiosity, and political party. Although the three novel predictors were all correlated (ranging from 0.31 to 0.51), the magnitude of these correlations suggest that they tap largely independent dimensions of personality and attitude.

These findings have significant implications for both understanding who may adopt COVID-19 protective behavior and how intervention messages might be tailored to accommodate or counter these beliefs. With regard to reactance, our findings indicate that a subset of the US population reflexively rejects the adoption of COVID-19 protective behaviors due to a general predisposition to act in the opposite direction of authority or resist any rules or public health recommendations they feel infringe upon their personal freedom. Individuals with this trait are prone to feeling their autonomy is being threatened by government regulations or public health recommendations and will restore their freedom by rejecting the recommended behavior or counterarguing with the content and source of related messages.

With regard to conspiracy beliefs, our findings indicate that a subset of the US population believes that government officials and the media inaccurately portray the *truth* about the COVID-19 epidemic, and those who possess these beliefs are less likely to adopt COVID-19 protective behaviors. Although

we did not specifically query this, it is likely this group would consider much of the mainstream media as *fake news*, making any data reports or behavioral guidelines suspect. Our findings are consistent with those of a recent study of 2501 British adults [14], which found that endorsement of COVID-19 conspiracy beliefs were significantly associated with lower self-reported adherence to recommended protective behaviors. These beliefs were also associated with general mistrust of government and authority, paranoia, vaccination conspiracy beliefs, religiosity, and climate change denial. There is a growing body of work showing an association between patterns of media consumption and endorsement of misinformation, including conspiracy beliefs. There is therefore a need to enhance media use skills and eHealth literacy in particular to help counter COVID-19–related misinformation [25,26].

Finally, individuals who believe that the COVID-19 epidemic is a signal that the end of times is nigh are less likely to adopt COVID-19 protective behaviors. Although the mechanism for this association merits elucidation, it seems plausible that individuals who believe in a coming apocalypse, particularly for those who believe they will be spared, might be less likely to adopt COVID-19 protective behaviors because, for this group, the ultimate outcome is viewed in a more positive light.

Intervention Implications

Designing tailored communication campaigns to encourage adoption of COVID-19 protective behavior for individuals who possess the beliefs we assessed poses substantial challenges. These individuals may be particularly immune to accepting new information and yielding their beliefs. The persistence of these beliefs may in part be due to having roots in deeper psychological attributes such as paranoia, mistrust, religious fundamentalism, or hostility. Health communication professionals may apply lessons learned from countering similar beliefs around climate change and vaccine hesitancy. One lesson learned from countering antivaccination beliefs is that simply providing corrective information not only may be ineffective but also could instigate further reactance, leading to entrenchment of antivaccine attitudes [27]. Thus, messages designed for individuals prone to reactance that minimize controlling language (eg, you must or you have to) and emphasize the individual's independence in adopting these behavioral recommendations may be more effective [28,29]. Messaging for those who possess conspiracy beliefs should similarly not assume that providing evidence contrary to these beliefs will alone alter behavior [29]. Other communication techniques such as *rolling with resistance*, a strategy used in motivational interviewing [30], which may manifest as agreeing or empathizing with some aspects of their belief system (eg, "the government is not always honest with the American people") should be considered. Messaging for those with apocalyptic beliefs may require using religious leaders as the message source and using scripture that would support the adoption of COVID-19 protection, perhaps as an act of kindness, respect, or following God's will. Interestingly, we found that higher scores of all three novel predictors were associated with lower COVID-19 knowledge, suggesting that lack of an accurate understanding of the virus may be driving some of these attitudes; although, it is also possible that these attributes may interfere with one's willingness or ability to seek and absorb accurate health information. Thus, efforts to improve COVID-19 knowledge among these subgroups, in particular designing messages that mitigate inherent resistance to absorbing and yielding to new information, may be an important part of the public health messaging strategy.

This paper provides insight that may inform public health communication efforts to reduce the transmission of COVID-19 among segments of the population that may not respond to general audience messages but whose adherence to recommendations are nonetheless needed to control the pandemic.

Limitations and Future Studies

Our data were cross-sectional, limiting directional inference. It is possible, for example, that behaviors might influence attitudes rather than the inverse. The sample was accrued entirely online, which introduces several potential sampling and response biases [31,32]. For example, our sample had a slightly lower percentage of non-Whites and a greater percentage of females and Democrats than the US population. Sample bias poses a lower threat to the validity of our findings, as we were primarily interested in exploring the association between variables rather than establishing the true prevalence of the attitudes and behaviors under study.

Our survey was administered on May 19-22, 2020. Late May appears to represent the high point of optimism about the pandemic in the United States compared to both the initial period of March and April and the second and third wave that occurred in October and November 2020. For example, according to the Gallup COVID Panel survey conducted late May 2020, the percentage of Americans who felt the COVID-19 situation was getting better was 42% [33]. Gallup had asked this question on several occasions between April and November, and this was approximately the peak for this variable for that period. The percentage feeling the situation was getting better began to drop sharply shortly after, hitting just 19% in June and remaining around only 20% through November 2020 [33]. Gallup similarly reported that, in early June, only 46% of Americans said they were very or somewhat worried about getting the virus compared to 58%-59% in July and August 2020 [34]. These data likely reflect optimism over what turned out to be short-lived falling infection and death rates during this period. These temporal patters are also reflected in data reported by the Pew Research Center, which found that in June 2020, 59% of Americans reported that they think the worst of the outbreak was still to come [35], compared to 73% who believed the worst was yet to come in April and 71% who reported the worst was yet to come in November 2020. Moreover, according to the Pew Research Center, the percent of Americans who reported wearing a mask or face covering all or most of the time when in stores and businesses over the past month was only 65% in June 2020 compared to 85% in August and 87% in November of 2020 [35]. How these background contextual factors and the specific timing of our survey administration might have impacted our results is difficult to determine. We acknowledge, however, that the pattern of our findings might have differed had we conducted our survey during times of greater overall

concern about COVID-19, higher rates of mask wearing, or more stringent lockdown restrictions. For example, during periods of greater perceived risk and more strict lockdowns, it is possible that the impact of the correlates we identified might have been attenuated, as they may have been overwhelmed by a greater concern of catching the disease. Alternatively, the strength of association might have been stronger in response to the increase in perceived risk and greater restrictions. We anticipate that future studies conducted during the COVID-19 period will elucidate the generalizability of our findings across the pandemic period.

There are other potential personality and attitudinal predictors of COVID-19 protective behaviors we did not measure including trait conspiracy orientation (we measured only COVID-19 conspiracy beliefs), mistrust of government, mistrust of science, paranoia, autonomy needs, independence, hostility, intelligence, media literacy, and vaccine hesitancy. How these constructs may relate to the three we focused on merits investigation. We did not query intentions regarding uptake of a potential future COVID-19 vaccine. Understanding how the factors identified here may also be associated with COVID-19 vaccine intentions merits investigation. Future studies are needed to replicate and extend our findings, including examination of how other psychosocial and demographic factors may interact with the three predictors we studied. Additionally, work is needed to determine how best to tailor messages, both on the group and individual level, based on these constructs.

Acknowledgments

This study was supported by the National Cancer Institute Grant P30CA046592-29-S4 (including Cancer Center Shared Resource: Center for Health Communications Research) and a Google Focus award (to LA).

Conflicts of Interest

None declared.

References

1. Centers for Disease Control and Prevention. 2020. URL: https://www.cdc.gov/coronavirus/2019-ncov/prevent-getting-sick/prevention.html [accessed 2020-07-29]
2. Chan E, Huang Z, Lo E, Hung K, Wong E, Wong S. Sociodemographic predictors of health risk perception, attitude and behavior practices associated with health-emergency disaster risk management for biological hazards: the case of COVID-19 pandemic in Hong Kong, SAR China. Int J Environ Res Public Health 2020 May 29;17(11):3869 [FREE Full text] [doi: 10.3390/ijerph17113869] [Medline: 32485979]
3. Howard M. Understanding face mask use to prevent coronavirus and other illnesses: development of a multidimensional face mask perceptions scale. Br J Health Psychol 2020 Nov;25(4):912-924 [FREE Full text] [doi: 10.1111/bjhp.12453] [Medline: 32588949]
4. Geldsetzer P. Use of rapid online surveys to assess people's perceptions during infectious disease outbreaks: a cross-sectional survey on COVID-19. J Med Internet Res 2020 Apr 02;22(4):e18790 [FREE Full text] [doi: 10.2196/18790] [Medline: 32240094]
5. Ko N, Lu W, Chen Y, Li DJ, Chang YP, Wang PW, et al. Cognitive, affective, and behavioral constructs of COVID-19 health beliefs: a comparison between sexual minority and heterosexual individuals in Taiwan. Int J Environ Res Public Health 2020 Jun 15;17(12):4282 [FREE Full text] [doi: 10.3390/ijerph17124282] [Medline: 32549345]
6. Brehm S, Brehm J. Psychological Reactance: A Theory of Freedom and Control. New York: Academic Press; 1981:A.
7. van Prooijen JW, van Vugt M. Conspiracy theories: evolved functions and psychological mechanisms. Perspect Psychol Sci 2018 Nov;13(6):770-788 [FREE Full text] [doi: 10.1177/1745691618774270] [Medline: 30231213]
8. Bogart L, Thorburn S. Are HIV/AIDS conspiracy beliefs a barrier to HIV prevention among African Americans? J Acquir Immune Defic Syndr 2005 Feb 01;38(2):213-218. [doi: 10.1097/00126334-200502010-00014] [Medline: 15671808]
9. Bogart L, Bird S. Exploring the relationship of conspiracy beliefs about HIV/AIDS to sexual behaviors and attitudes among African-American adults. J Natl Med Assoc 2003 Nov;95(11):1057-1065. [Medline: 14651372]
10. Sallam M, Dababseh D, Yaseen A, Al-Haidar A, Ababneh NA, Bakri FG, et al. Conspiracy beliefs are associated with lower knowledge and higher anxiety levels regarding COVID-19 among students at the University of Jordan. Int J Environ Res Public Health 2020 Jul 08;17(14):4915 [FREE Full text] [doi: 10.3390/ijerph17144915] [Medline: 32650409]
11. Ball P, Maxmen A. The epic battle against coronavirus misinformation and conspiracy theories. Nature 2020 May;581(7809):371-374. [doi: 10.1038/d41586-020-01452-z] [Medline: 32461658]
12. Georgiou N, Delfabbro P, Balzan R. COVID-19-related conspiracy beliefs and their relationship with perceived stress and pre-existing conspiracy beliefs. Pers Individ Dif 2020 Nov 01;166:110201 [FREE Full text] [doi: 10.1016/j.paid.2020.110201] [Medline: 32565592]
13. Biddlestone M, Green R, Douglas K. Cultural orientation, power, belief in conspiracy theories, and intentions to reduce the spread of COVID-19. Br J Soc Psychol 2020 Jul;59(3):663-673 [FREE Full text] [doi: 10.1111/bjso.12397] [Medline: 32592420]

14. Freeman D, Waite F, Rosebrock L, Petit A, Causier C, East A, et al. Coronavirus conspiracy beliefs, mistrust, and compliance with government guidelines in England. Psychol Med 2020 May 21:1-13 [FREE Full text] [doi: 10.1017/S0033291720001890] [Medline: 32436485]
15. Apocalypticism. Wikipedia. 2020. URL: https://en.wikipedia.org/wiki/Apocalypticism [accessed 2020-07-14]
16. Feinberg M, Willer R. Apocalypse soon? Dire messages reduce belief in global warming by contradicting just-world beliefs. Psychol Sci 2011 Jan;22(1):34-38. [doi: 10.1177/0956797610391911] [Medline: 21148457]
17. Tyminski R. Apocalyptic themes in times of trouble: when young men are deeply alienated. J Anal Psychol 2020 Feb;65(1):27-43. [doi: 10.1111/1468-5922.12572] [Medline: 31972895]
18. Dynata. URL: https://www.dynata.com/ [accessed 2021-04-09]
19. Shen L, Dillard JP. Psychometric properties of the Hong psychological reactance scale. J Pers Assess 2005 Aug;85(1):74-81. [doi: 10.1207/s15327752jpa8501_07] [Medline: 16083386]
20. Jonason P, Knowles H. A unidimensional measure of Hong's psychological reactance scale. Psychol Rep 2006 Apr;98(2):569-579. [doi: 10.2466/pr0.98.2.569-579] [Medline: 16796116]
21. Donnell A, Thomas A, Buboltz W. Psychological reactance: factor structure and internal consistency of the Questionnaire for the Measurement of Psychological Reactance. J Soc Psychol 2001 Oct;141(5):679-687. [doi: 10.1080/00224540109600581] [Medline: 11758045]
22. Hong S, Giannakopoulos E, Laing D, Williams N. Psychological reactance: effects of age and gender. J Soc Psychol 1994 Apr;134(2):223-228. [doi: 10.1080/00224545.1994.9711385] [Medline: 8201818]
23. Hong S, Page S. A psychological reactance scale: development, factor structure and reliability. Psychol Rep 2016 Sep 01;64(3_suppl):1323-1326. [doi: 10.2466/pr0.1989.64.3c.1323]
24. IBM SPSS software. IBM. 2017. URL: https://www.ibm.com/analytics/spss-statistics-software [accessed 2021-04-16]
25. Chong Y, Cheng H, Chan H, Chien W, Wong S. COVID-19 pandemic, infodemic and the role of eHealth literacy. Int J Nurs Stud 2020 Aug;108:103644 [FREE Full text] [doi: 10.1016/j.ijnurstu.2020.103644] [Medline: 32447127]
26. Brørs G, Norman C, Norekvål TM. Accelerated importance of eHealth literacy in the COVID-19 outbreak and beyond. Eur J Cardiovasc Nurs 2020 Aug;19(6):458-461 [FREE Full text] [doi: 10.1177/1474515120941307] [Medline: 32667217]
27. Sadaf A, Richards J, Glanz J, Salmon D, Omer S. A systematic review of interventions for reducing parental vaccine refusal and vaccine hesitancy. Vaccine 2013 Sep 13;31(40):4293-4304. [doi: 10.1016/j.vaccine.2013.07.013] [Medline: 23859839]
28. Vansteenkiste M, Williams G, Resnicow K. Toward systematic integration between self-determination theory and motivational interviewing as examples of top-down and bottom-up intervention development: autonomy or volition as a fundamental theoretical principle. Int J Behav Nutr Phys Act 2012 Mar 02;9:23 [FREE Full text] [doi: 10.1186/1479-5868-9-23] [Medline: 22385828]
29. Resnicow K, McMaster F. Motivational Interviewing: moving from why to how with autonomy support. Int J Behav Nutr Phys Act 2012 Mar 02;9:19 [FREE Full text] [doi: 10.1186/1479-5868-9-19] [Medline: 22385702]
30. Miller W, Rollnick S. Motivational Interviewing: Helping People Change, third edition. New York: Guilford Press; 2013.
31. Patten E, Perrin A. Who's left out in a web-only survey and how it affects results. Pew Research Center. 2015 Sep 22. URL: https://www.pewresearch.org/fact-tank/2015/09/22/who-s-left-out-in-a-web-only-survey-and-how-it-affects-results/ [accessed 2020-07-29]
32. Keeter S, McGeeney K. Coverage error in internet surveys: who web-only surveys miss and how that affects results. Pew Research Center. 2015. URL: https://www.pewresearch.org/methods/2015/09/22/coverage-error-in-internet-surveys/ [accessed 2020-07-29]
33. Saad L. Americans less amenable to another COVID-19 lockdown. Gallup. 2020 Nov 11. URL: https://news.gallup.com/poll/324146/americans-less-amenable-covid-lockdown.aspx [accessed 2020-12-08]
34. Brenan M. Roundup of gallup COVID-19 coverage. Gallup. 2020. URL: https://news.gallup.com/opinion/gallup/308126/roundup-gallup-covid-coverage.aspx [accessed 2020-12-08]
35. Funk C, Tyson A. Intent to get a COVID-19 vaccine rises to 60% as confidence in research and development process increases. Pew Research Center. 2020 Dec 03. URL: https://www.pewresearch.org/science/2020/12/03/intent-to-get-a-covid-19-vaccine-rises-to-60-as-confidence-in-research-and-development-process-increases/ [accessed 2020-12-08]

Edited by R Kukafka; submitted 13.08.20; peer-reviewed by G Williams, K Aguirre; comments to author 08.12.20; revised version received 09.12.20; accepted 06.04.21; published 20.04.21.

Please cite as:
Resnicow K, Bacon E, Yang P, Hawley S, Van Horn ML, An L
Novel Predictors of COVID-19 Protective Behaviors Among US Adults: Cross-sectional Survey
J Med Internet Res 2021;23(4):e23488
URL: https://www.jmir.org/2021/4/e23488
doi:10.2196/23488
PMID:33835930

Viewpoint

“Ask a Doctor About Coronavirus”: How Physicians on Social Media Can Provide Valid Health Information During a Pandemic

Dorthe Furstrand[1], MD, MSc; Andreas Pihl[2,3], MD; Elif Bayram Orbe[4], MD; Natasja Kingod[5], MSc; Jens Søndergaard[3], MD, PhD

[1]Section for Health Services Research, Department of Public Health, University of Copenhagen, Copenhagen, Denmark

[2]Roche Diagnostics, Copenhagen, Denmark

[3]The Research Unit of General Practice, Department of Public Health, University of Southern Denmark, Odense, Denmark

[4]Department of Internal Medicine, Herlev University Hospital, Herlev, Denmark

[5]Steno Diabetes Center Copenhagen, Copenhagen, Denmark

Corresponding Author:
Dorthe Furstrand, MD, MSc
Section for Health Services Research
Department of Public Health
University of Copenhagen
Øster Farimagsgade 5
Copenhagen, 1353
Denmark
Phone: 45 52675450
Email: furstrand@gmail.com

Abstract

In the wake of the COVID-19 pandemic, the information stream has overflowed with accurate information, misinformation, and constantly changing guidelines. There is a great need for guidance on the identification of trustworthy health information, and official channels are struggling to keep pace with this infodemic. Consequently, a Facebook group was created where volunteer medical physicians would answer laypeople’s questions about the 2019 novel coronavirus. There is not much precedence in health care professional–driven Facebook groups, and the framework was thus developed continuously. We ended up with an approach without room for debate, which fostered a sense of calmness, trust, and safety among the questioners. Substantial moderator effort was needed to ensure high quality and consistency through collaboration among the presently >200 physicians participating in this group. At the time of writing, the group provides a much-needed service to >58,000 people in Denmark during this crisis.

(J Med Internet Res 2021;23(4):e24586) doi:10.2196/24586

KEYWORDS

COVID-19; coronavirus; digital health literacy; eHealth literacy; Facebook; framework; health information; health literacy; health promotion; infodemic; infodemiology; mental health; misinformation; pandemic; patient-physician relationship; public health; social media; trust; web-based community

Introduction

The world is currently facing a COVID-19 infodemic, and the immense information load complicates the identification of trustworthy health information [1]. The substantial growth of media sources has resulted in a dilution of relevant and reliable information. Because of the novelty and worldwide spread of COVID-19, we learn more about the virus every day and witness almost daily changes to the clinical recommendations. This is confusing for the public and for health care professionals. In April 2020, Limaye et al [2] called for health care professionals to accommodate this issue by building trust on social media. This viewpoint describes a novel eHealth literacy project in Denmark, where we have built a network of >200 volunteer medical physicians who provide up-to-date knowledge to a broad audience in a Facebook group with thousands of members.

Health Information on Social Media

The Internet has become the primary source of health information for many people despite difficulties in verifying the reliability of web-based clinical evidence [3,4]. The emergence of Web 2.0 has further reoriented advancements from top-down information distribution to platforms for collaboration, dialogue, and content-sharing. One of the most

significant components of Web 2.0 is social media platforms such as Facebook with approximately 2.4 billion users [5]. Since Facebook launched its community pages function in 2010, it became possible for people to create health-related groups, which have become larger interactive communities [6]. Web-based communities provide a space for social support, sharing of experiential data, and a collective voice [7]. Layperson-friendly explanations of medical terms help patients have more positive experiences when visiting their health care providers, thus improving the physician-patient relationship [8]. However, negative reactions from health care professionals toward patients' social media activity might negatively affect patient-physician symmetries [9]; this underlines the need for health care professionals to engage on social media platforms to facilitate a positive physician-patient dynamic.

Owing to the COVID-19 pandemic, many new communities have been established on Facebook; however, this is accompanied with the risk of large-scale sharing of misinformation. The bombardment of emotionally evoking information makes it difficult for individuals to distinguish between accurate information and misinformation [10]. Although Facebook is mainly a lay-driven platform, health institutions, patients' societies, and health care professionals are increasingly creating their own groups to interact with patients on the internet. As this is a rather new phenomenon, few studies have investigated the effect of Facebook groups created and moderated by health care professionals. Initial experiences indicate difficulties in establishing, disseminating, and scaling such networks on Facebook [11].

The Facebook Group

In response to the infodemic, the Facebook group "Spørg en Læge om Coronavirus," which literally translates to "Ask a Doctor About Coronavirus," was created on March 15, 2020, by authors EBO and AP (Figure 1) [12]. This group was conceived owing to the sudden rise in waiting times on acute-care telehealth services. The continuously expanding group contains 57,000 members (approximately 1% of the entire population of Denmark) at the time of writing.

Figure 1. A screenshot from the Facebook group, "Spørg en læge om coronavirus," which literally translates to "Ask a doctor about coronavirus.".

The rules of the group are conservative. Group members can post a question, and physicians with administrator privileges can either reject or accept the request and answer the question. Questions are of different types, spanning from basic knowledge on, for example, the biological, statistical, and epidemiologic aspects of the practical applications of the proposed guidelines to more complex existential questions on fear, the future, and hope. Approximately 30% of the incoming questions are rejected because they violate the group rules, such as those on politically motivated comments, weblinks to misinformation, and personalized health information. Rejections to requests are typically accompanied by an explanation. In parallel to the main group, where laypeople ask questions and physicians provide answers, we have a closed administrative physicians-only group to debate and share guidelines, news, and warnings against popular myths.

Running a group with >200 volunteer physicians takes active management. Early on, we realized that rules and structure were essential for all member physicians to feel safe. We selected a narrow approach without room for discussion related to the answers; when a physician has answered a question, the thread is closed for further commenting and unauthorized answers are deleted. This approach was selected owing to logistic reasons: it would be too time consuming for physicians to ensure the validity and relevance of the information posted freely in a

thread. Surprisingly, the response from the users has been overwhelmingly positive. The calm and safe nature of the short threads of only physician-validated answers has proven a valuable factor, indicating the need for authoritative sources as a supplement to existing peer groups on Facebook. Unfortunately, the very rigid format of the group limits dialogue with the questioner. Furthermore, the practice has caused accusations of censorship. In the second wave of the COVID-19 pandemic, an increasing number of coronavirus skeptic members, verified through simultaneous membership in known government-critical or coronavirus skeptic groups, have joined the group and accused the group of censoring critical voices when closing threads and rejecting questions violating the rules; specifically those questions that are rejected because they contain weblinks to misinformation.

The ongoing stream of new questions provides us with a fine-tuned barometer of trending topics including circulating misinformation. Insecure members ask the group about the reliability of these topics before official information channels respond to the misinformation. We noticed that the users are highly perceptive to any inconsistency in the answers from different physicians, and we quickly identify any disruption in the sense of safety and security among the users. Fear and insecurity are directly articulated in the questions and comments, which confirms that feeling safe and in control is key to being eHealth-literate [13].

Meeting this need with consistency, empathy, and patience in our answers demands a considerable degree of active community building among the volunteer physicians. Key administrators are easily approachable for debriefing and conflict resolution among member physicians. In addition, a closed forum is used to discuss news and challenging questions and for a more unrestricted and informal conversation; this creates a safe space for critical feedback. The closed forum and personal messages among the volunteer physicians contribute to a supportive peer interaction and foster genuine interpersonal connections. We have witnessed the evolution of close relationships between physicians, even though they have never met.

The quality of the answers provided in the group is ensured largely by the community of the volunteer physicians and the approximately 30,000 daily active users. All answers are provided by named physicians, and supervision is accessible at all hours in the closed forum. Furthermore, physicians read one another's answers and users are quick to notice errors or inconsistencies in the answers. Lastly, other professionals such as pharmacists, nurses, and engineers have contributed with nuance or correction to answers related to their respective fields. This informal web of quality control has proven to ensure a generally high quality, although this has not been formally validated.

Volunteer physicians were recruited mainly via already established Facebook communities for only physicians with Danish authorization and through snowballing recruitment. Once the workflow was established and the legal implications were clarified, we had a running proof of concept. This, in turn, made recruitment of doctors easier as many of them were hesitant to participate before seeing a working setup. We recruited physicians with all experience levels and specialties. All were subsequently verified through the Danish Authorization Registry and internal Facebook relations with other physicians. We have mobilized not only working physicians, but also those excluded from contributing to the COVID-19 workforce owing to sick leave, maternity or paternity leave, pregnancy, or quarantine, all of whom worked from home on their own conditions. This provides an indispensable workforce for the group and is meaningful to physicians marginalized in this crisis.

Users were recruited through sheer diffusion. We do not have any control over the outreach of our group. We initially shared our group in our personal networks, and the group has naturally grown since then. Some answers have been shared in public and private networks, in closed groups, or on Twitter, and some have even been shared by Instagram influencers, thus expanding the awareness of the group and the service we provide. This positively reflects the degree of trustworthiness our group has achieved; however, this also highlights a point of consideration among all physicians that for all answers they provide, they should withstand publication out of context.

Especially in the beginning of the COVID-19 pandemic, a large proportion of the questioners presented with insufficient eHealth literacy to navigate the information environment of the pandemic [13]. Answers to many questions are accessible on public information websites, but the ability to obtain proper and updated information and to apply the guidelines for activities of daily life has been inadequate. As the pandemic has progressed, the difficulty level of questions has risen, which suggests that public information sources have succeeded in further disseminating basic information on the novel coronavirus among the general public.

At the other end of the spectrum, we noticed that many questioners show a high degree of health and science literacy. Their questions on complex preliminary scientific discoveries are embellished with distortionary "clickbait" headlines, indicating their limitations in interpreting such complex data. Such diversity in the questions underlines a need for physicians of different backgrounds to answer different questions in different tones and temperaments. Consequently, the challenge of maintaining consistency in the answers coexists with the large diversity in the group of physicians.

Future Perspectives

The Facebook group, "Spørg en Læge om Coronavirus," was created to counteract misinformation and foster a feeling of safety during a stressful time. The group has received no funding; nonetheless, 57,000 unique members have joined, which indicates a need for this type of health service. This group demonstrates classic one-to-one counseling in combination with one-to-many communication where answers are visible to all members, thereby illustrating the current discourse of health communication. Currently, a news article can stimulate a personal discussion in the comments section (one-to-many communication becoming one-to-one communication), while a recording of an individual consultation can be distributed on social media platforms (one-to-one communication becoming one-to-many communication). One can thus argue that we here

illustrate a future health communication premise that awaits almost every physician.

This group provides a proof of concept of a new way for health professionals to communicate and interact with the general public on social media platforms. The group is inspired by other Facebook groups and provides a template replicable in other similar or related initiatives. The Faroe Islands have already created the group, “Spyr ein Lækna um Korona,” which emulates the original initiative [14].

Our growing experience provides unique insights into the potential of Facebook in health communication; however, we cannot ignore the possibility of the distinctive information-seeking environment of the COVID-19 pandemic providing a favorable foundation for dissemination and upscaling of information. The extent and the validity of research on a health professional–driven social media platform should be further explored, preferably through a multidisciplinary approach and with an array of methodologies. Such studies could investigate developments in behavior, eHealth literacy, professional identity, and impact on physician-patient-relationships. Such a platform could also provide insights into patient adherence, navigation of health services, experiences with new treatments, management of common and rare diseases, and peer collaboration and communication. Therefore, further studies in this area would be highly relevant, and insights from this group should be further explored.

On World Patient Safety Day (September 17, 2020), the group "Spørg en læge om coronavirus” received the Danish Patient Safety Award 2020 in recognition of its effort to promote health understanding and a sense of safety through knowledge. This group was never intended to be a research study; the research potential was only recognized months after its launch. This group is part of a spontaneous health professional–driven emergency response at a unique time point where all community members have stepped up to contribute to their community in any way they can. This would limit the research potential of the data but adds a distinctive value of authenticity.

Acknowledgments

This study was made possible by the more than 200 volunteer medical physicians who contribute on a daily basis, with thousands of answers now accessible to the general public in the Facebook group “Spørg en læge om coronavirus.” The group functions solely on a volunteer basis without having received any funding. JS has sponsored the article processing fee for this viewpoint article.

Authors' Contributions

EBO, AP, and DF are core administrators and initiators (EBO and AP) of the Facebook group and conceived this viewpoint. EBO, AP, NK, and DF drafted the manuscript. JS provided insights on the direction of the content. All authors critically revised and approved the final version of the manuscript.

Conflicts of Interest

None declared.

References

1. Novel Coronavirus (2019-nCoV): Situation Report-13. World Health Organization. 2020 Feb 07. URL: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200202-sitrep-13-ncov-v3.pdf [accessed 2021-04-13]
2. Limaye RJ, Sauer M, Ali J, Bernstein J, Wahl B, Barnhill A, et al. Building trust while influencing online COVID-19 content in the social media world. Lancet Digit Health 2020 Jun;2(6):e277-e278 [FREE Full text] [doi: 10.1016/S2589-7500(20)30084-4] [Medline: 32322814]
3. Jones E, Sinclair J, Holt R, Barnard K. Social networking and understanding alcohol-associated risk for people with type 1 diabetes: friend or foe? Diabetes Technol Ther 2013 Apr;15(4):308-314. [doi: 10.1089/dia.2012.0327] [Medline: 23421853]
4. Wikgren M. Health discussions on the Internet. Libr Inf Sci Res 2001 Dec;23(4):305-317. [doi: 10.1016/S0740-8188(01)00091-3]
5. Myers J. Nearly a third of the globe is now on Facebook - chart of the day. World Economic Forum. 2019 Aug 05. URL: https://www.weforum.org/agenda/2019/08/facebook-users-social-media-internet/ [accessed 2020-06-18]
6. Kingod N. The tinkering m-patient: Co-constructing knowledge on how to live with type 1 diabetes through Facebook searching and sharing and offline tinkering with self-care. Health (London) 2020 Mar;24(2):152-168. [doi: 10.1177/1363459318800140] [Medline: 30207176]
7. Kingod N, Cleal B, Wahlberg A, Husted G. Online Peer-to-Peer Communities in the Daily Lives of People With Chronic Illness: A Qualitative Systematic Review. Qual Health Res 2017 Jan;27(1):89-99. [doi: 10.1177/1049732316680203] [Medline: 27956659]
8. van Uden-Kraan CF, Drossaert CHC, Taal E, Shaw BR, Seydel ER, van de Laar MAFJ. Empowering processes and outcomes of participation in online support groups for patients with breast cancer, arthritis, or fibromyalgia. Qual Health Res 2008 Mar;18(3):405-417. [doi: 10.1177/1049732307313429] [Medline: 18235163]

9. Smailhodzic E, Hooijsma W, Boonstra A, Langley DJ. Social media use in healthcare: A systematic review of effects on patients and on their relationship with healthcare professionals. BMC Health Serv Res 2016 Aug 26;16:442 [FREE Full text] [doi: 10.1186/s12913-016-1691-0] [Medline: 27562728]
10. Vestergaard M, Hendricks V. Fake News: når virkeligheden taber. Copenhagen: Gyldendal; 2017.
11. Cobb NK, Graham AL. Health behavior interventions in the age of facebook. Am J Prev Med 2012 Nov;43(5):571-572. [doi: 10.1016/j.amepre.2012.08.001] [Medline: 23079184]
12. Spørg en læge om coronavirus (Facebook group). Facebook. URL: https://www.facebook.com/groups/sporgomcoronavirus [accessed 2021-04-13]
13. Norgaard O, Furstrand D, Klokker L, Karnoe A, Batterham R, Kayser L, et al. The e-health literacy framework: A conceptual framework for characterizing e-health users and their interaction with e- health systems. Knowl Manag E-Learning 2015;7(4):522-540. [doi: 10.34105/j.kmel.2015.07.035]
14. Spyr ein lækna um korona (Facebook group). Facebook. URL: https://da-dk.facebook.com/groups/spyreinlakna/ [accessed 2021-04-13]

Edited by G Eysenbach; submitted 29.09.20; peer-reviewed by Z Ren, A Alasmari, J Hoogervorst; comments to author 18.11.20; revised version received 07.03.21; accepted 03.04.21; published 20.04.21.

Please cite as:
Furstrand D, Pihl A, Orbe EB, Kingod N, Søndergaard J
"Ask a Doctor About Coronavirus": How Physicians on Social Media Can Provide Valid Health Information During a Pandemic
J Med Internet Res 2021;23(4):e24586
URL: https://www.jmir.org/2021/4/e24586
doi:10.2196/24586
PMID:33835935

Review

Health Care Cybersecurity Challenges and Solutions Under the Climate of COVID-19: Scoping Review

Ying He[1], PhD; Aliyu Aliyu[2], MSc; Mark Evans[2], PhD; Cunjin Luo[3,4], PhD

[1]School of Computer Science, University of Nottingham, Nottingham, United Kingdom

[2]School of Computer Science and Informatics, De Montfort University, Leicester, United Kingdom

[3]School of Computer Science and Electronic Engineering, University of Essex, Colchester, United Kingdom

[4]Key Lab of Medical Electrophysiology, Ministry of Education, Institute of Cardiovascular Research, Southwest Medical University, Luzhou, China

Corresponding Author:
Cunjin Luo, PhD
School of Computer Science and Electronic Engineering
University of Essex
Wivenhoe Park
Colchester, CO4 3SQ
United Kingdom
Phone: 44 7493622995
Email: cunjin.luo@essex.ac.uk

Related Article:

This is a corrected version. See correction statement: https://www.jmir.org/2021/4/e29877

Abstract

Background: COVID-19 has challenged the resilience of the health care information system, which has affected our ability to achieve the global goal of health and well-being. The pandemic has resulted in a number of recent cyberattacks on hospitals, pharmaceutical companies, the US Department of Health and Human Services, the World Health Organization and its partners, and others.

Objective: The aim of this review was to identify key cybersecurity challenges, solutions adapted by the health sector, and areas of improvement needed to counteract the recent increases in cyberattacks (eg, phishing campaigns and ransomware attacks), which have been used by attackers to exploit vulnerabilities in technology and people introduced through changes to working practices in response to the COVID-19 pandemic.

Methods: A scoping review was conducted by searching two major scientific databases (PubMed and Scopus) using the search formula "(covid OR healthcare) AND cybersecurity." Reports, news articles, and industry white papers were also included if they were related directly to previously published works, or if they were the only available sources at the time of writing. Only articles in English published in the last decade were included (ie, 2011-2020) in order to focus on current issues, challenges, and solutions.

Results: We identified 9 main challenges in cybersecurity, 11 key solutions that health care organizations adapted to address these challenges, and 4 key areas that need to be strengthened in terms of cybersecurity capacity in the health sector. We also found that the most prominent and significant methods of cyberattacks that occurred during the pandemic were related to phishing, ransomware, distributed denial-of-service attacks, and malware.

Conclusions: This scoping review identified the most impactful methods of cyberattacks that targeted the health sector during the COVID-19 pandemic, as well as the challenges in cybersecurity, solutions, and areas in need of improvement. We provided useful insights to the health sector on cybersecurity issues during the COVID-19 pandemic as well as other epidemics or pandemics that may materialize in the future.

(J Med Internet Res 2021;23(4):e21747) doi:10.2196/21747

KEYWORDS

health care; security incidents; root causes; cybersecurity challenges; cybersecurity solutions; COVID-19; pandemics

Introduction

Background

COVID-19 has been an unprecedented challenge for the global health care system. It has further challenged the resilience of the health information system, which has affected our ability to achieve the global goal of health and well-being. The sector has become a primary target of adapted cybersecurity attacks [1,2]. To manage the pandemic and this extraordinary situation, the health sector has shifted its focus from the security of their systems and practices to their primary duty of delivering health care in order to save lives, placing themselves in a vulnerable situation. Attackers are taking advantage of the COVID-19 pandemic and have launched a number of cyberattacks against health care organizations [3-8]. Recent cyberattacks have impacted health care organizations such as Brno University Hospital [3], the US Department of Health and Human Services [4], the World Health Organization (WHO) [5], Gilead Sciences, Inc [6], hospitals in Romania [7], as well as the general supply chain of the health sector [8]. The health sector must be prepared to counteract cyberattacks in order to protect the availability of essential health care services as well as the confidentiality and integrity of health care information.

Cybercrime adapts to changes in the world situation very quickly. At the beginning of an escalation in the COVID-19 pandemic, malware cyberattackers identified common vulnerabilities and adapted their attacks to exploit these vulnerabilities. The current situation in the United Kingdom and worldwide provides a fertile breeding ground for various cyberattacks [9]. Cyberattackers are leveraging the increased reliance on remote working, decreased mobility, and the closure of borders between different countries, and the heightened demand for personal protective equipment (PPE) such as masks and gloves. The complex health care supply chain is also a target [10]. As a result, greater fear, uncertainty, and doubt is being experienced by the general population.

Rationale

There is some research reviewing the literature on cybersecurity in the health sector. Jalali et al [11] performed a systematic review of the literature on cybersecurity response plans in health care. Coventry et al [12] conducted a narrative review on trends in cyber threats and ways forward in the health sector. Kruse et al [13] systematically reviewed health care–related cyber threats and trends. Offner et al [14] reviewed cyber threats and mitigation strategies among Australian health care organizations. Sardi et al [15] performed a systematic review of cyber risk in health facilities. However, there is limited research on an in-depth review and analysis of key cybersecurity challenges and solutions, specifically in the health sector, in the context of a pandemic situation such as COVID-19.

Objective

Through a scoping review, this paper aims to identify the most prominent and significant methods of attack and threats that have affected the health sector during the COVID-19 pandemic, cybersecurity challenges, solutions, and areas that require further improvement. This research covers not only security-related matters as a result of the COVID-19 pandemic but also discusses inherent security challenges in health information systems that can be potentially exploited by attackers during the COVID-19 pandemic. It has implications for the whole spectrum of the health sector as a result of the increase in cybersecurity risks such as phishing, ransomware, and distributed denial-of-service (DDoS) attacks during the coronavirus crisis and in the long term.

Methods

Protocol and Registration

The review was performed according to the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) checklist, proposed by the Joanna Briggs Institute [16]. The aim of this review is to identify health sector cyberattacks, security challenges, and solutions. Before undertaking this review, a protocol was created detailing sources of information, search strategies, eligibility criteria, source selection, and data charting processes. The PRISMA-ScR checklist is presented in Multimedia Appendix 1.

Information Sources

A search of two major scientific databases (PubMed and Scopus) was performed to identify relevant articles. These include both original research articles and review articles.

Search

The search formula “(covid OR healthcare) AND cybersecurity” was used to search for articles. The articles identified should have either a COVID-cybersecurity core or a healthcare–cybersecurity core.

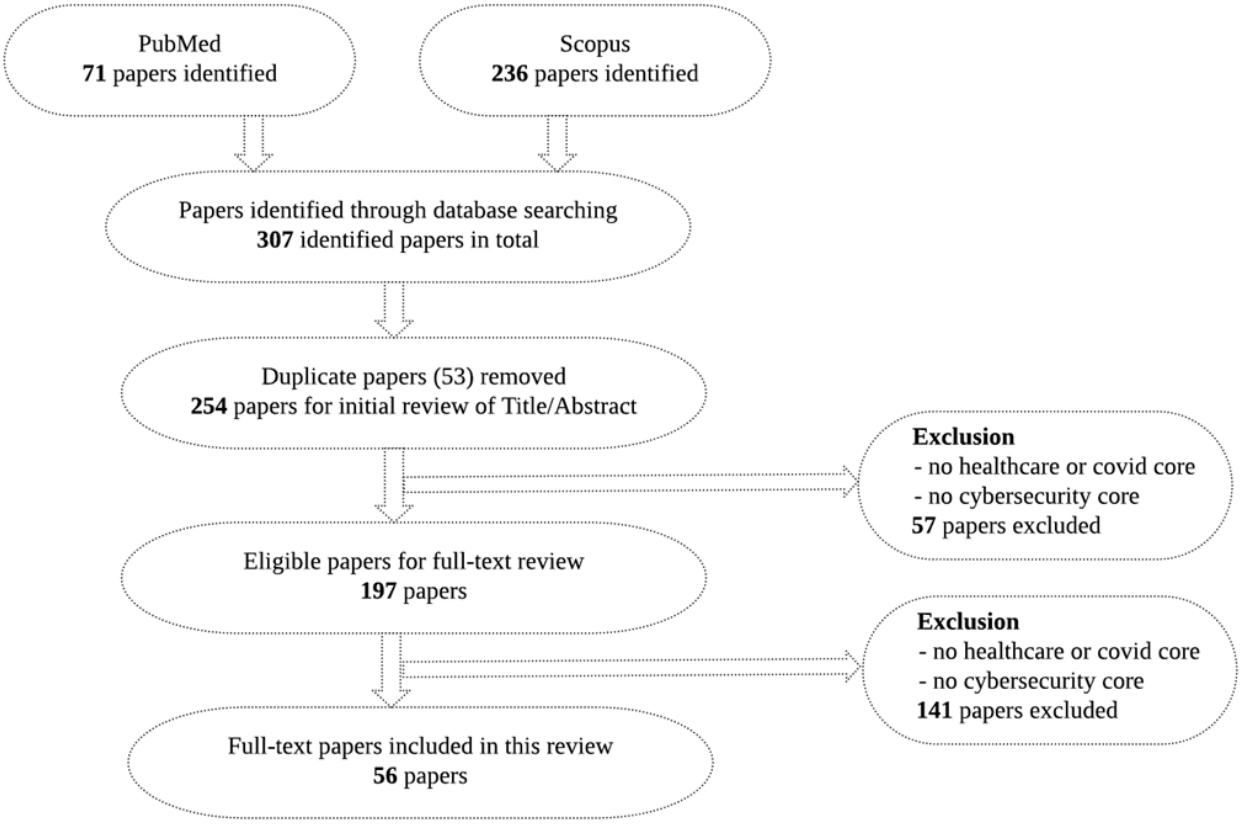
Eligibility Criteria

Only articles in English published in the last decade were included (ie, 2011-2020) in order to focus on current issues, challenges, and solutions. Reports, news articles, or websites were also included only when they are related directly to previously published work, or they were the only currently available information source at the time of manuscript preparation. Inclusion criteria were as follows: (1) relevance to health care cybersecurity and (2) coverage of well-discussed cybersecurity issues, challenges, and solutions.

Selection of Sources of Evidence

The selection process is illustrated in Figure 1. The results of the search were exported to the EndNote library. The title and abstract of each paper were analyzed by 2 of the authors to assess eligibility. In cases in which this was not obvious, all 4 authors examined the paper and, when necessary, read it to assess relevance. A total of 307 identified papers were screened and 53 duplicates were removed. An additional 57 papers were excluded for not focusing on the healthcare–cybersecurity core or the COVID-cybersecurity core in the abstract. Another 197 papers were excluded for lacking these cores in the full text. In total, 56 papers were included in the review.

Figure 1. Flowchart showing the article identification and selection process.



Data Charting Process

The data were extracted and stored in a standardized Microsoft Excel (Microsoft Corp) form. This was an iterative process whereby the charting table is continually updated. Data charting was carried out both independently and collectively by at least two authors to ensure the quality of the extracted key findings from the literature before being used in the analysis.

Data Items

Key data items, including title, abstract, authorship, aims, key findings related to the review objectives, evidence document, document type, year of publication, and location, were extracted.

Critical Appraisal Within Sources of Evidence

Although the Joanna Briggs Institute suggests that the critical appraisal is usually not needed for a scoping review, we had at least 2 authors check the quality of the source of evidence to ensure they were relevant, up to date, and from reputable sources. In cases in which this was not obvious, all 4 authors assessed the sources.

Synthesis of Results

By aggregating information from the selected literature, the results were analyzed and qualitatively presented in both tabular and descriptive formats (grouped into themes), which aligned with the objective and scope of the review.

Results

Four themes were observed across the selected literature: (1) health sector condition changes due to COVID-19, (2) health care cyberattacks during the COVID-19 pandemic, (3) health care cybersecurity challenges, and (4) health care cybersecurity controls.

Health Sector Condition Changes Due to COVID-19

The findings pertaining to changes in conditions in the health sector as a result of COVID-19 are summarized in Table 1. The main changes to health services caused by the COVID-19 pandemic include decreased mobility, border closures, and the increasing reliance on remote work, often carried out with little previous experience and planning. These conditions have made the health sector more vulnerable to potential cyberattacks [1,2,17].

Table 1. Health sector condition changes due to COVID-19.

Changes	Reference
The decreased mobility and border closures, and the increasing reliance on remote work, create challenges to health sector	Hakak et al [1], Williams et al [2], Schneck [17]
New technologies such as eConsultation services for patients and electronic multidisciplinary teams leave users open to a variety of attacks	Weil and Murugesan [18]
Health service staff often have limited experience in working remotely, leaving the sector vulnerable to cyberattacks, such as malwares	Boddy et al [9], Offner et al [14], Jalali et al [19], Hoffman [20], Ronquillo et al [21]
The health care industry significantly lags behind other industries in terms of cybersecurity and digital literacy is lacking among staff working from home, making it a prominent target	Sardi et al [15], Kim et al [22]
The increase in demand for certain goods such as PPE[a] makes health services and governments exposed to digital scams such as luring emails with the intention of stealing sensitive information	Schneck [17]

[a]PPE: personal protective equipment.

As health staff and patients are restricted in terms of movement due to the lockdown, the decrease in mobility and border closures make individuals and organizations turn to technology to provide essential health services such as appointments, diagnosis, and even operations. Examples are the use of eConsultation (electronic consultation) services for patients and electronic multidisciplinary teams. Although these technologies have their advantages, they leave users and receivers of these technologies open to a variety of attacks such as phishing campaigns and ransomware attacks [18].

Furthermore, health services staff often have limited previous experience with remote working and with planning for this change, which leaves the sector vulnerable to cyberattacks [9,14,19]. As health services make use of a variety of medical devices, interconnectivity and interoperability create issues as they are now being accessed from outside health services' internal network perimeter. The medium and mode of access creates problems as access to the sensitive parts of health services can be reached via unsecured network connections or unpatched systems by staff working remotely [19]. In addition, some medical devices use off-the-shelf software, such as commercial operating systems (eg, older versions of Windows). These systems are vulnerable to a large variety of threats such as malware, ransomware, etc [20,21]. Overall, the health care industry significantly lags behind other industries in terms of cybersecurity and coupled with a lack of digital literacy among staff mostly working from home, makes it a prominent target [15,22].

Additionally, the increase in demand for certain goods such as PPE and other protective merchandise such as masks, gloves, etc, are exposing health services and even governments to digital scams, especially in the form of phishing attacks. As health services are in need of these essential items, they can be targeted by adversaries via luring emails with the intention of stealing sensitive information [17].

Health Care Cyber Attacks During the COVID-19 Pandemic

Multiple cyberattacks occurred at the beginning of the global COVID-19 pandemic (early 2020) in the health sector. We selected well-documented cyberattacks with detailed information available, including root causes and consequences. The main findings are summarized in Table 2.

Table 2. Security incidents during the COVID-19 pandemic.

Security incidents	Type of attack	Impact
Brno University Hospital [3]	Ransomware	Postponement of surgeries, appointments, etc
US Department of Health and Human Services [4]	Distributed denial of service	Disruption to COVID-19 pandemic responses
World Health Organization [5]	Ransomware/phishing	Defacement and misinformation
Gilead Sciences, Inc [6]	Phishing	Impersonation and exfiltration
Hospitals in Romania [7]	Phishing/ransomware	Disruption and exfiltration
Health care supply chains [8]	Malware	Disruption of activities

Brno University Hospital in the Czech Republic, which is one of the country's main COVID-19 testing centers, was struck by ransomware, resulting in the postponement of surgeries. The ransomware infection was confirmed in the early hours of the day when the hospital decided to disconnect all computer networks. It was noticed that the ransomware infection was gradually replicating, and all the individual systems were failing. As a result, all computers had to be shut down. The hospital is reported to be still recovering capabilities, as it is not yet fully operational due to the attack [3]. The attack had an impact on the activities of the hospital as there was no database systems, that is, means of storing data; hence, staff have had to write and transfer their notes manually. This leads to slow processes and can potentially endanger lives in these trying times.

The US Department of Health and Human Services experienced a DDoS attack intended to disrupt the organization's responses to the COVID-19 pandemic. This attack targeted its servers by overloading it with millions of hits over several hours [4]. It was reported as a campaign of disruption aimed at hindering the response to the coronavirus pandemic as the targeted agency was tasked with protecting the health of citizens and delivering essential human services. Although the agency claimed the attack was not successful, and that the attackers did not infiltrate the internal network nor steal any data, this demonstrates that attacks like these can cause damage not just to the services of health agencies but also to the lives that depend on it, especially in times of emergencies.

Increased phishing website hacking attempts on the WHO and its partners led to the WHO putting out a warning to the general public to be more careful [5], as it has been reported that over 4000 coronavirus-related domains (ie, domains that contain words like "corona" or "covid") have been registered since the beginning of 2020. These registered domains were used by adversaries for phishing-related activities. Thus, the WHO incident was orchestrated by hackers in order to steal passwords. It was reported that a group of hackers created a malicious website posing as an email login portal for WHO employees in an attempt to steal their passwords. Although the WHO claims the attack was not successful, it still shows that phishing attacks can be leveraged to target health organizations.

Coronavirus vaccine manufacturer Gilead Sciences, Inc, was also targeted by hackers [6]. Staff at this pharmaceutical company were targeted via a fake email login page that was designed to steal passwords. It was reported that the attack was an attempt to compromise the email accounts of staff at the company using messages that impersonated journalists.

Hospitals in Romania experienced ransomware attacks by hackers as well [7]. The hackers were planning to use COVID-19–themed emails to infect these hospitals with ransomware. Their motivation was the protest against the COVID-19 quarantine measures of the country. The hackers owned malwares (eg, remote access trojans, ransomware, website defacements, and SQL injection tools) that can be used to bring down servers and steal information. It was reported that they intended to send emails about COVID-19 to hospitals to infect computers, encrypt files, and disrupt hospital activities. However, the attack was not as successful as the hackers were tracked down and arrested by Romanian law enforcement.

It has been reported that Interpol has cautioned agencies around the world about a significant rise in the global number of ransomware attacks explicitly targeting hospitals and health institutions [8]. It discovered that there was an increase in the number of attempted ransomware attacks on organizations in the 194 member countries. Additionally, a cyber warning was issued for key health care organizations involved in the coronavirus response both in the United Kingdom and the United States. A joint statement by the United Kingdom's National Cyber Security Centre (NCSC) and US Cybersecurity and Infrastructure Security Agency revealed that malicious cyber campaigns had been uncovered, with large-scale "password spraying" campaigns directed at health care bodies and medical research organizations in both nations [23].

Health care supply chains have not been omitted from these attacks; the US Federal Bureau of Investigation (FBI) issued a warning about a malware targeting this sector. The malware is called Kwampirs, a remote access Trojan that exploits network vulnerabilities of targeted organizations across the United States, Europe, Asia, and the Middle East [24]. The infected supply chain components included cyber-physical systems assets in health care organizations. The FBI alerted the health care sector against future cyberattacks, as Kwampirs have been historically targeting health care organizations.

The analysis of the above-mentioned incidents indicate that the health sector has become a primary target of cybersecurity attacks. Attackers are taking advantage of the COVID-19 pandemic and launching attacks, which are mainly ransomware, DDoS, phishing, and other type of malwares. The health care supply chain can be more vulnerable to cyberattacks especially during pandemics. The cyberattacks have resulted in negative impacts on the availability of essential health care services and challenged health care organizations in the protection of the confidentiality and integrity of health care information.

Health Care Cybersecurity Challenges

Selected papers discussing the main challenges of cybersecurity in the health sector were reviewed, and the main findings are summarized in Table 3.

Table 3. Key health sector security challenges and associated vulnerabilities.

Key challenges and published vulnerabilities	Reference
Remote work security assurance	
There are known security vulnerabilities with remote desktop protocols and virtual private networks	Argaw et al [10]
There are known attacks on health care system such as distributed denial-of-service attacks, malware, etc	Offner et al [14]
Cyberattacks target innumerable wireless connected devices in health care	Boddy et al [9]
Endpoint device management	
An endpoint device can provide an entry point to larger health care networks	Coventry et al [12]
The integration of new endpoint devices with outdated, legacy, or unsupported operating systems compromises interoperability and increases cybersecurity vulnerability	Kruse et al [13], Naidoo [25]
The health sector relies heavily on perimeter defense (antivirus, firewalls) for protection against cyber risk	Reagin and Gentry [26]
The factor that most influences cybersecurity in a hospital is endpoint complexity	Jalali and Kaiser [27]
Human factors in cybersecurity	
The majority of information security incidents are related to human error	Evans et al [28], Evans et al [29]
There is a statistically significant positive correlation between workload and the probability of health care staff opening a phishing email	Jalali et al [19]
The health sector lacks root cause analysis and cybersecurity incident prevention, especially those through unintentional human error	Evans et al [28], Evans et al [29]
Although some effort has been made to analyze human error (eg, use of IS-CHEC[a]), such approaches have not been widely adopted	Evans et al [30]
Lack of security awareness	
There is low awareness in the health sector of cyber risks	Gordon et al [31]
The most common action taken in response to breaches or attacks is additional staff training or communication	Furnell and Shah [32]
Health staff has poor awareness of consequences of behavior, and there is a lack of policies and reinforcement of secure behavior	Coventry et al [33]
There is a lack of pandemic-specific cybersecurity training campaigns, documented procedures, and guidance on revised procedures and technologies	Kaplan [34]
Inadequate board-level risk assessment communication	
There is a need for a matrix that can translate the strategic requirements of a health care system into prioritized cyber improvement needs	Barad [35]
There is a lack of understanding of security risks and its impact on organization-wide risk management	Tully et al [36]
There is a lack of appreciation among health care executive management of the business risk impact associated with cyber breaches	Jones and Katzis [37]
Inadequate business continuity plans	
Risks will continue to grow if cybersecurity is not designed into the product from the beginning of the product or project life cycle	Coventry and Branley [12]
The key security risks challenging business continuity are vendor dependence, inappropriate encryption configurations, and the inability to handle health information sharing and exchange with third-party and cross-border partners	Frontoni et al [38], Bhatia and Ibrahim [39], Natsiavas et al [40], Nalin et al [41]
The health sector lacks sophisticated data security tools compared to other industries	Walker-Roberts et al [42]
Cybersecurity capability is a strategic asset that every health organization must adopt, along with the concepts of building organizational resilience and the capacity to learn from mistakes	Jalali et al [11], Reagin and Gentry [26]
Lack of coordinated incident response	
The health sector tends to have a time lag between an attack occurring and detection of the breach	Coventry and Branley [12]
Current health care cyber defense is often reactive and undertaken after malicious attacks	Akinsanya et al [43]

Key challenges and published vulnerabilities	Reference
There is a lack of a coordinated incident response capacity to actively counteract constantly emerging and evolving malware threats	Chen et al [44]
Cybersecurity should be a team effort, from board members to front-line employees, with all being held accountable for cybersecurity	Pullin [45]
Limited budget and the need to deliver health care services without disruption	
There is a lack of experienced cybersecurity experts in the health care industry	Argaw et al [46]
There is a lack of a value-based system to weigh and balance benefits and risks in aspects of security, privacy, and adoption of technology	Boddy et al [9]
Vulnerable MCPS[b]	
Limited MCPS capability makes the health sector vulnerable to compromises	Almohri et al [47]
The reliance on the health care network increases cybersecurity risks to health care systems	Zheng et al [48]
Cyber threats can be introduced to the MCPS though vulnerable IoT[c] devices	Jimenez et al [49]

[a]IS-CHEC: Information Security Core Human Error Causes.

[b]MCPS: medical cyber-physical systems.

[c]IoT: internet of things.

The analysis shows that the main cybersecurity challenges of the health sector are remote work security assurance, endpoint device management, human errors, the lack of security awareness, inadequate senior-level security risk assessment, inadequate business continuity plans, the lack of coordinated incident response, constrains on budget and resources, and vulnerability of medical systems. These challenges cover not only the security-related matters as a result of the COVID-19 pandemic but also the inherent security challenges in the health sector that can be potentially exploited by attackers during the COVID-19 pandemic. It is imperative for the health care organizations to identify these challenges and take actions for prevention.

Remote Working Security Assurance

As remote working is now an integral element of health care service delivery, health staff are relying on enterprise remote desktop protocols and virtual private networks (VPN) to access internal networks. However, these come with certain risks that adversaries are looking to exploit. For example, the remote desktop protocol has a history of security issues and generally should not be publicly accessible without additional protections such as firewall, whitelist, and multifactor authentication [10]. Likewise, VPNs also have some known and unknown vulnerabilities, both on the client and server side, which have been exploited for years by cybercriminals [19]. The DDoS attacks on health care systems [14] and the innumerable wireless connected devices [9] have created further challenges to a remote work environment.

Endpoint Device Management

A number of endpoint devices, which comprises various patient-monitoring equipment that either connects to the internet or legacy-dispersed networks, are often unpatched [12]. This risk further increased during the pandemic as a result of organizations competing to procure internet of things (IoT) devices during the COVID-19 pandemic for their staff, which resulted in more employees than before using personal devices to perform work from home. From an enterprise architecture perspective, having tighter integration across the information technology (IT) environment is positive in terms of the organization being more agile; however, it makes the network vulnerable to cyberattacks such as email phishing, ransomware, DDoS, and network data breaches [13]. The integration of new endpoint devices with outdated legacy systems can increase vulnerabilities [13,25]. However, organizations overly rely on perimeter defense (antivirus, firewalls) and other forms of basic protection against cyberattacks [26]. By interviewing 19 C-Suite cybersecurity professionals, Jalali et al [27] also confirmed the factor that most influences cybersecurity in a hospital setting is endpoint complexity.

Human Factors in Cybersecurity

Existing research has shown that the majority of information security incidents are related to human error [28]. There is a tendency for human error when staff are busy focusing on saving lives and adjusting to new work environments and technologies. With sudden changes in working practices, being under stress for an extended period of time makes employees vulnerable to falling into malicious trickery and making mistakes [28]. According to Jalali at al [19], there is a statistically significant positive correlation between workload and the probability of a health care staff opening a phishing email. Naidoo et al [25] developed a multilevel influence model to explore how cybercriminals exploited the COVID-19 pandemic using social engineering techniques. However, the health sector lacks root cause analysis [28] to prevent human error related security incidents, especially those through unintentional human error [29]. Although some efforts have been made in applying the human reliability analysis technique in the context of information security (eg, Information Security Core Human Error Causes [IS-CHEC] [30]) to analyze human error, such approaches have not been widely adopted.

Lack of Security Awareness

Cybercriminals are exploiting people's anxieties during the COVID-19 pandemic. Gordon et al [31] identified that there is low awareness in the health sector of risks. Furnell et al [32] identified that the most common action taken in response to the most disruptive breaches or attacks is additional staff training or communication. Coventry et al [33] reported that health staff had poor awareness of the consequences of certain behaviors, and there is a lack of policies and reinforcement of secure behavior. However, increased cybersecurity awareness is required for the health sector to protect themselves and their patients from potential cyber threats such as phishing and ransomware. Due to the lack of prior planning and training to work under pandemic situations, health care staff require more training and support, such as pandemic-specific cybersecurity training campaigns, documented procedures, and guidance on revised procedures and technologies [34]. For example, health sector staff should be made aware of and able to flag phishing emails containing buzzwords during a pandemic, such as "WHO" or "donation." They should also be advised on how to validate trustworthy information sources in order to avoid ransomware attacks [1].

Inadequate Board-Level Risk Assessment Communication

There is a lack of understanding of security risks and its impact on organization-wide risk management, such as impacts on patient care and clinical outcomes [36]. The health sector lacks a matrix that can translate the strategic improvement needs of a health care system into prioritized information/cyber improvement needs [35]. Schwartz et al [37] identified that there is a lack of appreciation among health care executive management staff of the business risk impacts of cyber breaches.

Inadequate Business Continuity Plans

The health sector does not have enough data protection mechanisms; Walker-Roberts et al [42] confirmed that the health sector lacks sophisticated data security tools compared to other industries. Security is not built into its supply-chain and third-party vendors. Existing research shows that the key security risks challenging business continuity are vendor dependence, inappropriate encryption configurations, and the inability to handle health information sharing and exchange with third-party and cross-border partners [38-41]. Risks will continue to grow if cybersecurity is not integrated into the project life cycle from the beginning [12]. Cybersecurity capability is a strategic asset that every health organization must adopt, along with the concepts of building organizational resilience and the capacity to recover from incidents and learn from mistakes in order to maintain business continuity [11].

Lack of Coordinated Incident Response Involving Different Parties

As highlighted by Coventry and Branley [12], the health care sector has a exhibited a trend of having a time lag between the occurrence of an attack and its detection. In fact, this aids attackers by giving them more time to explore the network and conduct lateral movement, which increases the damage inflicted by security breaches. Current health care cyber defense response is often reactive and undertaken after malicious attacks [43], lacking a coordinated incident response capacity to counteract constantly emerging and evolving malware threats [44]. The failure of health care organizations in having a successful and secure backup mechanism in place makes it frail in terms of incident response and recovery [12]. Pullin et al [45] also confirmed that cybersecurity should be a team effort, with everyone from board members to front-line employees being held accountable for cybersecurity.

Limited Budget and the Need to Deliver Health Care Services Without Disruption

Although health care services are spending funds to become more integrated to deliver health care services without disruption [9], the necessary emphasis is not given to the security aspect in terms of upkeep (eg, keeping software updated and systems secure). However, this is reported to be due to a shortage in experienced cybersecurity experts within health care organizations with the required skills and experience to enable health care organizations to change their business operations at significant pace without undertaking the "usual" levels of cybersecurity assurance [46]. Boddy et al [9] identified the needs of a value-based system to weigh and balance the benefits and risks in aspects of security, privacy, and adoption of technology.

Vulnerable Medical Cyber-Physical Systems

Cybersecurity measures such as vulnerability scans or patch management are often not available or only possible by manufacturers [49]. Their basic limited capability makes them vulnerable to compromise [47]. Cybersecurity measures such as vulnerability scans or patch management are often not available or only accessible for manufacturers. Moreover, their connection and reliance upon the health care network significantly increase the cybersecurity risk to the entire health care system [48]. With the widespread use of IoT medical devices, cyber threats can be introduced to medical cyber-physical systems though vulnerable IoT devices [44].

Health Care Cybersecurity Controls

Selected papers discussing cybersecurity solutions present within the health sector were reviewed, and the main findings are summarized in Table 4.

Table 4. Crucial health sector security solutions.

Solution	Reference
Apply endpoint device management tools	
Apply perimeter-based defense (antivirus, firewalls) for protection against cyberattacks	Reagin and Gentry [26]
Restrict the technologies and devices used by health staff to remain compliant with security regulations such as HIPAA[a] during pandemics	Hoffman [20]
Adapt the NIST[b] approach to manage security IoT[c] medical devices	Kelly et al [50]
Secure the remote work environment	
Apply multifactor authentication	Argaw et al [10]
Apply a chaotic map–based authenticated security framework for remote point of care	Deebak et al [51]
Apply remote access monitoring such as the NHS[d] attack surface reduction rules	Zorz [52]
Apply perimeter security solution such as NHS Secure Boundary to enable secure access	NHS Digital [53]
The health care sector needs to ensure data protection mechanisms for securing system access and transmitting data	Rezaeibagha et al [54]
Raise security awareness	
Apply a holistic, integrated approach to improve staff awareness, competence, and mitigation of threats	Pullin [45], Sedlack [55]
Implement cybersecurity training programs and cybersecurity awareness campaigns	Gordon et al [56]
Apply the NCSC's[e] Board Toolkit to raise board-level security awareness	NHS Digital [57]
Provide comprehensive employee training and education to enable the identification and assessment of risks	Alzahrani [58]
Implement a positive organizational climate to influence people's behavior	Kessler et al [59]
Ensure business continuity	
Apply a self-assessment tool such as the NHS Data Security and Protection Toolkit	NHS Digital [60]
Embrace cybersecurity and a develop strong culture of cyber vigilance	Dameff et al [61]
Ensure business continuity through data backups, intrusion detection, and prevention systems	Rezaeibagha et al [54]
Apply a systematic risk assessment of the impacts on health care business operations	Kim et al [22]
Consider cybersecurity insurance in health care	Kabir et al [62]
Apply technical controls	
Apply network segmentation to isolate network traffic	Hakak et al [1]
Apply general technical controls including encryption, authentication, and authorization	Yaseen et al [63]
Apply homomorphic encryption that ensures strong security and privacy guarantees while enabling analysis of encrypted data and sensitive medical information	Raisaro et al [64]
Apply blockchain to facilitate health care interoperability	Narikimilli et al [65]
Apply cryptographic security to address data sharing and storage of patient information across network systems	Pussewalage and Oleshchuk [66]
Policies and legislations	
Laws and regulations can help to combat the issues of medical cyber-physical systems	Raisaro et al [64]
Security instructions and control designs should be tailored	Wang and Jones [67]
Regulatory changes or manufacturers should become more security-minded in the medical device design phase	Department of Health and Social Care, UK Government [68]
Policymakers may need to alter policies to allow new technological innovations to be applied to health care	Bhuyan et al [69]
The US Congress passed the 21st Century Cures Act to promote patient control over their own health information while protecting privacy and cybersecurity	Hoffman [20]
Incident reporting and cyber threat intelligence support	
NHS Digital issued two high-severity CareCERT alerts (BlueKeep and DejaBlue) and developed a high-severity alert process handbook to facilitate incident reporting and sharing	Department of Health and Social Care, UK Government [68]

Solution	Reference
Apply an evidence-based approach, such as the generic security template, for incident reporting and exchange	He and Johnson [70], He and Johnson [71]
Establish an international workforce to facilitate cyber threat reporting and exchange to combat pandemic-themed cyber threats	Hakak et al [1]
Cybersecurity guidance specific to COVID-19	
The NHS has added guidance on working from home securely in the context of COVID-19	NHS Digital [72]
The United Kingdom's Information Commissioner's Office created an information hub to assist individuals and organizations to manage data protection during the COVID-19 pandemic	Information Commissioner's Office [73]

[a]HIPAA: Health Insurance Portability and Accountability Act.

[b]NIST: National Institute of Standards and Technology.

[c]IoT: internet of things.

[d]NHS: National Health Service.

[e]NCSC: National Cyber Security Centre.

Apply Endpoint Device Protection

During the COVID-19 pandemic, health staff working from home may adopt telehealth technologies or IoT devices. This increases cybersecurity risks, as it expands the footprint for cyberattack to the use of new devices outside of the service providers' network [50]. Health staff are advised to restrict the technologies and devices they used to remain compliant with security regulations such as Health Insurance Portability and Accountability Act during the pandemics [20]. However, health care organizations mainly reply on perimeter defense (eg, antivirus, firewalls) for protection against the potential cyberattacks [26]. The National Institute of Standards and Technology (NIST) has recently released a draft security guide and recommendations for managing the security IoT devices, but it is unclear whether it will be enforced across the health sector [50].

Secure Remote Work Environment

Existing solutions include the use of multifactor authentication and the monitoring of the log activity of user accounts and revoking account access if no longer needed [10]. Deebak et al [51] proposed a chaotic map–based authenticated security framework for remote point of care. Health organizations such as those in the United Kingdom have started using services to monitor their remote access infrastructure constantly and to investigate anomalies. For example, the National Health Service (NHS) has employed attack surface reduction rules (eg, block macros, executable content, process creation) [52]. Furthermore, a more recent NHS Digital service, Secure Boundary, was introduced as a perimeter security solution to enable secure access for NHS staff and to provide security monitoring [53].

Raise Security Awareness

Health care organizations already have cybersecurity programs in place to increase levels of security awareness [45,55]. Existing solutions include the use of cybersecurity training programs and cybersecurity awareness campaigns [56]. In a cybersecurity campaign, the IT department sends out fake phishing emails to their staff and provides further training to those who fail to identify these emails [56]. In the United Kingdom, more than 100 NHS boards have completed cybersecurity training accredited by the Government Communications Headquarters since the WannaCry attack. Furthermore, the NCSC's Board Toolkit for the NHS provides additional information on ransomware and backups. NHS Digital also runs a cyber awareness campaign called the Keep I.T. Confidential campaign. Over 340 organizations have downloaded the materials since its launch in September 2019 [57]. However, there is not enough work on training programs tailored to the pandemic such as COVID-19–themed social engineering, although the world is realizing the importance of raising the awareness of COVID-19–related cyberattacks [58]. Existing research shows that positive organizational climate can influence people's behavior [59].

Ensure Business Continuity

Health care leadership must embrace cybersecurity and develop strong cultures of cybervigilance [61]. The health sector already has business continuity solutions in place such as data backups and intrusion detection and prevention systems [54]. NHS trusts have been asked to follow and meet the Cyber Essentials and government standards. NHS Digital has launched a Data Security and Protection Toolkit [60], a self-assessment tool for organizations that need to access NHS patient information and systems. The toolkit must be applied to ensure that organizations practice good cyber hygiene. Security risk assessment is essential to ensure business continuity. Kim et al [22] systematically assessed the impacts of cybersecurity threats on remote health care. Cybersecurity insurance in health care [62] should also be considered as a solution to ensure business continuity management, but it has not been widely adopted.

Apply Technical Controls

General technical controls applied by the health sector include encryption, authentication, and authorization to protect data from cyber threats [63]. Cryptographic security is used to address data sharing and storage of patient information across network systems [66]. Homomorphic encryption is applied to ensure robust security and privacy guarantees while enabling analysis of encrypted data and sensitive medical information [64]. Blockchain is also applied to facilitate health care interoperability due to its immutability, transparency, and decentralization [65]. Network segmentation and isolation also

need to be considered by the health sector [1]. With network segmentation, network traffic can be isolated and/or filtered to limit and/or prevent access between network zones. For example, in case of systems compromise, one should freeze any activity in the system, disconnect the infected machines from any external drive or medical device, and go offline from the network.

Policy and Legislation

The health sector already has security policies and legislation in place for cybersecurity management. Laws and regulations are available to protect medical cyber-physical systems [64]. Security controls need to be tailored according to regulation [67]. Manufacturers are also required to consider these regulations to design medical devices [68]. However, policymakers may need to alter policies to allow new technological innovations to be applied to health care [69]. The US Congress passed the 21st Century Cures Act to promote the interoperability of electronic health records and promote more patient control over one's own health information while protecting privacy and cybersecurity [20]. However, more efforts are needed on security policies or legislations in handling cybersecurity-related matters during pandemics like COVID-19.

Incident Reporting and Cyber Threat Intelligence Support

The health sector is required to report cybersecurity incidents to a supervisory authority, such as the national Computer Security Incident Response Team in the European Union. In the United Kingdom, there is government-approved support from the NCSC. NHS Digital has issued two high-severity CareCERT alerts in 2019 (BlueKeep and DejaBlue). After developing a high-severity alert process handbook, remediation went from 18 weeks for BlueKeep down to 3 weeks for DejaBlue [68]. He and Johnson [70,71] proposed a generic security template, which is an evidence-based argumentation approach to facilitate incident reporting and exchange. This approach was applied to a health care organization but has not been widely adopted. Hakak et al [1] identified the needs of establishing an international workforce to facilitate threat reporting and cyber threat intelligence (eg, attack vectors and countermeasures) exchange to combat pandemic-themed cyber threats. The health sector will benefit from such practices during pandemics in order to avoid similar incidents.

Cybersecurity Guidance Specific to COVID-19

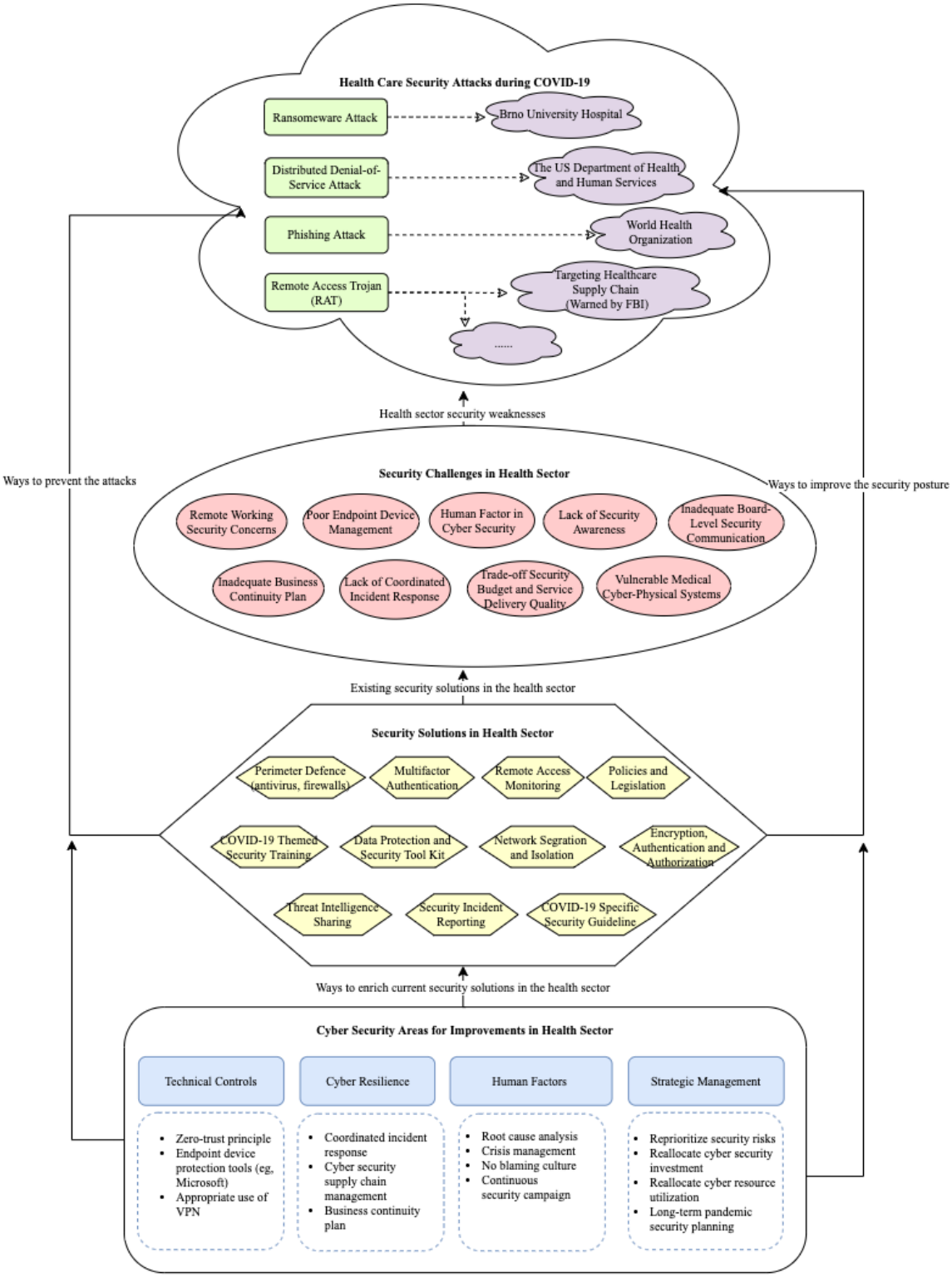
Some health care organizations have started providing security guidance specific to COVID-19 for their staff. For example, NHS Digital has added guidance on working from home security, ramping up its on-site support for trusts on risk mitigations, data backup, and threat response. They also offer the NHS the NCSC's Protective Domain Name Service free of charge [72]. Furthermore, governments also provide cybersecurity guidance to both individuals and organizations. For example, the United Kingdom's Information Commissioner's Office created an information hub in order to assist individuals and organizations to protect data during the COVID-19 pandemic [73].

Discussion

Summary of Evidence

Through a scoping review, this research identified key cybersecurity challenges, solutions adapted by the health sector, and areas to be improved in order to counteract the cyberattacks introduced through changes to working practices in the face of the COVID-19 pandemic. This review identified 9 main challenges in cybersecurity and 11 key solutions that health care organizations adapted to address these challenges. Based on our findings and analysis, we can conclude that the main challenges that the health sector faces due to the COVID-19 pandemic include increased reliance on remote working by staff, high demand for PPE by staff on the first line of defense, and decreased mobility due to the lockdown. Indeed, these changes have made the health sector vulnerable to potential cyberattacks. For example, remote work was taken up by users with little previous experience, and there was also no planning and cybersecurity-associated assurance prior to the shift. Furthermore, evidence can be seen from the security incidents that took place during the lockdown period such as those of Brno University Hospital, hospitals in Romania, etc. The health sector continues to face security challenges [1,17]. Challenges such as remote working security assurance, endpoint device management, inadequate business continuity plans, lack of security awareness, etc, are apparent in the health sector. There are some existing solutions employed by health care organizations, especially in the United Kingdom, such as remote access monitoring. Figure 2 summarizes the main findings from the literature review and highlights the gaps and vulnerabilities that were exploited during the cyberattacks that took place during the COVID-19 pandemic. However, there are still challenges and gaps to be addressed, as discussed below.

Figure 2. Security attacks, key security challenges, solutions, and areas to improve. FBI: Federal Bureau of Investigation; VPN: virtual private network.



Implications for Future Research

Although the health sector has made some efforts to address these challenges, more research is required in some domains.

Technical Controls

The health sector has applied some technical solutions to tackle cybersecurity challenges in order to secure the remote work environment and monitor endpoint applications. These include but are not limited to network security (eg, network segmentation), multifactor authentication, password protection, patching systems, and the use of intrusion detection and prevention systems. There are also innovative security solutions such as the zero-trust principle (ie, to treat all devices as untrustworthy before access or authorization can be considered). The use of VPNs is a popular technique in the remote work environment but is not always required. Health care organizations should avoid the abuse of VPNs and ensure it is

applied to specific tasks, such as for system admin use and medical diagnosis purposes through access to legacy systems (eg, patient records management systems) stored on private data servers. Future research should explore innovative solutions such as blockchain as it can facilitate health care interoperability due to its immutability, transparency, and decentralization. In general, the health sector significantly lags behind other sectors in terms of cybersecurity. Future research should borrow experience from general cybersecurity practices (eg, NIST guidelines) and adapt them according to the needs of the health sector, especially in the context of pandemics.

Cyber Resilience

In order to improve system resilience, health organizations have some business continuity planning in place for data protection and recovery but lack a systematic way to maintain cyber resilience [18]. The vulnerabilities in the cyber supply chain makes it difficult to recover from an incident caused by third parties [38-41]. In the case of impact on medical devices or clinical information systems, incident response should be coordinated with device manufacturers and vendors. Health care organizations have realized the importance of having a comprehensive view of cybersecurity management in order to prevent cyberattacks [18] but have not built this coordinated capacity. There is a lack of a cyber resilience program to evaluate vendors' capabilities around threat protection, particularly across email servers (phishing and ransomware), breadth of portfolio coverage in addressing cloud architecture, and endpoint security. Future research should focus on building a coordinated cybersecurity capacity in order to systematically assess vulnerabilities and respond to cyber threats.

Human Factors in Cybersecurity

People are likely to make mistakes, especially in the context of changes in their traditional way of working. Health care organizations are required to adopt a nonblaming culture in reporting incidents. The health sector should focus on root cause analysis [28] and prevent incidents from happening especially through unintentional human error. Published research has shown that the majority of information security incidents relate to human error [28,29], which is a vulnerability that attackers will look to exploit. A human error analytical approach such as IS-CHEC could be deployed both reactively, through integration within incident management practices [29,30], and proactively, through simple interaction with operational personnel [29], to detect current human error areas of weaknesses and apply associated remedial and preventative measures. Moreover, health care staff in the organization need to be educated and build awareness of the ongoing security situation during the COVID-19 pandemic. For example, in the case of infection, staff are required to disconnect from the network to contain the spread. Organizations should continuously raise awareness internally by launching campaigns even during a time of crisis (ie, to inform health staff not to open suspicious emails). Future research should focus on creating pandemic-themed security awareness campaigns. Moreover, a positive and empowering culture is also required (eg, by sharing the rate of people who did not click on phishing-negative emails during a training campaign). Experience can be borrowed from the organizational climate literature to positively influence people's behavior [59].

Strategic Cybersecurity Management

Although health care organizations have invested in cybersecurity to counteraction security attacks, further efforts are needed to reprioritize cybersecurity risk assessment during the COVID-19 pandemic, reallocate security investment, and optimize resource utilization to obtain adequate assurances. According to Argwa et al [46], health care organizations are advised to allocate more resources and funding to cybersecurity. Strategic cybersecurity investment is still an immature research area in health care largely due to boards' inability to fully understand and anticipate the direct and indirect impact on their health services. Further, there are language barriers between the technical team and the board [27]. Another reason is that the board finds it difficult to estimate the costs of investing and balancing these against potential benefits procured or impacts mitigated [8] as cybersecurity investments prevent potential losses but may not generate business benefits directly. Moreover, organizations should not only create security guidelines specific to the COVID-19 pandemic but also plan for the long term for remote working and spend efforts on strengthening their security mechanisms and cybersecurity crisis management capabilities. More research efforts are needed to support the top management teams of the health sector to understand the threat landscape and make better-informed decisions to allocate resources not just to provide services to staff and patients but also for protection and resilience, in order to continuously serve even in times of emergency such as the current pandemic and beyond.

Limitations

Contrary to systematic reviews, scoping reviews are used to identify knowledge gaps, scope a body of literature, and clarify concepts. However, some limitations should be considered. Scoping reviews usually provide descriptive information in order to address the objectives of the review, which often leads to less defined searches. This review mitigated this limitation by clearly defining the search terms and search formula. Scoping reviews are also at risk of bias from different sources. All 4 authors were involved in the article identification, selection, and analysis processes in order to reduce the risks of bias. Because of variability when conducting a scoping review, there is a need for methodological standardization to ensure the strength of evidence. This review followed the PRISMA-ScR to standardize the process and improve the strength of evidence. Another limitation is that this review included exact terms used to search the titles or abstracts of existing publications. Any articles that used different terms, (eg, "computer security") would not have been included. In addition, publications that were not written in English were excluded. Moreover, although this scoping review focused on health care, the solutions identified could be applied to other industries.

Conclusions

The COVID-19 pandemic has challenged the resilience of the health care information system. This research was motivated by the urgency of counteracting the cyberattacks that have recently happened to hospitals, pharmaceutical companies, the

US Department of Health and Human Services, and the WHO and its partners, etc. We performed a review on security challenges of the health sector and the solutions employed during COVID-19. We identified the root causes of the security incidents that have impacted the health sector during the COVID-19 pandemic, cybersecurity challenges, solutions, and areas in need of improvement. The results show that the main root causes of the security incidents that happened during the COVID-19 pandemic are mainly from phishing, ransomware, DDoS attacks, and malware. The main challenges faced by health care organizations are inadequate endpoint device management, lack of security awareness, insecure remote work environment, inadequate business continuity plans, lack of coordinated incident response, and difficulty in trading off security investment and service delivery quality. Needless to say, another major challenge is human error, both from the perspective of the health care worker at the frontline and those working from home. As the COVID-19 pandemic has shifted our priorities, there is a greater tendency for human error to occur when staff are preoccupied with saving lives, working in a strange or different environment, and using new or various technologies. With little or no experience and a lack of prior planning and training to work in such situations, health care workers require more than training and support, such as adequate time, documented procedures, and guidance on revised procedures and technology.

Although the health sector has made some efforts to address these challenges by applying technical measures, raising security awareness, enforcing policies, and developing COVID-19–specific guidelines, more research efforts are still required in some domains. Future research should focus on exploring enhanced technical controls through the adaption of general cybersecurity practices (eg, NIST guidelines); improving cyber resilience by building a coordinated cybersecurity capacity to systematically assess vulnerabilities of the complex health care supply chain and respond to cyber threats; reducing human-related security incidents by exploring human error reduction approaches and pandemic-themed awareness campaigns; and enhancing strategic cybersecurity management by exploring crisis management planning, security risks reprioritization, and the optimization of cybersecurity budget and resource reallocation.

Many health care organizations are applying a temporary solution to counteract cyber threats during the COVID-19 pandemic. These organizations should plan for the long term, provide adequate levels of cybersecurity resources to deal with fast-changing situations, and offer the required assurance within these changes. This paper provides useful insights for the health sector on their cybersecurity issues during the COVID-19 pandemic or other epidemic or pandemic situations in the future. Moreover, cybersecurity experience in other sectors can be borrowed and applied in the health sector.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (grant 61803318).

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA-ScR checklist.
[DOCX File , 25 KB - jmir_v23i4e21747_app1.docx]

References

1. Hakak S, Khan WZ, Imran M, Choo KR, Shoaib M. Have You Been a Victim of COVID-19-Related Cyber Incidents? Survey, Taxonomy, and Mitigation Strategies. IEEE Access 2020;8:124134-124144. [doi: 10.1109/ACCESS.2020.3006172]
2. Williams CM, Chaturvedi R, Chakravarthy K. Cybersecurity Risks in a Pandemic. J Med Internet Res 2020 Sep 17;22(9):e23692 [FREE Full text] [doi: 10.2196/23692] [Medline: 32897869]
3. Porter S. Cyberattack on Czech hospital forces tech shutdown during coronavirus outbreak. HealthCareITNews. 2020 Mar 19. URL: https://www.healthcareitnews.com/news/europe/cyberattack-czech-hospital-forces-tech-shutdown-during-coronavirus-outbreak [accessed 2021-04-05]
4. Stein S, Jacobs J. Cyber-attack hits U.S. health agency amid Covid-19 Outbreak. Bloomberg. 2020 Mar 16. URL: https://www.bloomberg.com/news/articles/2020-03-16/u-s-health-agency-suffers-cyber-attack-during-covid-19-response [accessed 2021-04-05]
5. Beware of criminals pretending to be WHO. World Health Organization. URL: https://www.who.int/about/communications/cyber-security [accessed 2021-04-05]
6. Stubbs J, Bing C. Exclusive: Iran-linked hackers recently targeted coronavirus drugmaker Gilead - sources. Reuters. 2020 May 8. URL: https://www.reuters.com/article/us-healthcare-coronavirus-gilead-iran-ex-idUSKBN22K2EV [accessed 2021-04-05]
7. Cimpanu C. Hackers preparing to launch ransomware attacks against hospitals arrested in Romania. ZDNet. 2020 May 15. URL: https://www.zdnet.com/article/hackers-preparing-to-launch-ransomware-attacks-against-hospitals-arrested-in-romania/ [accessed 2021-04-05]

8. Cybercriminals targeting critical healthcare institutions with ransomware. Interpol. 2020 Apr 4. URL: https://www.interpol.int/en/News-and-Events/News/2020/Cybercriminals-targeting-critical-healthcare-institutions-with-ransomware/ [accessed 2021-04-05]
9. Boddy A, Hurst W, Mackay M, El Rhalibi A. A study into data analysis and visualisation to increase the cyber-resilience of health care infrastructures. New York, NY: Association for Computing Machinery; 2017 Presented at: IML '17: Proceedings of the 1st International Conference on Internet of Things and Machine Learning; October; Liverpool, United Kingdom p. 1-7. [doi: 10.1145/3109761.3109793]
10. Argaw ST, Troncoso-Pastoriza JR, Lacey D, Florin M, Calcavecchia F, Anderson D, et al. Cybersecurity of Hospitals: discussing the challenges and working towards mitigating the risks. BMC Med Inform Decis Mak 2020 Jul 03;20(1):146 [FREE Full text] [doi: 10.1186/s12911-020-01161-7] [Medline: 32620167]
11. Jalali M, Russell B, Razak S, Gordon WJ. EARS to cyber incidents in health care. J Am Med Inform Assoc 2019 Jan 01;26(1):81-90 [FREE Full text] [doi: 10.1093/jamia/ocy148] [Medline: 30517701]
12. Coventry L, Branley D. Cybersecurity in healthcare: A narrative review of trends, threats and ways forward. Maturitas 2018 Jul;113:48-52. [doi: 10.1016/j.maturitas.2018.04.008] [Medline: 29903648]
13. Kruse CS, Frederick B, Jacobson T, Monticone DK. Cybersecurity in healthcare: A systematic review of modern threats and trends. THC 2017 Feb 21;25(1):1-10. [doi: 10.3233/thc-161263]
14. Offner KL, Sitnikova E, Joiner K, MacIntyre CR. Towards understanding cybersecurity capability in Australian healthcare organisations: a systematic review of recent trends, threats and mitigation. Intelligence and National Security 2020 Apr 22;35(4):556-585. [doi: 10.1080/02684527.2020.1752459]
15. Sardi A, Rizzi A, Sorano E, Guerrieri A. Cyber Risk in Health Facilities: A Systematic Literature Review. Sustainability 2020 Aug 27;12(17):7002. [doi: 10.3390/su12177002]
16. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Ann Intern Med 2009 Aug 18;151(4):264-9, W64. [doi: 10.7326/0003-4819-151-4-200908180-00135] [Medline: 19622511]
17. Schneck PA. Cybersecurity During COVID-19. IEEE Secur Privacy 2020 Nov;18(6):4-5. [doi: 10.1109/msec.2020.3019678]
18. Weil T, Murugesan S. IT Risk and Resilience—Cybersecurity Response to COVID-19. IT Prof 2020 May;22(3):4-10. [doi: 10.1109/MITP.2020.2988330]
19. Jalali MS, Bruckes M, Westmattelmann D, Schewe G. Why Employees (Still) Click on Phishing Links: Investigation in Hospitals. J Med Internet Res 2020 Jan 23;22(1):e16775 [FREE Full text] [doi: 10.2196/16775] [Medline: 32012071]
20. Hoffman D. Increasing access to care: telehealth during COVID-19. J Law Biosci 2020;7(1):lsaa043 [FREE Full text] [doi: 10.1093/jlb/lsaa043] [Medline: 32843985]
21. Ronquillo J, Erik Winterholler J, Cwikla K, Szymanski R, Levy C. Health IT, hacking, and cybersecurity: national trends in data breaches of protected health information. JAMIA Open 2018 Jul;1(1):15-19 [FREE Full text] [doi: 10.1093/jamiaopen/ooy019] [Medline: 31984315]
22. Kim D, Choi J, Han K. Risk management-based security evaluation model for telemedicine systems. BMC Med Inform Decis Mak 2020 Jun 10;20(1):106 [FREE Full text] [doi: 10.1186/s12911-020-01145-7] [Medline: 32522216]
23. Cyber warning issued for key healthcare organisations in UK and USA. National Cyber Security Centre. 2020 May 5. URL: https://www.ncsc.gov.uk/news/warning-issued-uk-usa-healthcare-organisations [accessed 2021-04-05]
24. Cimpanu C. FBI re-sends alert about supply chain attacks for the third time in three months. ZDNet. 2020 Mar 31. URL: https://www.zdnet.com/article/fbi-re-sends-alert-about-supply-chain-attacks-for-the-third-time-in-three-months/ [accessed 2021-04-05]
25. Naidoo R. A multi-level influence model of COVID-19 themed cybercrime. European Journal of Information Systems 2020 May 31;29(3):306-321. [doi: 10.1080/0960085x.2020.1771222]
26. Reagin MJ, Gentry MV. Enterprise Cybersecurity. Frontiers of Health Services Management 2018;35(1):13-22. [doi: 10.1097/hap.0000000000000037]
27. Jalali MS, Kaiser JP. Cybersecurity in Hospitals: A Systematic, Organizational Perspective. J Med Internet Res 2018 May 28;20(5):e10059 [FREE Full text] [doi: 10.2196/10059] [Medline: 29807882]
28. Evans M, He Y, Maglaras L, Janicke H. HEART-IS: A novel technique for evaluating human error-related information security incidents. Computers & Security 2019 Jan;80:74-89. [doi: 10.1016/j.cose.2018.09.002]
29. Evans M, He Y, Maglaras L, Yevseyeva I, Janicke H. Evaluating information security core human error causes (IS-CHEC) technique in public sector and comparison with the private sector. Int J Med Inform 2019 Jul;127:109-119. [doi: 10.1016/j.ijmedinf.2019.04.019] [Medline: 31128822]
30. Evans M, He Y, Luo C, Yevseyeva I, Janicke H, Zamani E, et al. Real-Time Information Security Incident Management: A Case Study Using the IS-CHEC Technique. IEEE Access 2019;7:142147-142175. [doi: 10.1109/access.2019.2944615]
31. Gordon WJ, Fairhall A, Landman A. Threats to Information Security — Public Health Implications. N Engl J Med 2017 Aug 24;377(8):707-709. [doi: 10.1056/nejmp1707212]
32. Furnell S, Shah JN. Home working and cyber security – an outbreak of unpreparedness? Computer Fraud & Security 2020 Aug;2020(8):6-12. [doi: 10.1016/s1361-3723(20)30084-1]

33. Coventry L, Branley-Bell D, Sillence E, Magalini S, Mari P, Magkanaraki A, et al. Cyber-risk in health carexploring facilitators and barriers to secure behaviour, in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). Switzerland: Springer; 2020 Presented at: 22nd International Conference on Human Computer Interaction; July 19–24; Copenhagen, Denmark p. 105-122 URL: https://researchportal.northumbria.ac.uk/files/31585347/1689_final.pdf [doi: 10.1007/978-3-030-50309-3_8]
34. Kaplan B. evisiting health information technology ethical, legal, and social issues and evaluation: telehealth/telemedicine and COVID-19. Int J Med Inform 2020 Nov;143:104239 [FREE Full text] [doi: 10.1016/j.ijmedinf.2020.104239] [Medline: 33152653]
35. Barad M. Linking Cyber Security Improvement Actions in Healthcare Systems to Their Strategic Improvement Needs. Procedia Manufacturing 2019;39:279-286. [doi: 10.1016/j.promfg.2020.01.335]
36. Tully J, Selzer J, Phillips JP, O'Connor P, Dameff C. Healthcare Challenges in the Era of Cybersecurity. Health Secur 2020 Jun 01;18(3):228-231. [doi: 10.1089/hs.2019.0123] [Medline: 32559153]
37. Jones RW, Katzis K. Cybersecurity and the Medical Device Product Development Lifecycle. Studies in Health Technology and Informatics 2017;238:76-79 [FREE Full text] [doi: 10.3233/978-1-61499-781-8-76]
38. Frontoni E, Mancini A, Baldi M, Paolanti M, Moccia S, Zingaretti P, et al. Sharing health data among general practitioners: The Nu.Sa. project. Int J Med Inform 2019 Sep;129:267-274. [doi: 10.1016/j.ijmedinf.2019.05.016] [Medline: 31445266]
39. Bhatia S, Ibrahim A. Understanding Security Risks When Exchanging Medical Records Using IHE. In: 17th International Conference on Information Technology–New Generations (ITNG 2020) / Advances in Intelligent Systems and Computing. Cham: Springer; 2020:477-481.
40. Natsiavas P, Rasmussen J, Voss-Knude M, Votis, Coppolino L, Campegiani P, et al. Comprehensive user requirements engineering methodology for secure and interoperable health data exchange. BMC Med Inform Decis Mak 2018 Oct 16;18(1):85 [FREE Full text] [doi: 10.1186/s12911-018-0664-0] [Medline: 30326890]
41. Nalin M, Baroni I, Faiella G, Romano M, Matrisciano F, Gelenbe E, et al. The European cross-border health data exchange roadmap: Case study in the Italian setting. J Biomed Inform 2019 Jun;94:103183 [FREE Full text] [doi: 10.1016/j.jbi.2019.103183] [Medline: 31009760]
42. Walker-Roberts S, Hammoudeh M, Dehghantanha A. A Systematic Review of the Availability and Efficacy of Countermeasures to Internal Threats in Healthcare Critical Infrastructure. IEEE Access 2018;6:25167-25177. [doi: 10.1109/access.2018.2817560]
43. Akinsanya OO, Papadaki M, Sun L. Current cybersecurity maturity models: How effective in health care cloud? 2019 Presented at: CERC 2019; Sept 15-17; Niagara Falls, ON p. 211-222 URL: http://ceur-ws.org/Vol-2348/paper16.pdf
44. Chen Y, Ding S, Xu Z, Zheng H, Yang S. Blockchain-Based Medical Records Secure Storage and Medical Service Framework. J Med Syst 2018 Nov 22;43(1):5. [doi: 10.1007/s10916-018-1121-4] [Medline: 30467604]
45. Pullin DW. Cybersecurity. Frontiers of Health Services Management 2018;35(1):3-12. [doi: 10.1097/hap.0000000000000038]
46. Argaw ST, Bempong N, Eshaya-Chauvin B, Flahault A. The state of research on cyberattacks against hospitals and available best practice recommendations: a scoping review. BMC Med Inform Decis Mak 2019 Jan 11;19(1):10 [FREE Full text] [doi: 10.1186/s12911-018-0724-5] [Medline: 30634962]
47. Almohri H, Cheng L, Yao D, Alemzadeh H. On threat modeling and mitigation of medical cyber-physical systems. 2017 Presented at: 2017 IEEE/ACM International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE); July 17-19; Philadelphia, PA p. 114-119 URL: https://ieeexplore.ieee.org/document/8010624 [doi: 10.1109/CHASE.2017.69]
48. Zheng G, Zhang A, Yang W, Valli C, Shankaran R, Orgun MA. From WannaCry to WannaDie: Security trade-offs and design for implantable medical devices. 2017 Presented at: 2017 17th International Symposium on Communications and Information Technologies (ISCIT); September 25-27; Cairns, QLD, Australia p. 1-5. [doi: 10.1109/iscit.2017.8261228]
49. Jimenez J, Jahankhani H, Kendzierskyj K. Health Care in the Cyberspace: Medical Cyber-Physical System and Digital Twin Challenges. In: Farsi M, Daneshkhah A, Hosseinian-Far A, Jahankhani H, editors. Digital Twin Technologies and Smart Cities. Internet of Things (Technology, Communications and Computing). Cham: Springer; 2020:79-92.
50. Kelly JT, Campbell KL, Gong E, Scuffham P. The Internet of Things: Impact and Implications for Health Care Delivery. J Med Internet Res 2020 Nov 10;22(11):e20135 [FREE Full text] [doi: 10.2196/20135] [Medline: 33170132]
51. Deebak BD, Al-Turjman F, Nayyar A. Chaotic-map based authenticated security framework with privacy preservation for remote point-of-care. Multimed Tools Appl 2020 Nov 12:1-26 [FREE Full text] [doi: 10.1007/s11042-020-10134-x] [Medline: 33204211]
52. Zorz Z. Vulnerable VPN appliances at health care organisations open doors for ransomware gangs. HelpNetSecurity. 2020 Apr 2. URL: https://www.helpnetsecurity.com/2020/04/02/vpn-healthcare-ransomware/ [accessed 2021-04-05]
53. NHS Secure Boundary: About our secure boundary service. NHS Digital. URL: https://digital.nhs.uk/cyber-and-data-security/managing-security/nhs-secure-boundary [accessed 2021-04-05]
54. Rezaeibagha F, Win KT, Susilo W. A systematic literature review on security and privacy of electronic health record systems: technical perspectives. Health Inf Manag 2015 Oct;44(3):23-38. [doi: 10.1177/183335831504400304] [Medline: 26464299]

55. Sedlack D. Understanding Cyber Security Perceptions Related to Information Risk in a Healthcare Setting. AIS eLibrary. 2016 Aug 11. URL: https://aisel.aisnet.org/amcis2016/ISSec/Presentations/42/ [accessed 2021-04-05]
56. Gordon W, Wright A, Glynn RJ, Kadakia J, Mazzone C, Leinbach E, et al. Evaluation of a mandatory phishing training program for high-risk employees at a US healthcare system. J Am Med Inform Assoc 2019 Jun 01;26(6):547-552 [FREE Full text] [doi: 10.1093/jamia/ocz005] [Medline: 30861069]
57. Cyber Associates Network: About the Cyber Associates Network. NHS Digital. URL: https://digital.nhs.uk/cyber-and-data-security/about-us/cyber-associates-network [accessed 2021-04-05]
58. Alzahrani A. Coronavirus Social Engineering Attacks: Issues and Recommendations. IJACSA 2020;11(5):9. [doi: 10.14569/ijacsa.2020.0110523]
59. Kessler SR, Pindek S, Kleinman G, Andel SA, Spector PE. Information security climate and the assessment of information security risk among healthcare employees. Health Informatics J 2020 Mar 14;26(1):461-473 [FREE Full text] [doi: 10.1177/1460458219832048] [Medline: 30866704]
60. Data Security and Protection Toolkit. NHS Digital. URL: https://digital.nhs.uk/data-and-information/looking-after-information/data-security-and-information-governance/data-security-and-protection-toolkit [accessed 2021-04-05]
61. Dameff CJ, Selzer JA, Fisher J, Killeen JP, Tully JL. Clinical Cybersecurity Training Through Novel High-Fidelity Simulations. J Emerg Med 2019 Feb;56(2):233-238. [doi: 10.1016/j.jemermed.2018.10.029] [Medline: 30553562]
62. Kabir UY, Ezekekwu E, Bhuyan SS, Mahmood A, Dobalian A. Trends and best practices in health care cybersecurity insurance policy. J Healthc Risk Manag 2020 Oct 22;40(2):10-14. [doi: 10.1002/jhrm.21414] [Medline: 32441812]
63. Yaseen M, Saleem K, Orgun MA, Derhab A, Abbas H, Al-Muhtadi J, et al. Secure sensors data acquisition and communication protection in eHealthcare: Review on the state of the art. Telematics and Informatics 2018 Jul;35(4):702-726. [doi: 10.1016/j.tele.2017.08.005]
64. Raisaro J, McLaren PJ, Fellay J, Cavassini M, Klersy C, Hubaux J, Swiss HIV Cohort Study. Are privacy-enhancing technologies for genomic data ready for the clinic? A survey of medical experts of the Swiss HIV Cohort Study. J Biomed Inform 2018 Mar;79:1-6 [FREE Full text] [doi: 10.1016/j.jbi.2017.12.013] [Medline: 29331453]
65. Narikimilli NRS, Kumar A, Antu AD, Xie B. Blockchain Applications in Healthcare – A Review and Future Perspective. In: Chen Z, Cui C, Palanisamy BB, Zhang LJ, editors. Blockchain – ICBC 2020. ICBC 2020. Lecture Notes in Computer Science, vol 12404. Cham: Springer; 2020:198-218.
66. Gardiyawasam Pussewalage HS, Oleshchuk VA. Privacy preserving mechanisms for enforcing security and privacy requirements in E-health solutions. International Journal of Information Management 2016 Dec;36(6):1161-1173. [doi: 10.1016/j.ijinfomgt.2016.07.006]
67. Wang L, Jones R. Big Data, Cybersecurity, and Challenges in Healthcare. 2019 Presented at: 2019 SoutheastCon; April 10-14; Huntsville, AL p. 1-6. [doi: 10.1109/SoutheastCon42311.2019.9020632]
68. Securing cyber resilience in health and care: progress update 2019. UK Government. 2019 Nov 4. URL: https://www.gov.uk/government/publications/securing-cyber-resilience-in-health-and-care-progress-update-2019 [accessed 2021-04-05]
69. Bhuyan SS, Kabir UY, Escareno JM, Ector K, Palakodeti S, Wyant D, et al. Transforming Healthcare Cybersecurity from Reactive to Proactive: Current Status and Future Recommendations. J Med Syst 2020 Apr 02;44(5):98. [doi: 10.1007/s10916-019-1507-y] [Medline: 32239357]
70. He Y, Johnson C. Improving the redistribution of the security lessons in healthcare: An evaluation of the Generic Security Template. Int J Med Inform 2015 Nov;84(11):941-949. [doi: 10.1016/j.ijmedinf.2015.08.010] [Medline: 26363788]
71. He Y, Johnson C. Challenges of information security incident learning: An industrial case study in a Chinese healthcare organization. Inform Health Soc Care 2017 Dec 09;42(4):393-408. [doi: 10.1080/17538157.2016.1255629] [Medline: 28068150]
72. COVID-19 cyber security support. NHS Digital. URL: https://digital.nhs.uk/cyber-and-data-security/covid-19-cyber-security-support [accessed 2021-04-05]
73. Data protection and coronavirus information hub. Information Commissioner's Office. URL: https://ico.org.uk/global/data-protection-and-coronavirus-information-hub/ [accessed 2021-04-05]

Abbreviations

DDoS: distributed denial of service
FBI: Federal Bureau of Investigation
IoT: internet of things
IT: information technology
NCSC: National Cyber Security Centre
NHS: National Health Service
NIST: National Institute of Standards and Technology
PPE: personal protective equipment

PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews
VPN: virtual private network
WHO: World Health Organization

Edited by C Basch; submitted 23.06.20; peer-reviewed by C Kruse, A Phillips; comments to author 17.11.20; revised version received 08.12.20; accepted 21.02.21; published 20.04.21.

Please cite as:
He Y, Aliyu A, Evans M, Luo C
Health Care Cybersecurity Challenges and Solutions Under the Climate of COVID-19: Scoping Review
J Med Internet Res 2021;23(4):e21747
URL: https://www.jmir.org/2021/4/e21747
doi:10.2196/21747
PMID:33764885

Original Paper

Participation in Virtual Urology Conferences During the COVID-19 Pandemic: Cross-sectional Survey Study

Menghua Wang[1*], MD; Banghua Liao[1*], MD; Zhongyu Jian[1], MD; Xi Jin[1], MD; Liyuan Xiang[1], MA; Chi Yuan[1], MD; Hong Li[1], MD; Kunjie Wang[1], MD, PhD

Department of Urology, Institute of Urology (Laboratory of Reconstructive Urology), West China Hospital, Sichuan University, Chengdu, China

[*]these authors contributed equally

Corresponding Author:
Kunjie Wang, MD, PhD
Department of Urology
Institute of Urology (Laboratory of Reconstructive Urology)
West China Hospital, Sichuan University
Chengdu
China
Phone: 86 189 8060 1848
Email: wangkj@scu.edu.cn

Abstract

Background: Due to the influence of the COVID-19 pandemic, conventional face-to-face academic conferences have been restricted, and many of these conferences have moved onto the internet.

Objective: The aim of this study was to investigate the virtual conferences in the field of urology during the COVID-19 pandemic and provide suggestions for better organization of such conferences.

Methods: A cross-sectional survey was conducted from May 30 to June 15, 2020, in China. Our team designed a 23-item questionnaire to investigate the conferences attended by urologists during the COVID-19 pandemic. SPSS 22.0 (IBM Corporation) was applied to analyze the data collected.

Results: A total of 330 Chinese urologists participated in our survey, and the response rate was 89.7% (330/368). Among the participants, 40.9% (135/330) were associate chief physicians. The proportion of participants who took part in conventional face-to-face academic conferences decreased from 92.7% (306/330) before the COVID-19 pandemic to 22.1% (73/330) during the pandemic (P<.001). In contrast, the proportion of urologists who took part in virtual conferences increased from 69.4% (229/330) to 90% (297/330) (P<.001). Most urologists (70.7%, 210/297) chose to participate in the virtual conferences at home and thought that a meeting length of 1-2 hours was most appropriate. Among the urologists, 73.7% (219/297) reported that their participation in the virtual conferences went smoothly, while the remaining respondents reported that they had experienced lags in video and audio streaming during the virtual conferences. When comparing conventional face-to-face conferences with virtual conferences, 70.7% (210/297) of the respondents thought that both conference formats were acceptable, while 17.9% (53/297) preferred virtual conferences and 11.5% (34/297) preferred conventional face-to-face meetings.

Conclusions: Virtual conferences are increasing in popularity during the COVID-19 pandemic; however, many aspects of these conferences could be improved for better organization.

(J Med Internet Res 2021;23(4):e24369) doi:10.2196/24369

KEYWORDS

virtual conference; COVID-19; survey

Introduction

Since the first case of COVID-19 was reported in Wuhan, China, the disease has spread worldwide rapidly [1]. To prevent the spread of COVID-19, many mass gathering events have been canceled or postponed, including the annual meetings of the European Association of Urology (EAU) and the American Urological Association [2,3]. In this condition, the role of virtual conferences is expanding [4], and many academic conferences moved onto the internet, such as the EAU20 Virtual Congress [5].

However, to our knowledge, there is no study investigating virtual conferences in the field of urology. In this study, we aimed to investigate the virtual meetings during the COVID-19 pandemic in China and provide advice on organizing efficient and high-quality virtual conferences according to the results of our survey.

Methods

Our team designed a structured questionnaire after reviewing the current relevant literature. A total of 23 items were finalized for this survey, covering demographics (2 items), traditional academic conferences (5 items), virtual (video) conferences (11 items), comparison between traditional and virtual conferences (3 items), and topics of academic conferences (2 items). We have provided the questionnaire in Multimedia Appendix 1.

We collected the data through the *Wenjuanxing* website [6] from May 30 to June 15, 2020. All participants accessed the questionnaire by clicking on the survey link. Participants were able to submit the questionnaires only when all the items were completed. To avoid repetition, IP restriction was applied, which meant that the survey could only be completed once from a single IP address.

Data are presented as number (percentage) for categorical variables, and chi-square tests were used to compare the distributions of categorical variables. All the statistical analyses were conducted using SPSS 20.0 (IBM Corporation).

Results

From May 30 to June 15, 2020, we sent the web-based questionnaire to 368 Chinese urologists; 330 of them completed the survey, resulting in a response rate of 89.7%. Among the participants, 96.4% (318/330) were based at secondary grade A hospitals and above. Regarding their academic titles, most of the participants were associate chief physicians (135/330, 40.9%), followed by attending physicians (113/330, 34.2%), chief physicians (61/330, 18.5%), and residents (21/330, 6.4%) (Table 1).

Table 1. Baseline information about the survey participants (N=330).

Characteristic	Value, n (%)
Hospital level[a]	
Tertiary grade A	168 (50.9)
Tertiary grade B	97 (29.4)
Secondary grade A	53 (16.1)
Secondary grade B	6 (1.8)
Primary and others	6 (1.8)
Academic title[b]	
Chief physician	61 (18.5)
Associate chief physician	135 (40.9)
Attending physician	113 (34.2)
Resident	21 (6.4)

[a]In China, hospitals are classified as primary, secondary, or tertiary institutions. Typically, primary hospitals contain less than 100 beds and secondary hospitals contain 100-500 beds, while tertiary hospitals have a bed capacity exceeding 500. According to their medical quality, etc, each hospital level is further subdivided into three subsidiary grades of A, B, and C, which results in a total of 9 levels.

[b]For the physicians' hierarchy in China, there are four levels in total, namely chief physician, associate chief physician, attending physician, and resident, from senior to junior.

The proportion of urologists participating in conventional face-to-face conferences decreased from 92.7% (306/330) before the COVID-19 pandemic to 22.1% (73/330) during the pandemic (P<.001). Generally, 89.5% (274/306) of the urologists were satisfied with the traditional conferences, while others (32/306, 10.5%) thought that live conferences were time-consuming (27/32, 84%,) and expensive (20/32, 62.5%).

The proportion of urologists participating in virtual conferences increased from 69.4% (229/330) before the COVID-19 pandemic to 90% (297/330) during the pandemic (P<.001). Of the participants, 99.0% declared that they were "very satisfied" or "satisfied" with the virtual conferences.

Most of the urologists (70.7%, 210/297) chose to take part in the virtual conferences at home. In terms of the length of the virtual conferences, most participants (62.3%, 185/297) reported that they usually spent 0.5-1.5 hours per meeting, and 66.3% (197/297) thought that 1-2 hours is the most appropriate duration for a meeting.

Regarding the software, Tencent Meeting, DingTalk, and Zoom were the most commonly used platforms in China. Of the urologists, 73.7% (219/297) reported that their participation in the virtual meetings went smoothly, while the remaining urologists said that due to limited bandwidth or unstable web-based platforms, they experienced lags in video and audio streaming during the virtual conferences. According to our survey, the urologists thought that a good web-based platform should provide on-demand video and should be stable and interactive.

When comparing conventional face-to-face conferences with virtual conferences, 70.7% (210/297) of the participants thought that both formats were acceptable, while 17.9% (53/297) preferred virtual conferences and 11.5% (34/297) preferred conventional face-to-face meetings. Compared with conventional face-to-face meetings, the urologists reported that virtual conferences were time-effective, less expensive, and more convenient; however, they also lacked interaction, body language, and Continuing Medical Education (CME) credits.

For the different topics of urological conferences, stones (286/330, 86.7%), laparoscopy (84.9%, 280/330), urological oncology (84.6%, 279/330), and prostatic diseases (74.9%, 247/330) were the four most popular topics, followed by andrology (177/330, 53.6%), female urology (140/330, 42.4%), reconstructive urology (130/330, 39.4%), and kidney transplantation (17/330, 5.2%). When the participants were asked about what they expected to learn about during urological conferences, clinical experience and surgical techniques were selected by 95.8% (316/330) and 94.2% (311/330) of participants, while the methodology of medical research and basic research progress were only chosen by 46.4% (153/330) and 39.1% (129/330) of the participants. Further analysis showed that there was no significant difference among the different levels of hospitals for interest in clinical experience (Figure 1), surgical technique (Figure 2), or basic research progress (Figure 3), while fewer urologists based at secondary grade A hospitals and below were interested in the methodology of medical research (Figure 4).

Figure 1. Proportions of participants interested in learning about clinical experience at urology conferences stratified by hospital level. *$P>.05$.

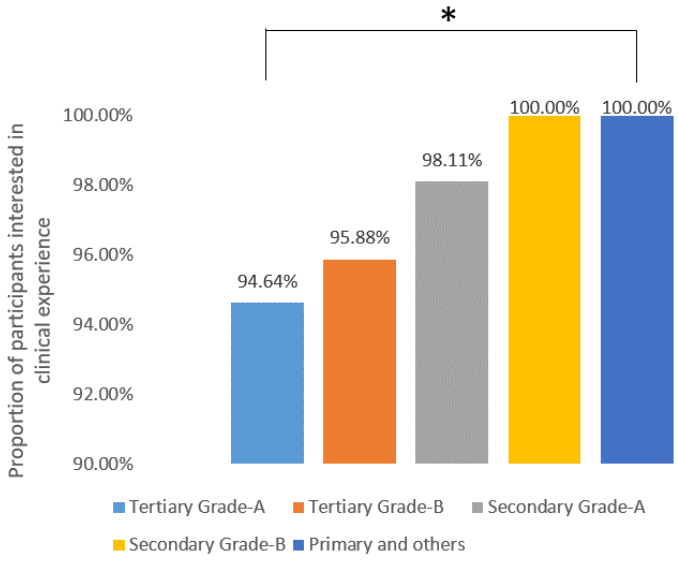


Figure 2. Proportions of participants interested in learning about surgical technique at urology conferences stratified by hospital level.*$P>.05$.

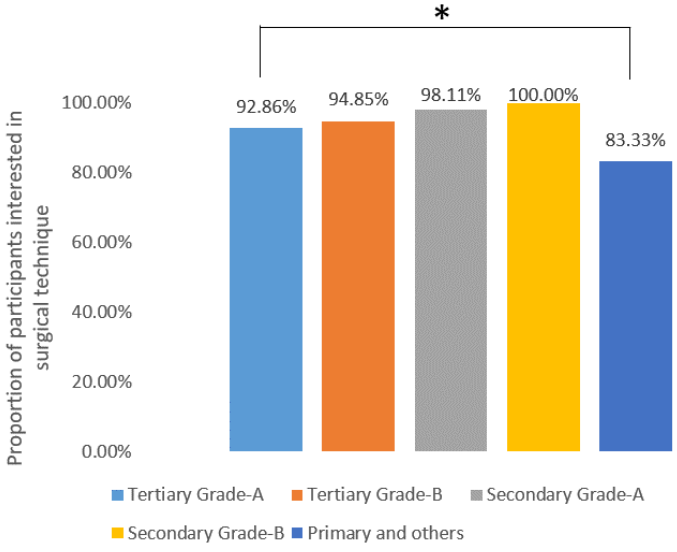


Figure 3. Proportions of participants interested in learning about basic research progress at urology conferences stratified by hospital level.*$P>.05$.

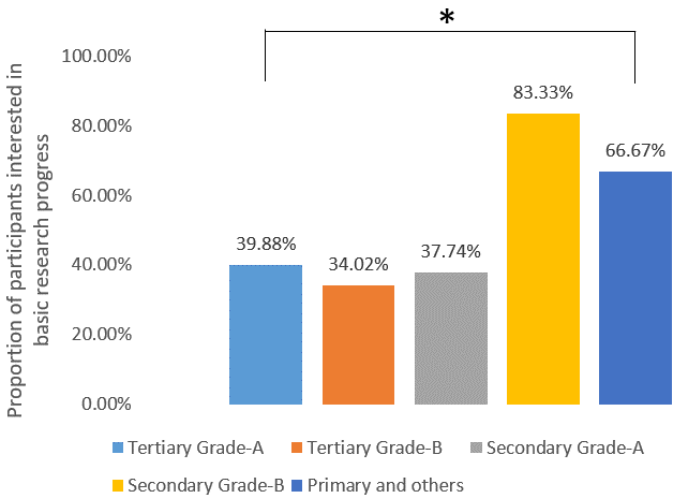
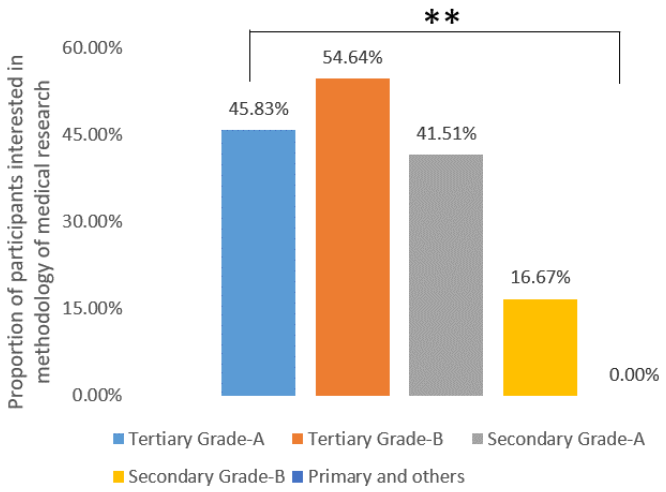


Figure 4. Proportions of participants interested in learning about the methodology of medical research at urology conferences stratified by hospital level. **$P<.05$.



Discussion

Principal Findings

As far as we know, this is the first survey to investigate urological conferences during the COVID-19 pandemic. A total of 330 Chinese urologists, with different academic titles and from different levels of hospitals, were included in our survey. Our study revealed that although virtual conferences are increasing in popularity during the COVID-19 pandemic, the organization of these virtual conferences needs further improvement.

During the COVID-19 pandemic, traditional face-to-face meetings have been restricted, and it is more important than ever to maximize virtual meetings. In our survey, the proportion of urologists who took part in virtual conferences increased from 69.4% before the COVID-19 pandemic to 90.0% during the pandemic, indicating a sharp increase in virtual conference participation. Additionally, due to the development of efficient

telecommunication networks, technological capabilities are no longer a hindrance, and these networks have been used in various areas [7,8]. Recently, the Human Genome Meeting, Global Genomic Medicine Collaborative, and Trans Tasman Radiation Oncology Group Annual Scientific Meeting were delivered virtually and achieved great success, with 97% of the participants being "very satisfied or satisfied" with the virtual style [9,10]; this was similar to our 99.0% proportion of satisfied participants, indicating that virtual conferences were well accepted by urologists. As a result, we should maximize the use of virtual conferences to ensure communication in the medical community during the COVID-19 pandemic.

When preparing for virtual conferences, the first task is to define the target audience, as different audiences may have different appetites. For example, in our study, the urologists who worked at secondary grade A hospitals and below had lower enthusiasm for the topic of medical research methodology. Second, according to our survey, most people chose to take part in conferences at home; therefore, avoiding early morning and late evening meetings would provide them with a better experience [11]. Web-based platform selection is another essential aspect of the success of virtual conferences. In our survey, Tencent Meeting, DingTalk, and Zoom were the most commonly used platforms. In addition to these platforms, webinar technology [12], Twitter [13], journal clubs, podcasts, *Intouch Vita*, etc, were used for virtual meetings [14]. However, regardless of which platform is chosen, it should ensure a smooth and stable video experience; this was what people were most concerned about according to our survey.

During virtual conferences, the duration of the sessions should be controlled, as the average attention span for adults is only approximately 15-20 minutes [15]. In our study, most of the urologists thought that 1-2 hours was the most appropriate length. Salomon and their team [16] designed each session at their virtual conference to last 2 hours, and this schedule was appreciated by most of the participants (135/181, 74.6%), which was in accordance with our respondents' preference of 1-2 hours. As a result, the duration of an academic meeting should be controlled within approximately 2 hours. Additionally, as mentioned before, one of the limitations of virtual conferences is the lack of interaction. To solve this problem, organizers could increase the time allotted to question-and-answer periods [10], permit attendees to turn on their cameras, provide virtual open discussion chat rooms [16], etc.

When a virtual conference has ended, the organizer should provide on-demand content such as video on the virtual platform. On-demand video is important for people in different time zones and is valuable for all participants to catch up on missed talks. Over 90% of the urologists in our survey thought that on-demand video was essential for their learning. Additionally, the physicians who attend virtual conferences should be provided with CME credits. In the era of information explosion, to enable physicians to remain current with their medical knowledge, CME has been implemented in many countries worldwide, and a variable number of CME credits is required [17,18]. Normally, CME credits can be earned through attendance at in-person academic conferences, workshops, symposiums, etc. However, according to our survey, some urologists reported that during virtual conferences, CME credits were not made available to participants, which in our view may reduce the enthusiasm of urologists to take part in these conferences. As a result, in the future, we believe that virtual conferences should also provide CME credits.

In terms of the comparison between virtual conferences and conventional face-to-face conferences, as mentioned before, both formats have their advantages and disadvantages. As a result, some researchers have proposed a new type of meeting called a hybrid meeting, which combines aspects of virtual and live conferences. In a survey conducted by Forrest et al [10], 85% of the delegates reported that they preferred hybrid meetings compared with virtual or physical meetings only. Additionally, Gomez Rivas et al [19] shared their preliminary experience of hybrid meetings that included both onsite and web-based participants. Moreover, we believe that this type of meeting would be the best choice for academic conferences in the future, as it combines the advantages of both formats.

Limitations

There are several limitations to our survey. First, our survey had a relatively small sample size. Second, most of the participants in this study were from tertiary hospitals. Despite these two limitations, our study provides a preliminary report of virtual conferences during the COVID-19 pandemic and provides many suggestions on how to organize an efficient and satisfying virtual conference, which is timely and useful.

Conclusion

Virtual conferences are increasing in popularity during the COVID-19 pandemic; however, many aspects of these conferences could be improved for better organization.

Acknowledgments

This work was funded by the 1.3.5 project for disciplines of excellence, West China Hospital, Sichuan University (ZYGD18011, ZYJC18015, ZY2016104).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Detailed questionnaire.

[DOCX File , 18 KB - jmir_v23i4e24369_app1.docx]

References

1. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, China Novel Coronavirus Investigating and Research Team. A novel coronavirus from patients with pneumonia in China, 2019. N Engl J Med 2020 Feb 20;382(8):727-733 [FREE Full text] [doi: 10.1056/NEJMoa2001017] [Medline: 31978945]
2. Porpiglia F, Checcucci E, Autorino R, Amparore D, Cooperberg MR, Ficarra V, et al. Traditional and virtual congress meetings during the COVID-19 pandemic and the post-COVID-19 era: is it time to change the paradigm? Eur Urol 2020 Sep;78(3):301-303 [FREE Full text] [doi: 10.1016/j.eururo.2020.04.018] [Medline: 32334884]
3. Puliatti S, Eissa A, Eissa R, Amato M, Mazzone E, Dell'Oglio P, et al. COVID-19 and urology: a comprehensive review of the literature. BJU Int 2020 Jun;125(6):E7-E14. [doi: 10.1111/bju.15071] [Medline: 32249538]
4. Rubinger L, Gazendam A, Ekhtiari S, Nucci N, Payne A, Johal H, et al. Maximizing virtual meetings and conferences: a review of best practices. Int Orthop 2020 Aug;44(8):1461-1466 [FREE Full text] [doi: 10.1007/s00264-020-04615-9] [Medline: 32445031]
5. Pang KH, Carrion DM, Rivas JG, Mantica G, Mattigk A, Pradere B, European Society of Residents in Urology. The impact of COVID-19 on European health care and urology trainees. Eur Urol 2020 Jul;78(1):6-8 [FREE Full text] [doi: 10.1016/j.eururo.2020.04.042] [Medline: 32376133]
6. Wenjuanxing. Website in Chinese. URL: https://www.wjx.cn/ [accessed 2021-04-13]
7. Hong Z, Li N, Li D, Li J, Li B, Xiong W, et al. Telemedicine during the COVID-19 pandemic: experiences from western China. J Med Internet Res 2020 May 08;22(5):e19577 [FREE Full text] [doi: 10.2196/19577] [Medline: 32349962]
8. Fisk M, Livingstone A, Pit SW. Telehealth in the context of COVID-19: changing perspectives in Australia, the United Kingdom, and the United States. J Med Internet Res 2020 Jun 09;22(6):e19264 [FREE Full text] [doi: 10.2196/19264] [Medline: 32463377]
9. McDowell L, Goode S, Sundaresan P. Adapting to a global pandemic through live virtual delivery of a cancer collaborative trial group conference: The TROG 2020 experience. J Med Imaging Radiat Oncol 2020 Jun;64(3):414-421. [doi: 10.1111/1754-9485.13047] [Medline: 32304265]
10. Forrest ARR, Repetto GM, Reichardt JKV. Human genetics and genomics meetings going virtual: practical lessons learned from two international meetings in early 2020. Hum Genomics 2020 Jul 08;14(1):27 [FREE Full text] [doi: 10.1186/s40246-020-00275-3] [Medline: 32641116]
11. Terhune KP, Choi JN, Green JM, Hildreth AN, Lipman JM, Aarons CB, et al. Ad astra per aspera (through hardships to the stars): lessons learned from the First National Virtual APDS Meeting, 2020. J Surg Educ 2020;77(6):1465-1472 [FREE Full text] [doi: 10.1016/j.jsurg.2020.06.015] [Medline: 32646812]
12. Jell A, Vogel T, Ostler D, Marahrens N, Wilhelm D, Samm N, et al. 5th-generation mobile communication: data highway for Surgery 4.0. Surg Technol Int 2019 Nov 10;35:36-42. [Medline: 31694061]
13. Mackenzie G, Murray AD, Oliver CW. Virtual attendance at an international physical activity meeting using Twitter: how can data visualisation provide a presence? Br J Sports Med 2018 Mar;52(6):351-352. [doi: 10.1136/bjsports-2016-097373] [Medline: 28254745]
14. Porpiglia F, Checcucci E, Amparore D, Verri P, Campi R, Claps F, et al. Slowdown of urology residents' learning curve during the COVID-19 emergency. BJU Int 2020 Jun;125(6):E15-E17 [FREE Full text] [doi: 10.1111/bju.15076] [Medline: 32274879]
15. Bradbury NA. Attention span during lectures: 8 seconds, 10 minutes, or more? Adv Physiol Educ 2016 Dec 01;40(4):509-513 [FREE Full text] [doi: 10.1152/advan.00109.2016] [Medline: 28145268]
16. Salomon D, Feldman MF. The future of conferences, today: are virtual conferences a viable supplement to "live" conferences? EMBO Rep 2020 Jul 03;21(7):e50883. [doi: 10.15252/embr.202050883] [Medline: 32510697]
17. Kearney P, Simoons M, Ryden L, Kirchhof P, Pries A, O'Morain C, et al. The medical profession, industry, and continuing medical education: finding the balance that's right for patients. Am J Med 2019 Aug;132(8):921-925. [doi: 10.1016/j.amjmed.2019.02.026] [Medline: 30853472]
18. Ghanem O, Logghe HJ, Tran BV, Huynh D, Jacob B. Closed Facebook groups and CME credit: a new format for continuing medical education. Surg Endosc 2019 Feb;33(2):587-591. [doi: 10.1007/s00464-018-6376-9] [Medline: 30105596]
19. Gómez Rivas J, Carrion DM, Tortolero L, Veneziano D, Esperto F, Greco F, et al. Scientific social media, a new way to expand knowledge. What do urologists need to know? Actas Urol Esp 2019 Jun;43(5):269-276. [doi: 10.1016/j.acuro.2018.12.003] [Medline: 31014549]

Abbreviations

CME: Continuing Medical Education
EAU: European Association of Urology

Edited by G Eysenbach; submitted 17.09.20; peer-reviewed by R Watson, G Kernohan; comments to author 18.11.20; revised version received 30.01.21; accepted 07.04.21; published 21.04.21.

Please cite as:
Wang M, Liao B, Jian Z, Jin X, Xiang L, Yuan C, Li H, Wang K
Participation in Virtual Urology Conferences During the COVID-19 Pandemic: Cross-sectional Survey Study
J Med Internet Res 2021;23(4):e24369
URL: https://www.jmir.org/2021/4/e24369
doi:10.2196/24369
PMID:33844635

Short Paper

Decline of Psychological Health Following the Designation of COVID-19 as a Pandemic: Descriptive Study

Darpan I Patel[1], BSc, MA, PhD; Yazmin Gamez[1], BSN; Lajja Shah[2], MASLP; Jaini Patel[2], MPT

[1]Biobehavioral Research Laboratory, School of Nursing, University of Texas Health Science Center at San Antonio, San Antonio, TX, United States
[2]Department of Rehabilitation & Sports Medicine, Sir H N Reliance Foundation Hospital, Mumbai, India

Corresponding Author:
Darpan I Patel, BSc, MA, PhD
Biobehavioral Research Laboratory
School of Nursing
University of Texas Health Science Center at San Antonio
7703 Floyd Curl Drive
San Antonio, TX
United States
Phone: 1 210 567 0362
Email: pateld7@uthscsa.edu

Abstract

Background: COVID-19 was declared a pandemic by the World Health Organization on March 11, 2020, and as of this writing, Texas, United States, has reported >675,000 cases with over 14,000 deaths. Many of the preventive measures implemented during the pandemic can increase sedentary lifestyles, which can lead to the development of chronic diseases, including obesity, among the general population and cause serious threats to people's physical health and overall quality of life. Individuals with pre-existing comorbidities are at an increased risk of COVID-19 and may hence have higher levels of stress.

Objective: This study aimed to investigate the relationship between physical activity levels and mental health status on an individual level and to compare them between those with and those without comorbidities in a cohort of Texas residents, before and after COVID-19 was declared a pandemic.

Methods: An electronic survey was disseminated throughout various regions of Texas. In total, 160 individuals were asked questions about their demographic characteristics, time spent on daily physical activities, and daily mental health status before and after COVID-19 was declared a pandemic. Frequency distributions and descriptive statistics were analyzed.

Results: Overall, 94 (58%) participants reported having ≥1 medical condition, and 31 (13.1%) had >3 medical conditions. Physical activity levels among participants with ≥1 pre-existing comorbidity drastically—but not significantly—decreased, as evident from a 10% increase in sedentary lifestyles after COVID-19 was declared a pandemic. On the contrary, we observed a 9% increase in the number of individuals without a pre-existing comorbidity who reported 30-60 min of physical activity per week. There was a 2-fold increase in the number of participants reporting more frequent feelings of nervousness, too much worry, trouble relaxing, and the fear of something awful happening after the pandemic. More specifically, individuals with pre-existing medical conditions reported, on average, a 10% higher incidence of feelings of stress, anxiety, and sadness compared to their healthy counterparts after COVID-19 was declared a pandemic.

Conclusions: Stressful life conditions and chronic comorbidities are risk factors that can affect mental health and reduce the ability to perform activities of daily life. Therefore, when implementing pandemic protocols, municipalities should consider providing mental health support to their citizens to protect them from this rather inconspicuous adverse effect.

(J Med Internet Res 2021;23(4):e24964) doi:10.2196/24964

KEYWORDS

anxiety; COVID-19; descriptive study; mental health; pandemic; physical health; quality of life; stress

Introduction

COVID-19 was declared a pandemic by the World Health Organization on March 11, 2020 [1], and Texas has reported >675,000 cases and 14,000 deaths as of this writing. The routines of individuals have been impacted since the beginning of the pandemic owing to the implementation of restrictions and preventative measures such as social distancing, self-isolation, stay-at-home orders, and closure of businesses and recreational facilities. Many of these protocols can have a negative effect, promoting a sedentary lifestyle and decreasing physical activity levels, which can lead to the development or exacerbation of chronic diseases and mental health issues [2-5]. This study aimed to explore the relationship between physical activity levels and mental health status at an individual level in a cohort of Texas residents before and after COVID-19 was declared a pandemic. We hypothesize that the declaration of COVID-19 as a pandemic, which led to the implementation of social distancing and self-isolation protocols, would negatively impact mental health, increase stress and anxiety, and decrease physical activity levels among all study participants.

Methods

Recruitment

Texas residents aged ≥18 years were recruited through social media and email, using convenience and snowball sampling over a 2-week period in July 2020. Respondent identities were anonymized, and the study was exempt from ethics review for the use of human subjects by the institutional review board of the University of Texas Health Science Center at San Antonio.

Survey Instrument

A shortened survey instrument initially developed by Flanagan et al [6] was used to capture respondent demographics, medical history, and physical and mental health status. Questions related to physical and mental health were constructed in a manner to determine health status before and after the implementation of COVID-19 quarantine protocols. Study data were collected and managed using Research Electronic Data Capture (REDCap) [7] tools hosted at the University of Texas Health Science Center at San Antonio. REDCap is a secure web-based software platform designed to support data capture for research studies, providing (1) an intuitive interface for validated data capture, (2) audit trails for tracking data manipulation and export procedures, (3) automated export procedures for seamless data downloads to common statistical packages, and (4) procedures for data integration and interoperability with external sources.

Demographics and Medical History

Data on the participants' gender (male, female, or no response), age, marital status, employment status, household number, ethnicity, and race were collected. Participants were asked to indicate whether they have been diagnosed with any chronic medical conditions, including cardiovascular, respiratory, gastrointestinal, genitourinary, kidney, hematologic, infectious, dermatological, ophthalmologic, endocrine, musculoskeletal, oncologic, or neurologic disease. The demographic characteristics of the survey participants are presented in Table 1.

Table 1. Demographic characteristics of the survey participants (N=160).

Characteristic	Number of participants, n (%)
Race	
American Indian or Alaskan native	0 (0)
Asian	10 (6.3)
Black or African American	3 (1.9)
Native or Pacific Islander	3 (1.9)
White	134 (83.8)
Not reported	8 (5.0)
Unknown	2 (1.3)
Ethnicity	
Hispanic or Latino	65 (40.9)
Not Hispanic or Latino	86 (53.9)
Not reported	7 (4.5)
Unknown	2 (1.3)
Gender	
Female	138 (86.3)
Male	20 (12.5)
Not reported	2 (1.3)
Age (years)	
18-25	17 (10.9)
26-35	44 (27.6)
36-45	30 (18.6)
46-55	34 (21.2)
56-65	28 (17.3)
66-75	7 (4.4)
Pre-existing diagnosed medical conditions (n=94; percentages reported considering a sample of 160 participants)	
1	52 (32.5)
2	21 (13.1)
3	13 (8.1)
≥4	8 (5.0)

Questions on Physical Activity

Participants were asked to provide details regarding their physical activity levels before and after COVID-19 was declared a pandemic. Specifically, they were asked to provide details regarding their physical activity or exercise minutes per day. Participants were provided the following categorical variables as choices: 0-30 minutes, 30-60 minutes, 60-90 minutes, 90-120 minutes, or >120 minutes of daily physical activity.

Questions Related to Psychological Status and Mental Health

Participants were asked if they were worried about their physical health with respect to the COVID-19 pandemic. They were also asked about generalized stress, anxiety, and sadness before and after the COVID-19 pandemic. Furthermore, 8 questions were related to the participants (1) feeling nervous, anxious, or on edge; (2) not being able to stop of control worry; (3) worrying too much about different things; (4) having trouble relaxing; (5) being restless and finding it difficult to sit still; (6) becoming easily annoyed or irritable; (7) feeling afraid that something awful might happen; and (8) having difficulty getting things done because of these mental health issues. Participants selected either (1) not at all, (2) several days, (3) over half the days, or (4) nearly every day. Each response was converted to a numerical score for analysis.

Statistical Analysis

Descriptive statistics were calculated to summarize the response frequency. Respondents who declined to answer a question were considered missing and excluded from the calculated proportions. We report all summary statistics. For paired testing

of questions that asked for pre– and post–COVID-19 comparisons, we used a paired samples *t* test to analyze mean values. The significance level was set at Cronbach α=.05, and all tests were 2-sided. All data were processed using SPSS for Windows (version 23.0, SPSS Inc).

Results

Participant Demographics

In total, 160 individuals responded to the social media posts. The vast majority of respondents were married, white, non-Hispanic or -Latino females. Overall, 65 (41%) respondents identified themselves as Hispanic or Latino (Table 1), which was more than the 31% estimated census of Hispanic or Latinos in Texas. More than half of the respondents (n=91, 57%) were aged <45 years and had full-time jobs (n=117, 73%), while 30 (19%) indicated that they were unemployed.

Furthermore, 94 (58%) participants reported having ≥1 pre-existing medical condition, and 21 (13.6%) respondents had >3 pre-existing medical conditions. Medical conditions with the highest frequency included hypertension (22.3% of all responses), respiratory problems (14.0%), endocrine disorders (13.4%), and gastrointestinal problems (12.7%). Participant responses to questions on pre-existing medical conditions are presented in Table 2.

Table 2. Distribution of the responses of participants with pre-existing medical conditions (N=160).

Pre-existing medical condition	Number of responses, n (%)[a]
Cardiac or heart disease	6 (3.8)
Respiratory problems	22 (14.0)
Gastrointestinal problems	20 (12.7)
Genitourinary or kidney disease	5 (3.2)
Hematologic condition	10 (6.4)
Infectious disease	1 (0.6)
Dermatologic condition	13 (8.3)
Ophthalmologic condition	7 (4.5)
Endocrine conditions	21 (13.4)
Diabetes	10 (6.4)
Musculoskeletal conditions	8 (5.1)
Hypertension	35 (22.3)
Cancer	6 (3.8)
Neurologic or psychiatric condition	14 (8.9)

[a]Among 160 respondents, 94 reported having ≥1 pre-existing medical condition. Individuals were asked to mark all medical conditions they have been diagnosed with, allowing them to select >1 response. Percentages are in relation to the total sample of 160 survey responses.

Slight Reduction in Physical Activity in Response to the Declaration of COVID-19 as a Pandemic

No significant differences in self-reported physical activity were observed for the entire cohort of respondents. Similarly, no significant differences were observed in the subgroups of individuals with (*P*=.81) or without (*P*=.91) pre-existing medical conditions. Physical activity was slightly—but not significantly (*P*=.81)—reduced among participants with ≥1 pre-existing condition, as evident from a 10% increase in sedentary lifestyles after COVID-19 was declared a pandemic. On the contrary, we observed a 9% increase in the number of individuals without a pre-existing conditions reporting 30-60 min of physical activity per week.

Negative Effect of the COVID-19 Pandemic on Psychological Health

Respondents were asked how the quarantine protocols in Texas impacted their psychological health (Table 3). We observed significant increases in feelings of fear, annoyance, restlessness, worry, nervousness, sadness, anxiety, and stress (*P*<.001) among the survey respondents. Specifically, participants reported more frequently feeling nervous, worrying too much, having trouble relaxing, and feeling afraid something awful might happen after the pandemic. This negative impact on psychological well-being interfered with their ability to accomplish their work, tasks, or interact with other people. Similar results were obtained among individuals with and those without pre-existing medical conditions. The only exception is that individuals without pre-existing medical conditions reported that the COVID-19 pandemic had no significant impact on their ability to relax (*P*=.08).

Table 3. Changes in psychological health resulting from the COVID-19 pandemic.

Items	All respondents			Respondents with pre-existing medical conditions			Respondents without pre-existing medical conditions		
	Pre–COVID-19, mean (SD)	Post–COVID-19[a], mean (SD)	*P* value	Pre–COVID-19, mean (SD)	Post–COVID-19, mean (SD)	*P* value	Pre–COVID-19, mean (SD)	Post–COVID-19, mean (SD)	*P* value
Feeling nervous, anxious, or on edge	1.78 (0.77)[b]	2.36 (0.91)	<.001	1.88 (0.77)	2.53 (0.98)	<.001	1.63 (0.75)	2.13 (0.74)	<.001
Not being able to stop or control your worry	1.57 (0.82)	1.95 (0.93)	<.001	1.63 (0.84)	2.04 (1.03)	<.001	1.41 (0.80)	1.82 (0.77)	.003
Worrying too much about different things	1.73 (0.81)	2.31 (1.02)	<.001	1.78 (0.75)	2.40 (1.06)	<.001	1.68 (0.90)	2.18 (0.98)	<.001
Trouble relaxing	1.70 (0.89)	2.09 (1.03)	<.001	1.81 (0.91)	2.29 (1.06)	<.001	1.58 (0.86)	1.84 (0.95)	0.08
Being so restless that it is difficult to sit still	1.27 (0.62)	1.55 (0.81)	<.001	1.31 (0.63)	1.64 (0.91)	.002	1.21 (0.62)	1.45 (0.65)	.05
Becoming easily annoyed or irritable	1.68 (0.67)	2.07 (0.86)	<.001	1.70 (0.68)	2.16 (0.88)	<.001	1.68 (0.66)	2.00 (0.81)	.01
Feeling afraid as if something awful might happen	1.41 (0.70)	2.04 (0.99)	<.001	1.45 (0.73)	2.12 (1.01)	<.001	1.38 (0.68)	1.92 (1.01)	<.001
How difficult did these factors make it for you to do your work, take care of things, or get along with other people	1.22 (0.73)	1.69 (0.86)	<.001	1.34 (0.92)	1.77 (0.95)	<.001	1.19 (0.66)	1.62 (0.72)	<.001

[a]After COVID-19 was declared a pandemic by the world Health Organization on March 11, 2020.

[b]Participants selected either of the following responses: (1) not at all, (2) several days, (3) over half the days, or (4) nearly every day. Each response was converted to a numerical score for analysis. For the question on difficulty level, participants selected either of the following responses: (0) I did not experience any of the conditions, (1) not difficult at all, (2) somewhat difficult, (3) very difficult, and (4) extremely difficult.

Discussion

Principal Findings

After COVID-19 was declared a pandemic, quarantine and social distancing protocols were implemented to help mitigate disease spread in Texas. The implementation of these protocols was accompanied by several adverse effects on quality of life, which were independent of the SARS-CoV-2 infection. With the requirement for more individuals to stay at home and in isolation from friends, family, and colleagues, physical and mental health can depreciate. Mental health was drastically affected in our respondents owing to social isolation measures implemented by local municipalities in Texas. Overall, our survey respondents reported feeling more stress, anxiety, and sadness after COVID-19 was declared a pandemic. However, the long-term implications of quarantine protocols on mental health remain unknown. However, evidence from natural disasters suggest that mental health symptoms peak in the following months and can persist for years [8]. For example, 5% of the population of Texas affected by hurricane Ike met the criteria for major depressive disorder in the months following the storm [9]. Therefore, while a significant impact is demonstratable approximately 4 months into the quarantining protocol, there is potential for even greater impacts in the subsequent months.

Quarantine protocols can have a serious impact on physical health, leading to the development and progression of chronic diseases [10]. The American College of Sports Medicine recommends that most adults engage in moderate-intensity exercise training for more than 30 minutes for more than 5 days a week [11]. Contrary to Tinson et al [2], participants with no pre-existing medical conditions increased their physical activity after COVID-19 was declared a pandemic, with more individuals meeting or exceeding the recommended physical activity guidelines of the American College of Sports Medicine [11]. In this study, physical activity levels slightly—but not significantly—decreased among participants with pre-existing medical conditions, which indicates a potential risk of decreased overall health, potentially resulting in the exacerbation of medical conditions. Given that the risk of COVID-19 is significantly higher in subgroups of the population, which have pre-existing medical conditions, such subgroups are expected to decrease the time they spend outside of home in fear of unintended exposure to the virus.

Limitations

Our study is limited by the self-reported nature of our survey and the convenience and snowball sampling method used to recruit participants. Further, without the use of geocoding, including zip codes, we are limited in our ability to generalize our results to different regions in Texas. Finally, our survey was

distributed only in English, although Spanish is also a dominant language in several areas of Texas.

Conclusions

The quarantine protocols implemented in Texas led to significant levels of stress and anxiety in our survey respondents. Physical activity can be considered a coping mechanism to alleviate the psychological distress people may feel. Psychological distress experienced during the COVID-19 pandemic is likely to have long-term clinical implications.

Acknowledgments

DIP received support in the form of grants from the Mays Cancer Center at University of Texas Health Science Center at San Antonio (National Cancer Institute grant# P30CA054174), the Department of Education (grant# P031S150048), San Antonio Claude D. Pepper Older Americans Independence Center (National Institute on Aging grant# P30 AG044271), and Molecular Transducers of Physical Activity Consortium Adult Clinical Center (grant# U01AR071150).

Authors' Contributions

DIP conceptualized the study. DIP, LS, and JP developed the survey questions. YG developed the web-based survey database. DIP and YG performed the data analysis. All authors contributed significantly to the study. DIP and YG drafted the manuscript, and LS and JP revised the manuscript. All authors approved the final version of this manuscript.

Conflicts of Interest

None declared.

References

1. WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020. World Health Organization. 2020 Mar 11. URL: https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020 [accessed 2021-04-05]
2. Tison GH, Avram R, Kuhar P, Abreau S, Marcus GM, Pletcher MJ, et al. Worldwide Effect of COVID-19 on Physical Activity: A Descriptive Study. Ann Intern Med 2020 Nov 03;173(9):767-770 [FREE Full text] [doi: 10.7326/M20-2665] [Medline: 32598162]
3. Ammar A, Brach M, Trabelsi K, Chtourou H, Boukhris O, Masmoudi L, et al. Effects of COVID-19 Home Confinement on Eating Behaviour and Physical Activity: Results of the ECLB-COVID19 International Online Survey. Nutrients 2020 May 28;12(6):1583 [FREE Full text] [doi: 10.3390/nu12061583] [Medline: 32481594]
4. Pancani L, Marinucci M, Aureli N, Riva P. Forced social isolation and mental health: A study on 1006 Italians under COVID-19 lockdown. PsyArXiv Preprint posted online April 5, 2020. [doi: 10.31234/osf.io/uacfj]
5. Kumar A, Nayar KR. COVID 19 and its mental health consequences. J Ment Health 2021 Feb 27;30(1):1-2. [doi: 10.1080/09638237.2020.1757052] [Medline: 32339041]
6. Flanagan E, Beyl RA, Fearnbach SN, Altazan AD, Martin CK, Redman LM. The Impact of COVID-19 Stay-At-Home Orders on Health Behaviors in Adults. Obesity (Silver Spring) 2021 Feb;29(2):438-445 [FREE Full text] [doi: 10.1002/oby.23066] [Medline: 33043562]
7. How REDCap is being used in response to COVID-19. REDCap. URL: https://www.project-redcap.org/ [accessed 2021-04-06]
8. Chapman D, Perry GS, Strine TW. The vital link between chronic disease and depressive disorders. Prev Chronic Dis 2005 Jan;2(1):A14 [FREE Full text] [Medline: 15670467]
9. Goldmann E, Galea S. Mental health consequences of disasters. Annu Rev Public Health 2014;35:169-183. [doi: 10.1146/annurev-publhealth-032013-182435] [Medline: 24159920]
10. Tee CA, Salido EO, Reyes PWC, Ho RC, Tee ML. Psychological State and Associated Factors During the 2019 Coronavirus Disease (COVID-19) Pandemic Among Filipinos with Rheumatoid Arthritis or Systemic Lupus Erythematosus. Open Access Rheumatol 2020;12:215-222 [FREE Full text] [doi: 10.2147/OARRR.S269889] [Medline: 33061689]
11. Garber C, Blissmer B, Deschenes MR, Franklin BA, Lamonte MJ, Lee IM, American College of Sports Medicine. American College of Sports Medicine position stand. Quantity and quality of exercise for developing and maintaining cardiorespiratory, musculoskeletal, and neuromotor fitness in apparently healthy adults: guidance for prescribing exercise. Med Sci Sports Exerc 2011 Jul;43(7):1334-1359. [doi: 10.1249/MSS.0b013e318213fefb] [Medline: 21694556]

Edited by C Basch; submitted 12.10.20; peer-reviewed by P Jimenez, R Ho; comments to author 23.12.20; revised version received 23.02.21; accepted 18.03.21; published 22.04.21.

Please cite as:
Patel DI, Gamez Y, Shah L, Patel J
Decline of Psychological Health Following the Designation of COVID-19 as a Pandemic: Descriptive Study
J Med Internet Res 2021;23(4):e24964
URL: https://www.jmir.org/2021/4/e24964
doi:10.2196/24964
PMID:33793408

Original Paper

Loss of Smell and Taste in Patients With Suspected COVID-19: Analyses of Patients' Reports on Social Media

Sachiko Koyama[1], PhD; Rumi Ueha[2], MD, PhD; Kenji Kondo[2], MD, PhD

[1]Department of Chemistry, Indiana University, Bloomington, IN, United States

[2]Department of Otolaryngology and Head and Neck Surgery, Faculty of Medicine, The University of Tokyo, Tokyo, Japan

Corresponding Author:
Sachiko Koyama, PhD
Department of Chemistry
Indiana University
800 E Kirkwood Ave
Bloomington, IN, 47405-7102
United States
Phone: 1 812 345 6155
Email: apodemusmice@gmail.com

Abstract

Background: The year 2020 was the year of the global COVID-19 pandemic. The severity of the situation has become so substantial that many or even most of the patients with mild to moderate symptoms had to self-isolate without specific medical treatments or even without being tested for COVID-19. Many patients joined internet membership groups to exchange information and support each other.

Objective: Our goal is to determine the benefits and limits of using social media to understand the symptoms of patients with suspected COVID-19 with mild to moderate symptoms and, in particular, their symptoms of anosmia (loss of the sense of smell) and ageusia (loss of the sense of taste). The voluntary reports on an internet website of a membership group will be the platform of the analyses.

Methods: Posts and comments of members of an internet group known as COVID-19 Smell and Taste Loss, founded on March 24, 2020, to support patients with suspected COVID-19 were collected and analyzed daily. Demographic data were collected using the software mechanism called Group Insights on the membership group website.

Results: Membership groups on social media have become rare sources of support for patients with suspected COVID-19 with mild to moderate symptoms. These groups provided mental support to their members and became resources for information on COVID-19 tests and medicines or supplements. However, the membership was voluntary, and often the members leave without notification. It is hard to be precise from the free voluntary reports. The number of women in the group (6995/9227, 75.38% as of October 12, 2020) was about three times more than men (2272/9227, 24.62% as of October 12, 2020), and the peak age of members was between 20-40 years in both men and women. Patients who were asymptomatic other than the senses comprised 14.93% (53/355) of the total patients. Recovery of the senses was higher in the patients who were asymptomatic besides having anosmia and ageusia. Most (112/123, 91.06%) patients experienced other symptoms first and then lost their senses, on average, 4.2 days later. Patients without other symptoms tended to recover earlier (P=.02). Patients with anosmia and ageusia occasionally reported distorted smell and taste (parosmia and dysgeusia) as well as experiencing or perceiving the smell and taste without the sources of the smell or taste (phantosmia and phantogeusia).

Conclusions: Our analysis of the social media database of suspected COVID-19 patients' voices demonstrated that, although accurate diagnosis of patients is not always obtained with social media–based analyses, it may be a useful tool to collect a large amount of data on symptoms and the clinical course of worldwide rapidly growing infectious diseases.

(J Med Internet Res 2021;23(4):e26459) doi:10.2196/26459

KEYWORDS

COVID-19; anosmia; ageusia; free reports on social media; symptomatic; asymptomatic; recovery of senses; symptom; social media; smell; taste; senses; patient-reported; benefit; limit; diagnosis

Introduction

Recent outbreaks of COVID-19, caused by SARS-CoV-2, have escalated into a worldwide pandemic. Over 29.4 million people in the United States alone have contracted the virus, causing over 532,000 deaths as of March 13, 2021.

According to the World Health Organization (WHO), "serious symptoms" are defined as "difficulties in breathing or shortness of breath, chest pain or pressure, loss of speech or movement" [1]. The Centers for Disease Control and Prevention (CDC) reported that over 80% of patients with COVID-19 have mild to moderate levels of illness [2] and 20% of the patients develop severe to critical conditions. Although there have been a number of studies published regarding the symptoms and their clinical course for COVID-19, most of them have focused on the patients with severe symptoms, and there have been few reports on patients with COVID-19 with mild to moderate conditions. Depending on the country, especially where the numbers of cases have been high and beds at hospitals are limited, most of the patients with mild to moderate symptoms of suspected COVID-19 have been self-isolating themselves with minimum access to medical treatments due to the large number of patients with severe conditions. Previously there were criteria that were required to be present (fever and coughs) to get tested, although later it became clearer that the symptoms vary largely and that made many suspected patients with various symptoms not be allowed to get tested. The patients with COVID-19 with mild to moderate symptoms are thus the most invisible population of patients who have received the least attention from the public and from medical care facilities.

Recently it has been shown that there is a high percentage (40% to as high as 96%) of asymptomatic [3-6] or presymptomatic [7] COVID-19 positive patients. In addition, more recently, there have been reports on the loss of the senses of smell (anosmia) and taste (ageusia) in COVID-19 positive patients with otherwise asymptomatic to mild levels of other symptoms [8-11] and as one of the early stage symptoms [9]. The ability to sense smell and taste showed a correlation with the severity of other symptoms as well, and odors related to chemesthesis were also found to be impaired by contraction of SARS-CoV-2 [12,13]. Studies have shown that 98% of patients with COVID-19 showed some smell dysfunction [14], indicating that smell dysfunction is a major biomarker of COVID-19. In a survey comparing COVID-19 positive and negative participants, smell dysfunction was found to be the best predictor of COVID-19 [15,16]. There are also papers reporting that asymptomatic patients with COVID-19 stay contagious longer than patients with symptoms [17]. These studies suggest that there is an urgent need to understand the asymptomatic to mild and moderate symptoms of patients with COVID-19 with anosmia and ageusia. Asymptomatic conditions sometimes mean that the patients actually have symptoms but did not realize or notice them because they are mild or because these symptoms were not included in the list of well-known symptoms like coughs and fevers. Understanding of the symptoms and the progression of COVID-19, especially in the mild cases of COVID-19, may help the subjectively asymptomatic patients notice any subtle symptoms that others have previously ignored. This includes the symptoms of anosmia and ageusia. The understanding of the onset and progression of anosmia and ageusia in otherwise asymptomatic patients and that of patients with mild or moderate symptoms may contribute to containment of the virus by enabling the detection of patients with COVID-19 at an earlier stage after contracting SARS-CoV-2.

Near the start of the pandemic, a group to support patients who lost their senses of olfaction or taste due to contraction of COVID-19, known as *COVID-19 Smell and Taste Loss*, was founded on a social media site. The posts from the members were recorded at that time to provide *precision advice*, which was based on the concept of the *Precision Medicine Initiative*. *Precision Medicine* is an approach for disease treatment that considers individual variability. In the case of a social media group, one of the ways to provide *precision advice* is to take into consideration the history of each person's previous posts on their symptoms and progress. The record of the posts at the early stage after the group was founded was conducted with the purpose of understanding the symptoms of the new disease in each person, which became the base of this study. To understand the new disease, COVID-19, we recorded all the posts on symptoms whether they were related to the senses or not.

The aim of this study was to analyze the posts and comments on this social media membership group to establish the benefits and limits of using social media to understand COVID-19 and to investigate the symptoms of mild to moderate COVID-19 patients, especially focusing on anosmia and ageusia. As earlier studies reported, these symptoms are highly observed in patients with COVID-19 [14]. They are early signs of COVID-19 [9], and in some SARS-CoV-2 infected patients, they were the only symptoms of COVID-19 [9,11]. There are only a few papers so far using the reports of patients obtained from social media [18,19]. This is one of the few reports on the mildly to moderately affected patients, many of whom did not have enough symptoms to receive COVID-19 tests and medical treatment, and therefore, clinical information is limited.

Methods

Recording of Cases

A membership group called *COVID-19 Smell and Taste Loss* was established on March 24, 2020, by the charity organization AbScent, which is based in England, United Kingdom. Polls were occasionally conducted to determine various aspects of the members, for example, hospitalization for COVID-19. The information on the age, sex, and the countries and cities where the members were located was obtained through the software mechanism called *Group Insights*. The reports posted were recorded daily as they were in Excel (Microsoft Corporation) spreadsheets with the records of the dates posted and the names, COVID-19 test results if available, a short summary of the symptoms, starting dates of anosmia and ageusia symptoms if available, and recovery dates from these symptoms if available. When the same person posted on a separate day, the posts were recorded under the name of the person with the records of the dates they were posted. Recording of the posts and comments for this study lasted from March 24 to July 1, 2020. Posts and comments without specific content related to their own

symptoms were not recorded (eg, expressing their sympathy, excitement, or approval saying "Yes," "congratulations!" "Thank you," or "nobody knows").

Institutional Review Board

The study is approved as exempt (Institutional Review Board protocol number 10082020-1, Exempt Category #3)

Case Studies

Because, in this group, members freely reported their cases, the level of detail for each member was not consistent. We classified the reports by whether the patient had other symptoms than anosmia and ageusia, and whether they recovered their senses, and then selected several reports that described their symptoms in detail.

Statistical Analysis

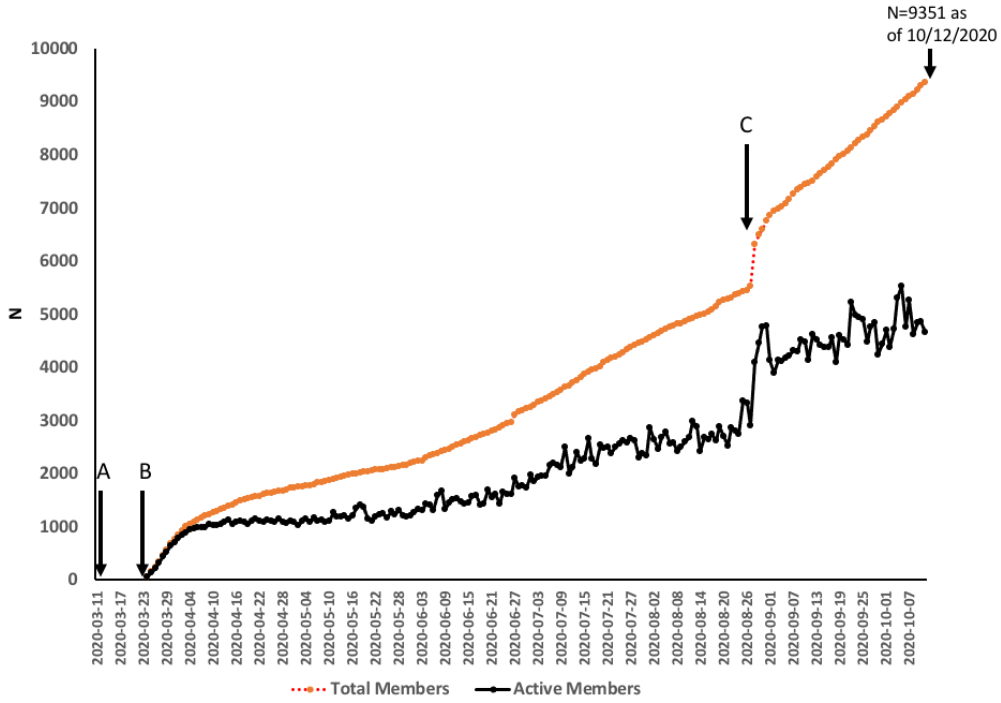
Fisher exact test was used to statistically evaluate the recovery rate among groups.

Results

Demographic Data

Figure 1 shows the time course since the pandemic started. On March 11, 2020, the WHO declared the rapidly spreading SARS-CoV-2 virus a pandemic and warned that it might spread to many countries around the world (Figure 1A). The disease caused by the virus, COVID-19, was first considered to cause fever and coughs as the main symptoms. However, soon it was discovered that there were more symptoms, including the loss of the senses of smell (anosmia) and taste (ageusia). The membership group on social media studied in this project was founded on March 24 (Figure 1B), and the number of members reached about 1000 during the first week (Figure 1). This immediate increase of the members indicates the rapid spread of the disease and the large number of people who suddenly lost their senses of smell and taste at the early stage of the pandemic. The increase in the number of members showed another surge around the time when the second wave in the number of cases started in the autumn of 2020. This may be due to both the increased publicity after the introduction of the group in the media (Figure 1C) and the large number of people with COVID-19–induced anosmia and ageusia. This suggests the existence of a large number of patients who lost their senses during this pandemic.

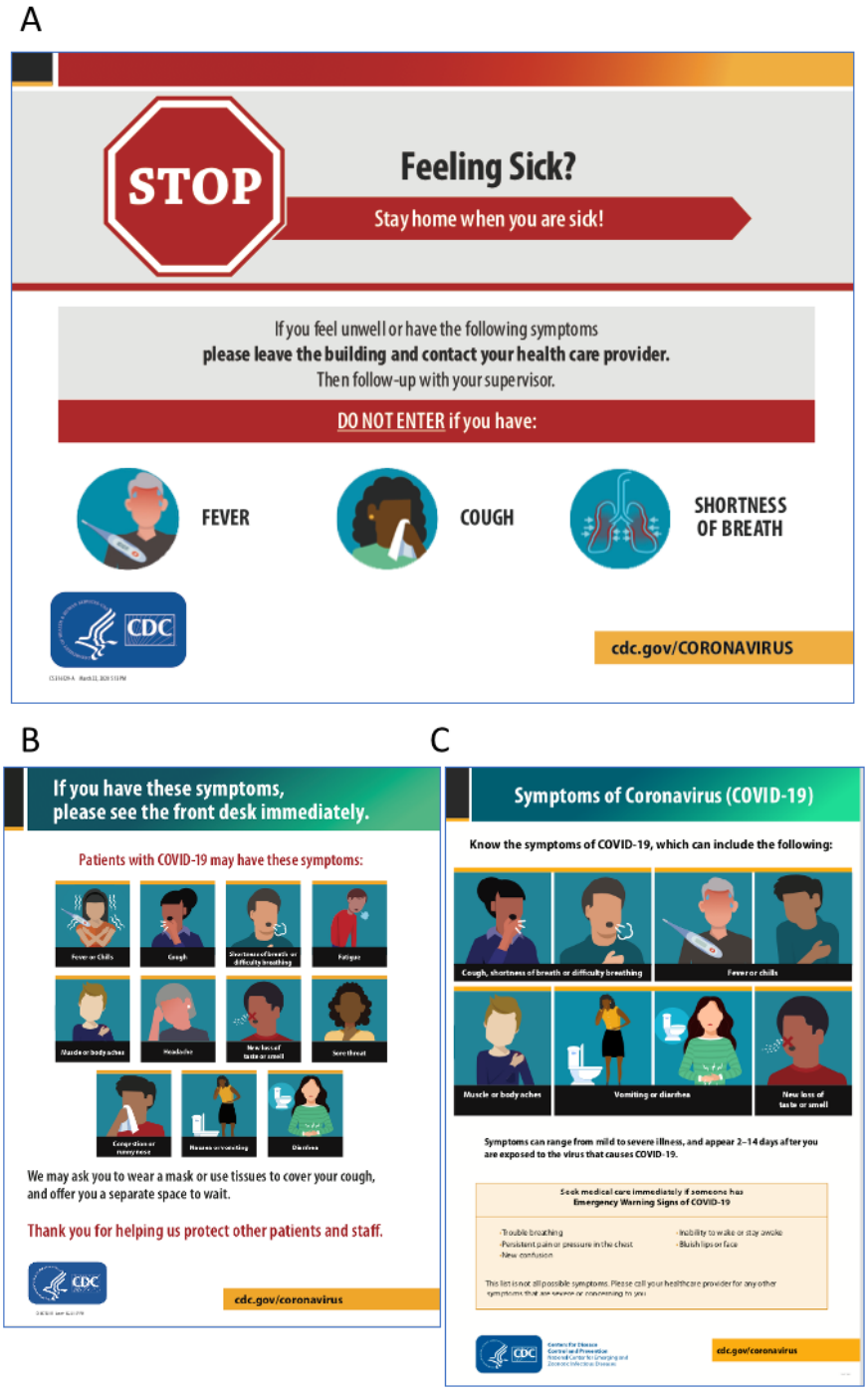
Figure 1. Increase in the number of members (red) and active members who make posts or comments to other posts (black). A: Date of the World Health Organization announcement on the pandemic; B: COVID-19 Smell and Taste Loss membership group founded; C: increase of over 1000 members within a few days following the interview of the founder of the group on the BBC.



What were considered the symptoms of COVID-19 have changed from the earlier stage of the pandemic in spring 2020 to summer and fall 2020. Figure 2 shows the outdoor and indoor signs made by the CDC in March, June, and July [20]. In the sign generated in March (Figure 2A), only the symptoms of fever, coughs, and shortness of breath are described, whereas in the June and July signs, the variety of symptoms has increased by more than double, and now it is well known that not every COVID-19 positive patient has fever, coughs, or shortness of breath (Figure 2B and C). In the poll asking if they got the COVID-19 tests immediately after the onset of symptoms, about 70% (n=259) of the 705 people answered that they could not get tested at all or could not get tested immediately and received the antibody tests later (Figure 3A; could not get tested: n=259, 36.74%; antibody test later: n=247, 35.04%). Analyses of the people who added comments (n=54) to the poll revealed that

most of the patients with COVID-19 who could get tested immediately after the onset of symptoms was due to their jobs (essential workers, mostly) or those who contracted the virus after June when the criteria to get tested expanded (compare Figure 2A and 2B; Figure 3B bottom; job: n=12, 22.22%; after June: n=26, 48.15%). The comments by the people who replied that they could not get tested immediately and received antibody tests later or could not get any test revealed that they were mostly the people who had an onset of symptoms before June (Figure 3B, top two bars).

Figure 2. Indoor and outdoor signs released by the CDC as free to be printed and used. CDC: Centers for Disease Control and Prevention.



Thus during the early stage of the pandemic, there were many people who got sick and self-diagnosed themselves with COVID-19 (suspected COVID-19 patient). This inability to get a COVID-19 test immediately after the onset of symptoms also resulted in the inability to get immediate medical treatments because medicines were not prescribed without an official diagnosis. We asked a question about the way they spent time after the onset of symptoms and found that about 90% (n=204,) of the 221 members who replied either stayed at home with generic medicines and supplements (n=162, 73.73%) or stayed at home without any medicines (n=40, 18.10%; Figure 4). Only 6.79% (n=15) replied that they stayed at home with medicines prescribed by doctors. Although there were some responses showing that there were members who experienced hospitalization in a separate poll (see Figure 3), none of them responded to the poll on the medical treatments, and thus, the

option of medical treatment following hospitalization is not shown in Figure 4. Although there is some missing information on the hospitalized people in the results, Figure 4 clearly shows the invisible nature of the patients with mild to moderate symptoms, which also suggests that the number of cases worldwide in the countries where availability of COVID-19 tests was limited or had criteria to get tested is highly likely to be far more than the officially reported numbers.

Figure 3. Summary of the reply to the question "were you able to get tested immediately after the onset of the symptoms of COVID-19" (A) and classification of the comments related to their selection. (B) Most often listed reasons that they could or could not get tested were the time of onset (before June or after June when the Centers for Disease Control and Prevention announced changes in the typical symptoms and included more symptoms), their jobs as essential workers or supervisors' arrangements, pre-existing health conditions, and hospitalization because of symptoms.

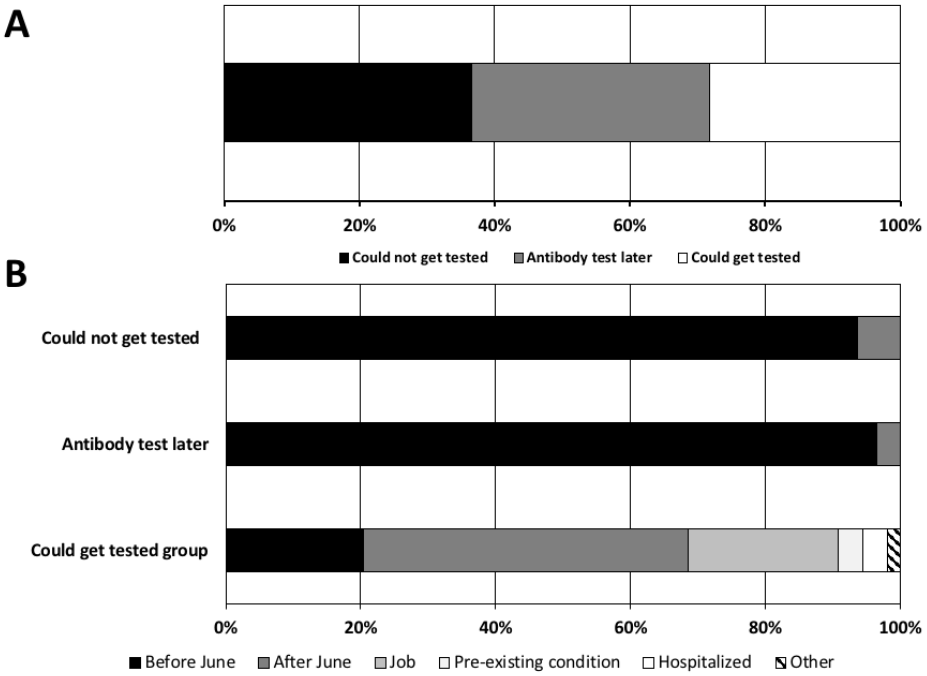
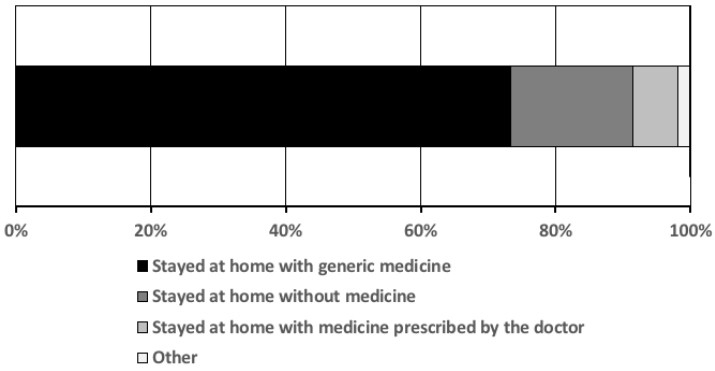
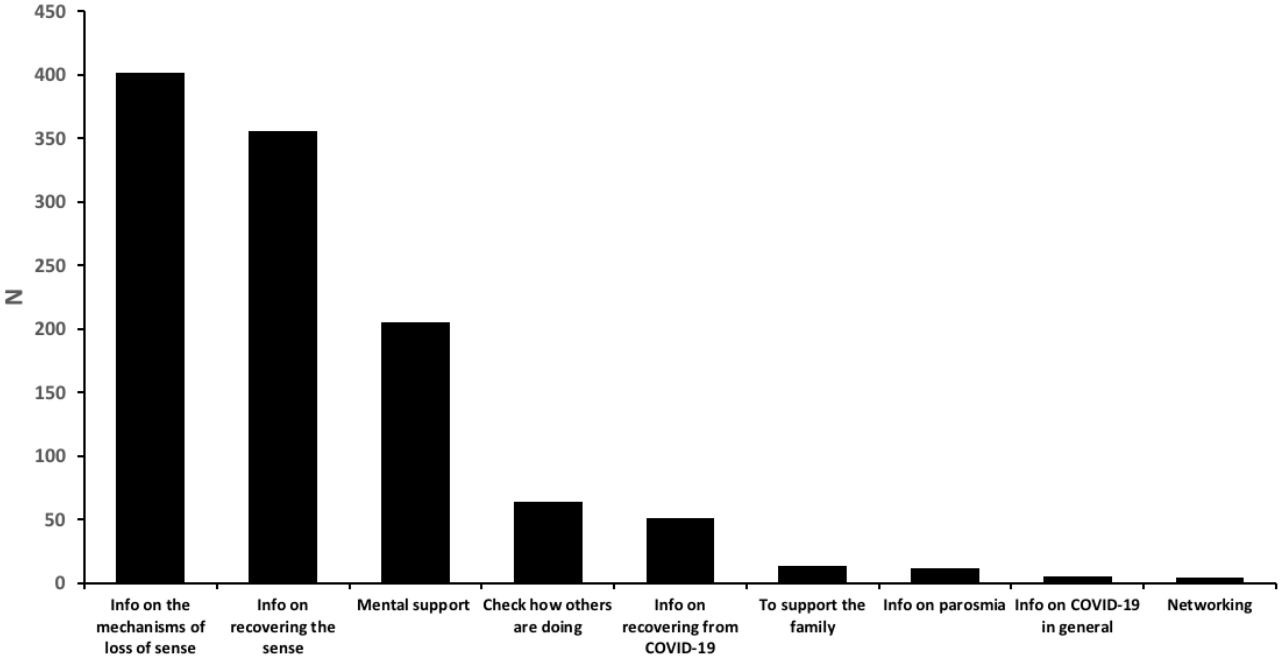


Figure 4. Availability of medical treatments. The result of another choice "hospitalized and then stayed at home with medicines prescribed by the doctors" is not shown as no one selected the option.



The self-isolation situation, staying at home, itself can cause anxiety. With symptoms of suspected COVID-19 and without a COVID-19 test and medical treatment, it is not hard to imagine that this situation causes high anxiety. It is possible during self isolation to obtain information from the internet and, if available, to join a patient group to exchange information. We asked the members the reasons that they joined the group. We found that the top three reasons were *to obtain the information about the mechanisms of sense loss* (n=401), *information on how to recover the senses* (n=355), and *mental support* (n=205), which accounted for 86.65% (n=961) of the 1109 responses (Figure 5). The sum of the responses related to information (5 out of 9 options; n=823, 74.21%) and mental support of themselves or their family (2 out of 9 options; n=218, 19.66%) accounted for over 93% of the reasons to join the group. If we consider that the two other options, *networking* and *checking how others are doing*, are relatively similar to mental support, it is possible to say that obtaining information and mental support are the two major reasons for joining the group.

Figure 5. Reasons to join the membership group. Multiple selections allowed and adding options was allowed.



Demographic data for the registered members (n=9227 as of October 12, 2020) of the social media site is shown in Figure 6. The number of women in the group (n=6995, 75.38%) was about three times more than men in the group (n=2272, 24.62%). This number included the members who were administrators of the group or scientists, and not patients, although the estimated percentage of these nonpatient members in the group was minimal. There was a peak in the number of members in the age range between 25-34 years in both men and women (Figure 6). This high number in the rather young ages could be due to the use of social media, although there are multiple other reports on patients with COVID-19 with anosmia and ageusia, which are not using social media, that show the same tendency [6]. On the other hand, in other membership groups of AbScent on social media, the average age is higher, which also suggests that this is specific to COVID-19 and not due to the use of social media. We do not know yet if this is due to the tendency that older-aged people develop severe conditions when they contract COVID-19 and are hospitalized or reach a severe condition such that they are unable to use the internet, or if the symptoms of the senses are specific to patients of younger ages.

Figure 6. Number of members of the COVID-19 Smell and Taste Loss group sorted by age and sex.

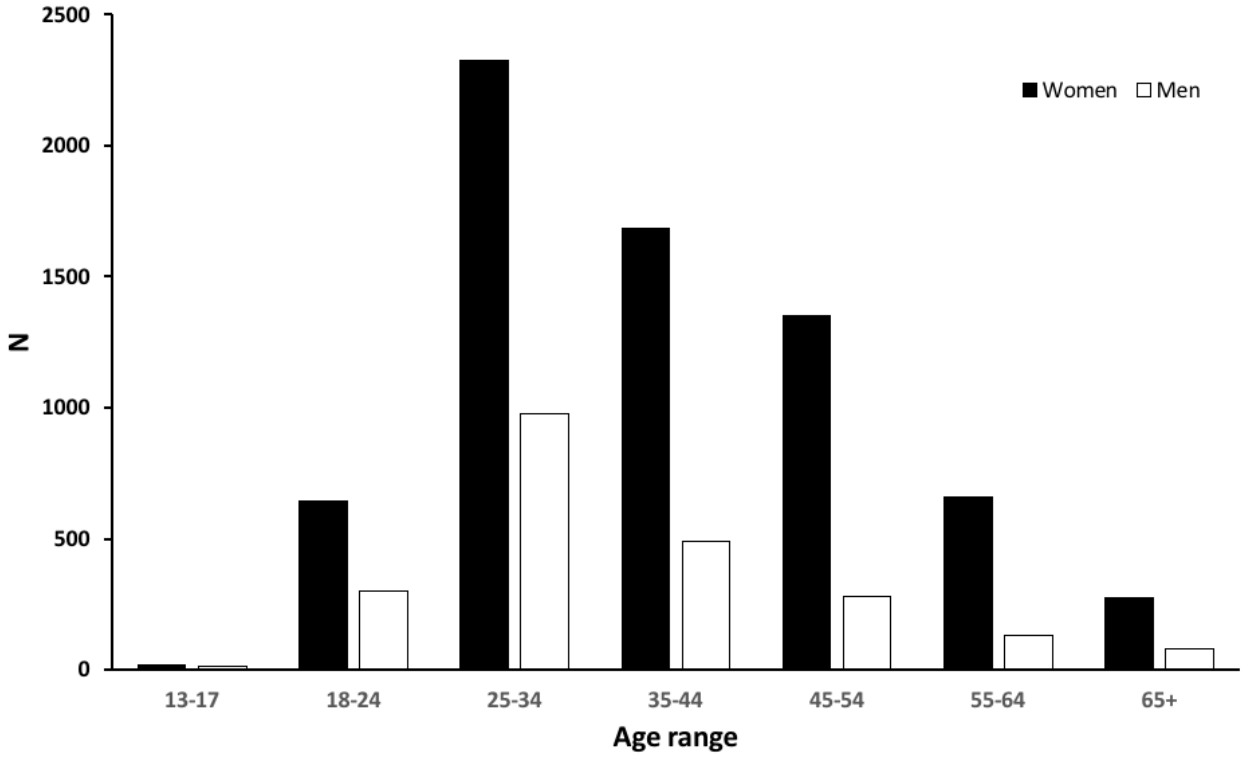
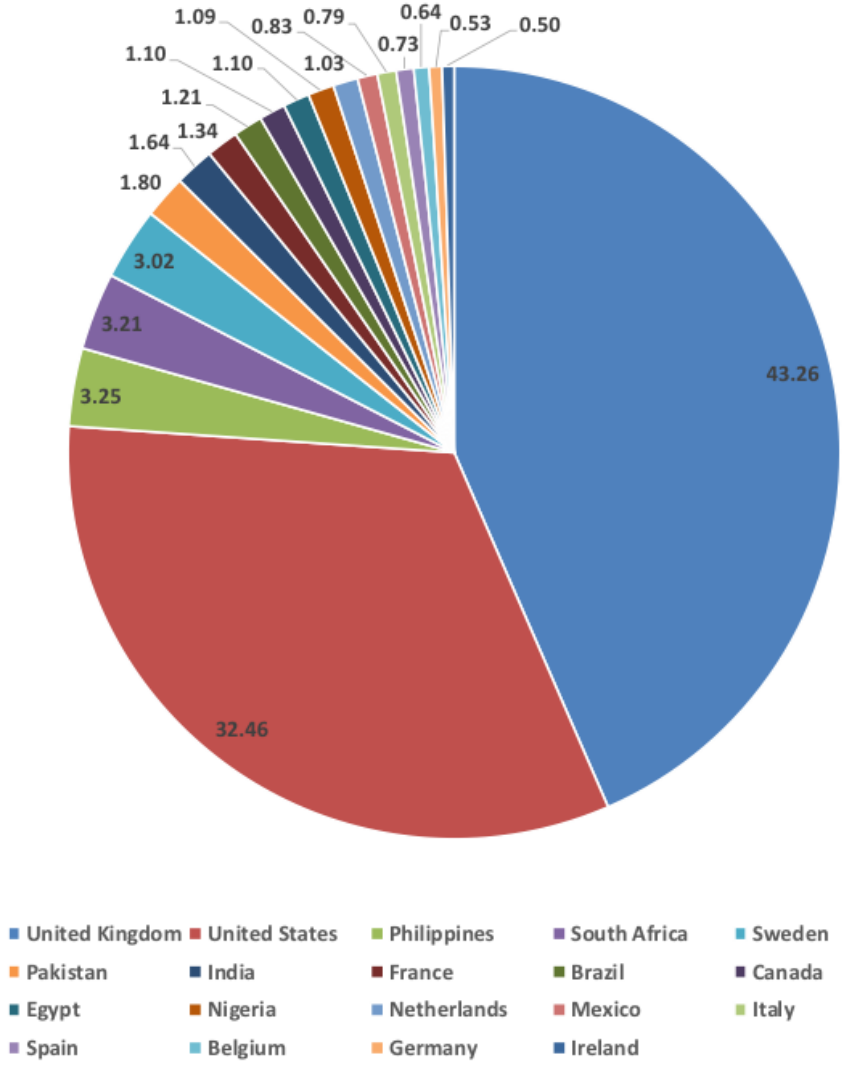


Figure 7 shows the top 20 countries of the members as of October 12, 2020. Where the patients were located may have had some influence on the symptoms, considering the mutation of the reported virus [21,22]. There were a total of 99 countries that had members in the group. The fact that it was possible to join the group from anywhere the internet was available, if there was no federal restriction on joining a social media group, was one of the benefits for suspected patients staying at home without medical treatments. The top 2 countries, however, were the United Kingdom and the United States, comprising 75.7% (United Kingdom: n=3744; United States: n=2809; total of the 20 countries: n=8654 members) of the number of members of the top 20 countries. This suggests that it is easy for the people in the countries where English is used to join the social media group. This may explain the names of the third and fourth top countries as well. The Philippines and South Africa are the third and fourth countries, respectively, in the number of members, and English is one of the official languages in these countries. There were rather few members from the countries in Europe even though the pandemic was severe there (eg, Italy: n=68, 0.79%), which could be due to the language barrier.

Figure 7. Top 20 countries with a large number of members. Total number of members in these 20 countries was 8654 as of October 12, 2020. The numbers shown inside or outside of the figure indicate the percentage of members of each country.



Symptoms

In December 2019, when the first reports on the outbreak started in Wuhan, China, the major symptoms reported were shortness of breath, cough, fever, and diarrhea. However, it has become clear that many patients who contracted the virus are asymptomatic yet contagious [3-5]. In addition, the symptoms of the symptomatic patients are more varied than first reported (Figure 2). Many patients reported symptoms that suggested encephalitis (headaches, foggy brain, memory loss, hallucinations, pain in the upper end of the nose suggesting inflammation in the olfactory bulb, ear pain, and pain in the back of eyes; Table 1). In addition, some patients also reported a skin rash and tingling in the legs (Table 1), which also has been reported by other recent publications [23].

Table 1. Symptoms other than anosmia or ageusia reported by COVID-19–induced anosmia or ageusia patients.

Region and symptoms	Patients, n
Head	
Headache, migraine, pressure in head	143
Dizziness, light head, vertigo	25
Depression, anxiety	22
Foggy brain	8
Insomnia	7
Forgetful	2
Hallucination	1
Disoriented	1
Nightmare	1
Nose	
Congested sinus, pressure in sinus, stuffy nose	67
Burning nose, nose pain, inflammation in nose, nasal pressure	51
Running nose	21
Sneeze	11
Bloody nose	8
Dry nose	4
Dried glue mucus	2
Respiratory system	
Cough	91
Sore throat, burning sensation in throat	47
Chest pain, chest tightness, burning lungs	43
Short breath, difficulties in breathing	31
Congested throat	4
Wheezing, harsh voice	3
Throat congestion	1
Eyes	
Eye pain, sore eyes	18
Dry eye	3
Light sensitivity	2
Pink eye, red eye	2
Watery eye	2
Vision affected, difficulties in distance vision	1
Mouth	
Dry mouth, sore mouth	6
Tooth pain, tooth sensitivity	3
Tingling in tongue, sensation in tongue	2
Film on tongue, thrush	2
Dry lips	1
Burning mouth	1
Ear	
Ear pain, ringing ears, tinnitus	20

Region and symptoms	Patients, n
Muffled ear, ear congestion, ear infection	4
Deafness in ear	1
Gastrointestinal system	
Nausea, upset stomach	17
Diarrhea, off stomach	14
No appetite	9
Stomachache	1
Other	
Fatigue, drained, winded	117
Fever, feverish	81
Body ache include all parts other than in the head area, muscle ache	74
Chills, shiver, chills and sweating	19
Rash on body	8
Malaise	6
Pins and needles in leg	6
Sensitive skin, sore skin pain	3
Burning up	2
Numb at toe, numb at hands	2
Face hurt, pressure in face	4
Facial palsy	2
Skin wrinkle	1
Dry skin	1
High blood pressure	1

During the time period that the posts from the members were recorded (March to the end of June 2020), there were 355 members who posted reports of their smell or taste dysfunction and other symptoms they had. Of these 355 members, the percentage of patients without other symptoms comprised 14.93% (n=53), and 85.07% (n=302) of them had other symptoms than the loss of their senses. Almost 10% had recovered their senses by the time this report was summarized (recovered vs not recovered: n=34, 9.58% and n=321, 90.42%). The percentage of patients who recovered their senses was higher in the patients who did not have other symptoms besides anosmia and ageusia (recovered with other symptoms vs without other symptoms: 25/302, 8.28% and 9/53, 16.98%).

There were 123 members who reported that they had other symptoms and reported the timing of the onset of the loss of senses and other symptoms. Of the 123 members, 112 (91.06%) members reported that other symptoms started first. The average difference of the days between the onset of other symptoms and the onset of the loss of senses was 4.2 days. There were 6 (4.9%) members who reported that the onset of the loss of senses occurred on the day other symptoms started, and 5 (4.07%) members who reported that the loss of senses occurred before other symptoms started. The average difference of the days between the onset of the loss of senses and other symptoms was 1.6 days when the loss of senses preceded the onset of other symptoms.

The duration of time until recovery varied largely, with some people regaining their senses within a week and some people who lost their senses at the beginning of the pandemic and still had not recovered their senses at the end of June 2020. As we found the tendency that patients without other symptoms had a rather higher recovery rate, we examined if these differences correlated with the duration to recover their senses as well. Table 2 shows the 34 patients who recovered their senses in the order of the days until recovery. When the 34 patients were divided into the top 50% and the bottom 50% (fast recovery group and slow recovery group; n=17 for both), the average days to recover was 12.9 days for the fast recovery group and 41.3 days for the slow recovery group. In the fast recovery group (n=17), there were 8 (61.54%) patients without other symptoms, whereas in the slow recovery group (n=17), there was only 1 (7.69%) patient who did not have other symptoms, and there was a statistically significant difference in the duration to recovery (Fisher exact test was used: P=.02; Figure 8). These results indicate that patients without other symptoms than the loss of the senses had the tendency to recover earlier. This, however, did not indicate that patients without other symptoms will always recover their senses faster. Of the 321 patients who reported they had not recovered their senses by the time we

summarized this report, there were 44 (13.71%) patients who reported they did not have other symptoms. This suggests that there may be several causations in the loss of senses (ie, one of these could be the severity of the overall COVID-19–related symptoms and some other factors could also be involved, affecting the recovery process).

Patients with anosmia and ageusia occasionally reported distorted smell and taste (parosmia and dysgeusia) as well as the smell and taste without the sources of the smell or taste (phantosmia and phantogeusia). Table 3 summarizes the reports from the patients. Although there are various ways to describe the nature of phantosmia and phantogeusia, a smell and taste of smoke and burnt material was most frequently reported.

Table 2. Days until the recovery of senses sorted by the days and the symptoms.

Order	Until recovery of senses (days)	Symptoms other than anosmia and ageusia
1	5	Low grade fever, fatigue, insomnia
2	9	No other symptoms
3	9	No other symptoms
4	11	Fatigue, chest pain, short breath, no fever, no cough
5	12	No other symptoms
6	13	No other symptoms
7	13	No other symptoms
8	14	Cough, nausea
9	14	No other symptoms
10	14	Slight fever, runny nose
11	14	Fever, aches, chills, sore throat, headaches, rash
12	14	No other symptoms
13	14	No other symptoms
14	15	Nose pain, fatigue, headaches
15	15	Fatigue
16	17	Cold, runny nose, fatigue, sore eyes, congestion
17	17	Fatigue, cough
18	21	Burning nose, fatigue, eye pain, shiver
19	21	Fatigue, sore throat
20	23	Fever
21	27	Dizziness, disorientation, foggy brain, insomnia, ear congestion
22	28	No other symptoms
23	30	Back pain, sore throat
24	31	Burning nose, headache, fatigue, sinus infection
25	32	Nose is super dry and inflammation is high
26	35	Backache, low fever, very mild nasal congestion and mild difficulty breathing.
27	38	No fever, chills, coughs, headache, muscle ache, fatigue
28	42	Fatigue, feverish, cough
29	42	unwell, sore throat, breathless, headaches, fatigue, nausea, tinnitus
30	47	Fever, headache, fatigue, sore skin
31	56	Cough, fever, muscle ache, lethargy
32	69	Fatigue, headaches, fizzy, burning nose
33	77	Fatigue, headaches
34	83	Fatigue, headache, nasal pain

Figure 8. Recovery of senses. Earliest 50%: patients who lost their senses and recovered; longer 50%: patients who lost their senses and took a longer time to recover.

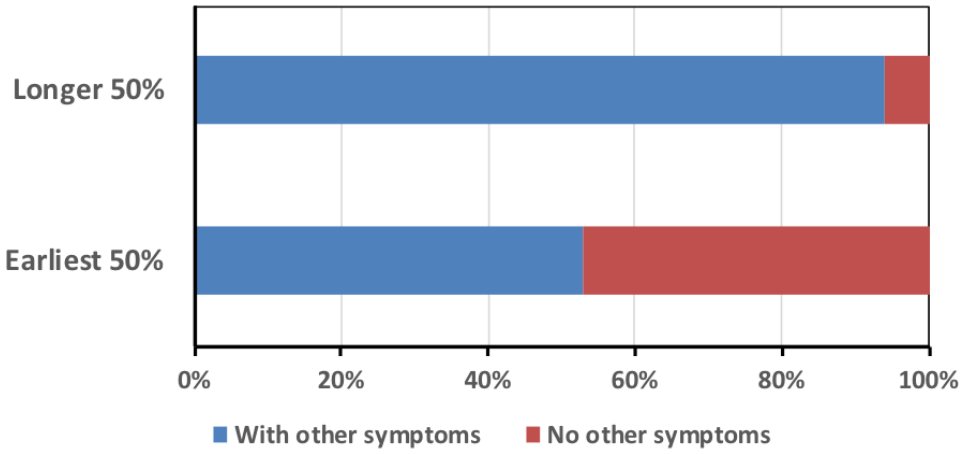


Table 3. Types of phantosmia and phantogeusia.

Types of phantosmia and parosmia	Reported cases, n
Smoke, ashtray, cigarette, burnt, fire, dusty	43
Chemical	7
Ammonia, vinegar	6
Metallic	5
Garbage, rotten	3
Skunk	1
Other expressions: weird, strange, distorted	6

Case Studies

Here, we show several examples of reports from the patients. In selecting the examples, we first classified the 355 members who posted reports of their smell or taste dysfunction during March and June 2020 into the patients who recovered (fast recovery group and slow recovery group) and had not recovered, and those with or without other symptoms than anosmia and ageusia.

Case 1 represents a fast recovery group patient who recovered early and did not have other symptoms. The length of time until recovery of smell and taste was 9 days. This patient was not tested for COVID-19. The dates shown in Textbox 1 were added when they were recorded in Excel.

Textbox 1. Case 1: fast recovery group without other symptoms (original posts).

April 7, 2020 (Tuesday)

- "Woke up and Lost all sense of smell on Thursday. Can't taste can only taste salt or sweet. Smell seems to be coming back it's now at about 10% maybe lower because I can only smell garlic and onion powder. Can't smell perfume at all. I am day 6. I am praying it's returning. Brother in law lost his smell for 11 days and it's returned fully"

April 8, 2020 (Wednesday)

- "I woke this morning and I can smell my toast in the toaster and also zoflora cleaning product can also taste the toast as well I would say it's come back 30% so happy it's improving daily thank god."
- "Quick update lost all sense of smell last Thursday complete loss no taste or smell. Yesterday i could smell spices on close inspection. This morning I could smell my toast in the toaster and also taste it it's still not 100% but I would say it's now at about 30% smell 50% taste it is getting better each day I'm now on day 7 since losing it so the general time frames are generally 7-14 days but if yours isn't then don't panic it will come back in time."
- "It came back together once my smell returned yesterday taste returned but can taste more than I can smell it's 40% smell 60% taste which am hoping will improve daily I'm On day 7 too."

April 9, 2020 (Thursday)

- "Hi all update to help anyone currently feeling scared and alone try and not to panic it will come back mine is now back 75% I'm day 7 I cried tonight when I ate dinner because I could taste and smell it very scary when all sense goes but the joy when it returns we often take this sense for granted please don't worry it will return"

April 11, 2020 (Saturday)

- "100% back"

Here is another example from the fast recovery group, but the patient had other symptoms. The patient reported recovery of the senses in 13 days (however, at the time point of summarizing this report, we have heard that this patient started to experience parosmia starting from about 2 months after the recovery of the senses). The dates shown in Textbox 2 were added when they were recorded in Excel.

Textbox 2. Case 2: fast recovery group with other symptoms (original posts).

March 28, 2020 (Saturday)

- "I've noticed this other weird phenomenon and wanted to see if you all were experiencing it too. Since I got my first possible CV19 symptom 13 days ago (slight cough), whenever I shower my skin on my hands wrinkle up super fast...like within 5 minutes of being in the shower. It usually take a good 30 minutes for that to happen to me. Now I barely shampoo and my hands are all wrinkled. My wife noticed it happening to her too although she has had no symptoms. Anyone else notice this weird thing? Ps...my smell is coming back! It's about 40-50% today and I'm on day 7. I hope it continues!!!"

March 29, 2020 (Sunday)

- "I'm on day 8 and yesterday was the weirdest day. I had the strange sense of a smell whenever I inhaled, like you, but mine was like a burning sour smell. And I got a slight metallic taste when I was eating food. Not going to lie, it was pretty awful. In fact, at the end of the day I started feeling sick to my stomach with the feeling that I had been smelling sour milk for hours and hours. Luckily today, that's completely gone. Now I'm back to smelling things about 50%. I think it's called Phantosmie. I don't know, I'm going to do more research today."

April 3, 2020 (Friday)

- "I just wanted to write a positive post for those who are feeling anxious and scared. I've got my smell and taste back! It is GLORIOUS. Everything is so flavorful and the smells are incredible. I lost both senses 13 days ago and had absolutely no improvement for 5-6 days. I was so scared that it might never come back and it was very depressing. On day 5 or 6 I could just get a wiff of a few things that I passed by maybe at about 5%. It stayed there for a couple days and then slowly started to get better. I personally had a lot of ups and downs where one day it might be at 60% and then the next day back at 30%. I was feeling rather frustrated because I couldn't tell if I was really getting better or not. Well I'm happy to say that the last two days I've been able to taste everything I've eaten and smelled things I haven't smelled in almost two weeks!! I know it might drop again before being at 100% consistently but I feel hope that I will fully heal for the long haul. Those of you who are where I was a week or so ago...don't give up. Try to stay hopeful even though it's hard. I want this to encourage anyone who feels nervous right now. It's a possibly that things will be ok.

  "PS...I'm pretty sure I had CV19 but couldn't get tested because I got don't fit the criteria."
- "I had a dry cough and then a full head cold and I lost my senses once my cold started to get better. I did have one day of burning nose and chemical taste right before I started really getting better with my smell and taste."

The next example is a patient of the slow recovery group who recovered after 42 days and had other symptoms. The dates shown in Textbox 3 were added when they were recorded in Excel.

Textbox 3. Case 3: slow recovery group with other symptoms (original posts).

April 8, 2020 (Wednesday)

- "I started getting covid symptom April 1st and on day 4 my sense of smell and taste suddenly 100% dissapeared. Now this was only 4 days ago but it already feels like a lifetime...I can't even remember what it's like to taste food. I'm just wondering, I've seen that this could be caused by a zinc deficiency? Coincidently I had just started taking zinc when the covid symptoms started so I was taking it, but now that I've actually started researching it apperently a certain form of zinc (I believe it was a nasal spray) had caused permenant loss of smell for many people so now I'm kindof hesitant to even take the pills...So confusing!!!"

May 16, 2020 (Saturday)

- "just wanted to share my story before I leave this group to provide hope to people. I got COVID symptoms April 1st (weakness, fatigue, feverish, dry cough) and woke up April 4th with zero taste or smell. I was never actually tested. I had 5 days with absolutely nothing (which was absolutely horrible, I can't even imagine people who have to deal with that for more than a month it was literally all I thought about) then I was finally able to get a tiny whiff of one of my essential oils (lemon grass haha) which gave me hope that it was not permenant. It was soooo slow coming back, like I would say it only improved max 2% a day. It is only yesterday that I realized I think I am finally at 100% back to normal, so I would estimate it took a month to slowly come back. Throughout that month I would wake up almost every day with that phantom "chemically plasticy burning" smell/sensation which gradually reduced in strength, I do remember smelling it only a few days ago. Now I guess my recovery is sort of on the fast side, which I would attribute to my healthy lifestyle. I generally eat a plant focused diet, with minimal meat, dairy and processed foods. Since I was sick I have had barely any alcohol and obviously no smoking. In the beginning I was taking vitamins daily such as vit C, D, zinc and oil of oregano. I also tried to incorporate many anti-inflammatory foods into my diet such as tumeric, garlic, ginger. Who knows how much this actually helped but I'm sure it didn't hinder."

The last example is of a patient who has not recovered their senses yet. The dates shown in Textbox 4 were added when they were recorded in Excel.

Textbox 4. Case 4: not recovered, with other symptoms.

March 29, 2020 (Sunday)

- "March 13th - sudden and complete loss smell and taste. Before that date I have been suffering some headaches. March 16 to 18 – fatigue and dizziness. Now I feel ok physically but smell and taste is still gone. After 12-13 days later I was able to smell coffee grains but faintly. My testers are coffee, vicks, perfume for sniffing. I'm about 10% and it comes and goes back. Taste was completely gone I could not even identify salty, spicy etc. Now I can identify them but flavor is missing. I feel 15-20% improvement but not so sure."

April 2, 2020 (Thursday)

- "I could smell Vicks around day 15-16. Now in day 20. Coffee, garlic, mint etc. are OK but rest is not good"

April 14, 2020 (Tuesday)

- "Day 32 still weak. It comes and goes even during the day. Early mornings and evenings are worse. Still need to get close to be able to smell and taste is way behind."

April 24, 2020 (Friday)

- "Hi all! It has been around 10 weeks for me and I cannot even say the percentage :) How do you define it? For example, I feel 100% for mint, coffee or other strong things but 0% for basil interestingly - once upon a time it was my fav :D Besides, all bad smells are totally gone. Sometimes I can smell strong smells in the air - if there is garlic in the food for example. And as many others I'm experiencing fluctuations hour by hour and day by day. Still feel improving but sometimes I feel quite desperate and it's really diffucult to stay always postive! Good luck to all of us!"

Discussion

Benefits and Limits of Social Media

The year 2020 has become a tragic year with the pandemic that has over 67 million cases and over 1.54 million deaths worldwide as of December 6, 2020. At the early stage of the pandemic, many suspected COVID-19 patients could not get tested because of the overwhelming number of patients with severe conditions at the hospitals. This suggests that the actual number of patients with COVID-19 could be far more than the officially reported number. Many of the suspected COVID-19 patients with mild to moderate symptoms self-isolated themselves without medical treatment. These suspected COVID-19 patients had to depend on the internet to obtain information, and many of them joined membership groups on social media, which were founded to support patients with COVID-19.

In this study, we analyzed the information in the social media database of suspected COVID-19 patients, with particular attention to their olfactory and gustatory dysfunction. About 70% of them could not get tested at all or could not get tested immediately and received the antibody tests later. Furthermore,

about 90% of them either stayed at home with generic medicines or supplements, or stayed at home without any medicines. These results suggest the limited access of the patients with mild to moderate symptoms to medical services. In the aspect of olfactory and gustatory dysfunction, it was demonstrated that patients who were asymptomatic other than the senses comprised about 15% of patients. Recovery of the senses was higher in the patients who were asymptomatic besides having anosmia and ageusia. Most patients experienced other symptoms first and then lost their senses, on average, 4.2 days later. Patients without other symptoms tended to recover earlier (P=.02). Patients with anosmia and ageusia occasionally reported distorted smell and taste (parosmia and dysgeusia) and experiencing or perceiving the smell and taste without the sources of the smell or taste (phantosmia and phantogeusia). Olfactory dysfunction of a large portion of patients with COVID-19 persisted at least several weeks to months.

Through the analyses of the reports, we have noticed the benefits and limits of using social media (Textbox 5). There are many benefits for patients, which suggest that these membership groups can be a beneficial system for patients to provide information and mental support. For researchers, there are some limits that require improvements to use social media membership groups to enhance understanding of the COVID-19 symptoms and conditions of the patients. These limits are due to the free nature of the involvement in the activities of the group, which causes difficulties in following up with each patient, and control of the conditions. The reports are based on each patient's subjective experience that the same symptom at similar severity could be experienced as a difficult experience by one person and as an easy experience by another. Some modifications in the way of organizing these membership groups may be able to reduce these limits for the researchers in the future. Demographic data revealed that there were about three times more women than men, and the range of age showed a peak in the age group of 25-34 years. The number of members in each age group decreased in the groups of higher age. These results on the demographic trends may reflect the trends that the symptoms of the younger people are milder compared to older people, and thus, they stayed home without medical treatments. However, there are also possibilities that women used social media more than men and that younger people used social media more than older people to obtain information on COVID-19. There were, in fact, a few cases that wives who joined the group instead of the husbands with suspected COVID-19 and mothers who joined the group instead of their small children who were suspected COVID-19 patients. These are the possible aspects of the social site structures that can skew the information.

Textbox 5. Benefits and limits of using social media.

Benefits

- (Patient) Possible to post *immediately* without waiting, compared to doctor visits, which usually require appointments
- (Patient) Patients who have mild/moderate symptoms can obtain and exchange *information*
- (Patient) Patients who have mild/moderate level of symptoms can *mentally support* each other
- (Patient) Possible to join from anywhere in the world if there is access to internet
- (Researcher) If the number of members increase, it is possible to obtain large numbers of data
- (Researcher) Possible to understand the patients' conditions, emotional responses, their experiences with their doctors from their posts

Limits

- (Patient) Possibility to join is limited to the areas where there is access to internet and to countries without federal restrictions to join internet groups
- (Researcher) There is no control group
- (Researcher) Some people are tested for COVID-19 and some are not, and we do not know which person is tested by the reports unless the person writes this each time
- (Researcher) Cannot get the same people to respond to each poll
- (Researcher) Everything is based on self-reporting and it is difficult to get precise information on the symptoms to diagnose
- (Researcher) Members often leave without notification and it is difficult to follow up

Symptoms

Among the symptoms, we especially focused on smell and taste dysfunction, which is now known as one of the major biomarkers of COVID-19. The use of social media allowed members to join the group from around the world. Although there is a limit in the range of information that can be obtained through social media (eg, accurate diagnosis of the patients), the use of it enabled us to obtain the symptoms and the progression of these symptoms from the patients who did not or could not visit the hospitals.

The current analysis demonstrated that about 90% of the 355 patients in the *COVID-19 Smell and Taste Loss* group that reported their subjective smell or taste disturbance did not recover, at least completely, for months. These results are in sharp contrast to the previously reported findings that 80% of smell and taste dysfunction in patients with COVID-19 is recovered within a few weeks [9,24-26]. This discrepancy may

be due to several reasons: the differences in the patient populations (ie, the patients in the previous studies were mostly in-patients whereas over 99% of the patients of this report had never been hospitalized and most of them did not or could not have access to medical treatments at a hospital), this is a self-selected group who joined because of their smell loss, and one limit of social media, which is the inability to confirm the recovery unless the patients post their recovery to the group.

The pathophysiology of human olfactory dysfunction is divided into three major categories. One is an airflow problem, where the airflow toward the olfactory cleft is blocked by mucosal swelling or mucus hypersecretion [27-29]. The second is a sensorineural problem, where the perception of odorants is disturbed due to damage to the olfactory mucosa, such as olfactory neural degeneration, downregulation of the odorant molecules on the olfactory neurons, or impaired transport of the odorant molecules to the olfactory receptors due to secretory dysfunction of Bowman glands [27-29]. The third is a central nervous problem, in which the olfactory pathway from the olfactory bulb to the olfactory cortex is damaged [27-29]. Airflow problems could be resolved in a relatively short time as mucosal swelling by viral infection decreases in several days to weeks, and usually the smell problem could fluctuate in its severity during the time course, depending on the nasal condition. In contrast, sensorineural and central nervous problems are usually severe, do not fluctuate, and take a longer time (months to years) for recovery. Postviral olfactory disorder, an olfactory problem developing after an upper respiratory viral infection, is usually considered as a combination of a sensorineural and central nervous system problem [27-29]. In contrast to such current concepts, previous studies reported that a large portion of patients with COVID-19 recovered from their olfactory dysfunction within a few weeks [9,24-26], suggesting that the problem was due to an airflow problem. In fact, a case report of olfactory dysfunction in a patient with COVID-19 showed mucosal swelling in a restricted area of the olfactory cleft on magnetic resonance images [30]. However, our analysis, as well as some more recent studies [31], demonstrated that a larger number of those with olfactory dysfunction were not resolved in such short time frames, suggesting that a neural problem could be also involved in COVID-19–induced olfactory dysfunction. Previous reports suggest that the virus may migrate along the peripheral and central nervous system. Studies using animal models have shown that human coronaviruses can make axonal transports enabling neuron-to-neuron propagation [32]. Development of parosmia, a sign of neural olfactory dysfunction, supports this hypothesis.

Our analysis also demonstrated that the smell problem of patients with COVID-19 tends to recover sooner when the patients only have smell dysfunction, not multiple COVID-19 symptoms. The reason for this finding is unclear, but one possible explanation is that, for the outpatients without access to medical treatments, recovery from olfactory dysfunction is faster if the patients have a lower viral load and therefore fewer symptoms. In other words, olfactory dysfunction may be one of the most sensitive indicators of SARS-CoV-2 infection. In fact, the percentage of patients showing anosmia or ageusia was especially high in the asymptomatic to mild level patients [9]. This indicates that anosmia and ageusia are several of the important symptoms to diagnose COVID-19 in the early stages, and to detect patients who are asymptomatic or have mild symptoms [12,15]. The official classification of the levels of illness is based on the ability to breathe, chest pain, and ability to talk or move [1], although often the ability to breathe becomes the major criteria. The level called *mild* refers to the lack of these signs of severity, although often the pains and severity of symptoms in patients classified to have mild or moderate symptoms are not mild or moderate (Table 1). Some patients with mild symptoms have fatigue and pain in various parts of the body for a prolonged length of time [33]. This indicates that patients with COVID-19 need to be diagnosed as soon as possible to enable immediate treatment not only to avoid the symptoms turning into a severe condition but also to avoid the prolonged time of these symptoms.

As described in this study, social media can become an important platform to collect data from outpatients, especially in the era of infectious disease pandemics, when hospitals are overwhelmed with patients that have severe conditions and patients with mild or moderate symptoms are often overlooked. Following the outbreak of COVID-19, virtual visits to doctors have started to become more popular, which can provide medical service without in-person visits. The benefits of social media are that patients can post their symptoms and questions whenever they are in need without appointments and can receive responses from the public, either patients, nurses, or doctors, through the internet, which enables real-time monitoring of the infected condition, patient symptoms, and the clinical course. In the United States, there is a project called DETECT, which uses a smartphone and a wristband to sense changes in the heart rate, which can be an indicator of changes in health conditions [34]. This device also allows the movement of people to be tracked, providing epidemiologically useful data. However, DETECT cannot distinguish between changes due to COVID-19 or other causes. The use of social media in addition to these other media (ie, virtual doctor visits and portable tracking devices) can become a strong monitoring method in the era of infectious disease pandemics. What is necessary to further strengthen the merits of social media could be the establishment of a system that serves as a route from the social media to, for example, virtual doctors or clinics nearby. At the current stage, obtaining enough information to provide diagnosis and treatments through social media sites contains challenging aspects for medical professionals. A system that guides patients from social media to the adequate medical support they need, to counselling, to other types of support systems, and the establishment of the rules in detail to organize such systems will help social media to become one of the new tools for the medical system in the near future.

Principal Results

Our analysis of a social media database of COVID-19 patients' posts demonstrated that olfactory dysfunction of a large portion of patients with COVID-19 persisted at least several weeks to months. This information is important in understanding the pathophysiology of COVID-19–induced anosmia. Although accurate diagnosis of patients is not always obtained with social media–based analysis, it may be a useful tool to collect a large

amount of data on symptoms and the clinical course of worldwide, rapidly growing, infectious diseases.

Limitations

Our study explores the benefits and limits of using social media membership groups in studying the symptoms of COVID-19. The patients are suspected COVID-19 patients and not all of them have been tested for COVID-19, but all have experienced loss of the senses of smell and taste.

Comparison With Prior Work

This is the first study to determine the benefits and limits of using social media in studying the symptoms of COVID-19.

Conclusions

Social media membership groups can function as a source of information and mental support for patients with COVID-19. Although some improvements in the system are necessary for its use for research purposes, it is a rare source of a large amount of data for researchers as well.

Acknowledgments

SK would like to express her special gratitude to Dr Amy Poehlman who edited the English of the paper and provided comments. SK would like to thank Christine Kelly, the founder of the social media membership group *COVID-19 Smell and Taste Loss* for her comments to the draft. The authors also wish to thank the members of *COVID-19 Smell and Taste Loss* who participated in this project. This study was partially supported by MSD Life Science Foundation, Public Interest Incorporated Foundation.

Conflicts of Interest

None declared.

References

1. Coronavirus: symptoms. World Health Organization. URL: https://www.who.int/health-topics/coronavirus#tab=tab_3 [accessed 2021-04-09]
2. Interim clinical guidance for management of patients with confirmed coronavirus disease (COVID-19). Centers for Disease Control and Prevention. URL: https://www.cdc.gov/coronavirus/2019-ncov/hcp/clinical-guidance-management-patients.html [accessed 2021-04-09]
3. Oran DP, Topol EJ. Prevalence of asymptomatic SARS-CoV-2 infection: a narrative review. Ann Intern Med 2020 Sep 01;173(5):362-367 [FREE Full text] [doi: 10.7326/M20-3012] [Medline: 32491919]
4. Poletti P, Tirani M, Cerada D, Trentini F, Guzzetta G, Sabatino G, ATS Lombardy COVID-19 Task Force, et al. Probability of symptoms and critical disease after SARS-CoV-2 infection. arXiv. Preprint posted online on June 15, 2020.
5. Sakurai A, Sasaki T, Kato S, Hayashi M, Tsuzuki S, Ishihara T, et al. Natural history of asymptomatic SARS-CoV-2 infection. N Engl J Med 2020 Aug 27;383(9):885-886 [FREE Full text] [doi: 10.1056/NEJMc2013020] [Medline: 32530584]
6. Yang R, Gui X, Xiong Y. Comparison of clinical characteristics of patients with asymptomatic vs symptomatic coronavirus disease 2019 in Wuhan, China. JAMA Netw Open 2020 May 01;3(5):e2010182. [doi: 10.1001/jamanetworkopen.2020.10182] [Medline: 32459353]
7. Hao X, Cheng S, Wu D, Wu T, Lin X, Wang C. Reconstruction of the full transmission dynamics of COVID-19 in Wuhan. Nature 2020 Aug;584(7821):420-424. [doi: 10.1038/s41586-020-2554-8] [Medline: 32674112]
8. Gane S, Kelly C, Hopkins C. Isolated sudden onset anosmia in COVID-19 infection. A novel syndrome? Rhinology 2020 Jun 01;58(3):299-301. [doi: 10.4193/Rhin20.114] [Medline: 32240279]
9. Lee Y, Min P, Lee S, Kim S. Prevalence and duration of acute loss of smell or taste in COVID-19 patients. J Korean Med Sci 2020 May 11;35(18):e174 [FREE Full text] [doi: 10.3346/jkms.2020.35.e174] [Medline: 32383370]
10. Spinato G, Fabbris C, Polesel J, Cazzador D, Borsetto D, Hopkins C, et al. Alterations in smell or taste in mildly symptomatic outpatients with SARS-CoV-2 infection. JAMA 2020 May 26;323(20):2089-2090 [FREE Full text] [doi: 10.1001/jama.2020.6771] [Medline: 32320008]
11. Zhang Q, Shan KS, Abdollahi S, Nace T. Anosmia and ageusia as the only indicators of coronavirus disease 2019 (COVID-19). Cureus 2020 May 01;12(5):e7918. [doi: 10.7759/cureus.7918] [Medline: 32494532]
12. Parma V, Ohla K, Veldhuizen MG, Niv MY, Kelly CE, Bakke AJ, GCCR Group Author, et al. More than smell-COVID-19 is associated with severe impairment of smell, taste, and chemesthesis. Chem Senses 2020 Oct 09;45(7):609-622 [FREE Full text] [doi: 10.1093/chemse/bjaa041] [Medline: 32564071]
13. Iravani B, Arshamian A, Ravia A, Mishor E, Snitz K, Shushan S, et al. Relationship between odor intensity estimates and COVID-19 prevalence prediction in a Swedish population. Chem Senses 2020 May 22:bjaa034 [FREE Full text] [doi: 10.1093/chemse/bjaa034] [Medline: 32441744]
14. Moein ST, Hashemian SM, Mansourafshar B, Khorram-Tousi A, Tabarsi P, Doty RL. Smell dysfunction: a biomarker for COVID-19. Int Forum Allergy Rhinol 2020 Aug;10(8):944-950 [FREE Full text] [doi: 10.1002/alr.22587] [Medline: 32301284]

15. Gerkin R, Ohla K, Veldhuizen M, Joseph PV, Kelly CE, Bakke AJ, GCCR Group Author. Recent smell loss is the best predictor of COVID-19 among individuals with recent respiratory symptoms. Chem Senses 2021 Jan 01;46:bjaa081 [FREE Full text] [doi: 10.1093/chemse/bjaa081] [Medline: 33367502]
16. Menni C, Valdes AM, Freidin MB, Sudre CH, Nguyen LH, Drew DA, et al. Real-time tracking of self-reported symptoms to predict potential COVID-19. Nat Med 2020 Jul;26(7):1037-1040 [FREE Full text] [doi: 10.1038/s41591-020-0916-2] [Medline: 32393804]
17. Long Q, Tang X, Shi Q, Li Q, Deng H, Yuan J, et al. Clinical and immunological assessment of asymptomatic SARS-CoV-2 infections. Nat Med 2020 Aug;26(8):1200-1204. [doi: 10.1038/s41591-020-0965-6] [Medline: 32555424]
18. Merchant RM, Asch DA, Crutchley P, Ungar LH, Guntuku SC, Eichstaedt JC, et al. Evaluating the predictability of medical conditions from social media posts. PLoS One 2019;14(6):e0215476 [FREE Full text] [doi: 10.1371/journal.pone.0215476] [Medline: 31206534]
19. Shen C, Chen A, Luo C, Zhang J, Feng B, Liao W. Using reports of symptoms and diagnoses on social media to predict COVID-19 case counts in mainland China: observational infoveillance study. J Med Internet Res 2020 May 28;22(5):e19421 [FREE Full text] [doi: 10.2196/19421] [Medline: 32452804]
20. COVID-19: print resources. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/coronavirus/2019-ncov/communication/print-resources.html?Sort=Date%3A%3Adesc&Search=symptoms [accessed 2021-04-09]
21. Lokman SM, Rasheduzzaman M, Salauddin A, Barua R, Tanzina AY, Rumi MH, et al. Exploring the genomic and proteomic variations of SARS-CoV-2 spike glycoprotein: a computational biology approach. Infect Genet Evol 2020 Oct;84:104389 [FREE Full text] [doi: 10.1016/j.meegid.2020.104389] [Medline: 32502733]
22. Shen Z, Xiao Y, Kang L, Ma W, Shi L, Zhang L, et al. Genomic diversity of severe acute respiratory syndrome-coronavirus 2 in patients with coronavirus disease 2019. Clin Infect Dis 2020 Jul 28;71(15):713-720 [FREE Full text] [doi: 10.1093/cid/ciaa203] [Medline: 32129843]
23. Young S, Fernandez AP. Skin manifestations of COVID-19. Cleve Clin J Med 2020 May 14:1 [FREE Full text] [doi: 10.3949/ccjm.87a.ccc031] [Medline: 32409442]
24. Hopkins C, Surda P, Whitehead E, Kumar BN. Early recovery following new onset anosmia during the COVID-19 pandemic - an observational cohort study. J Otolaryngol Head Neck Surg 2020 May 04;49(1):26 [FREE Full text] [doi: 10.1186/s40463-020-00423-8] [Medline: 32366299]
25. Lechien JR, Chiesa-Estomba CM, De Siati DR, Horoi M, Le Bon SD, Rodriguez A, et al. Olfactory and gustatory dysfunctions as a clinical presentation of mild-to-moderate forms of the coronavirus disease (COVID-19): a multicenter European study. Eur Arch Otorhinolaryngol 2020 Aug;277(8):2251-2261 [FREE Full text] [doi: 10.1007/s00405-020-05965-1] [Medline: 32253535]
26. Pellegrino R, Cooper KW, Di Pizio A, Joseph PV, Bhutani S, Parma V. Corona viruses and the chemical senses: past, present, and future. Chem Senses 2020 May 14:bjaa031 [FREE Full text] [doi: 10.1093/chemse/bjaa031] [Medline: 32407464]
27. Hummel T, Landis BN, Hüttenbrink KB. Smell and taste disorders. GMS Curr Top Otorhinolaryngol Head Neck Surg 2011;10:Doc04. [doi: 10.3205/cto000077] [Medline: 22558054]
28. Hummel T, Whitcroft KL, Andrews P, Altundag A, Cinghi C, Costanzo RM, et al. Position paper on olfactory dysfunction. Rhinol Suppl 2017 Mar;54(26):1-30. [doi: 10.4193/Rhino16.248] [Medline: 29528615]
29. Miwa T, Ikeda K, Ishibashi T, Kobayashi M, Kondo K, Matsuwaki Y, et al. Clinical practice guidelines for the management of olfactory dysfunction - secondary publication. Auris Nasus Larynx 2019 Oct;46(5):653-662 [FREE Full text] [doi: 10.1016/j.anl.2019.04.002] [Medline: 31076272]
30. Politi LS, Salsano E, Grimaldi M. Magnetic resonance imaging alteration of the brain in a patient with coronavirus disease 2019 (COVID-19) and anosmia. JAMA Neurol 2020 Aug 01;77(8):1028-1029. [doi: 10.1001/jamaneurol.2020.2125] [Medline: 32469400]
31. Bagheri SH, Asghari A, Farhadi M, Shamshiri AR, Kabir A, Kamrava SK, et al. Coincidence of COVID-19 epidemic and olfactory dysfunction outbreak in Iran. Med J Islam Repub Iran 2020;34:62 [FREE Full text] [doi: 10.34171/mjiri.34.62] [Medline: 32974228]
32. Dubé M, Le Coupanec A, Wong AHM, Rini JM, Desforges M, Talbot PJ. Axonal transport enables neuron-to-neuron propagation of human coronavirus OC43. J Virol 2018 Sep 01;92(17):e00404-e00418 [FREE Full text] [doi: 10.1128/JVI.00404-18] [Medline: 29925652]
33. Horowitz J. Surviving COVID-19 may not feel like recovery for some. The New York Times. 2020 May 10. URL: https://www.nytimes.com/2020/05/10/world/europe/coronavirus-italy-recovery.html?smid=fb-nytimes&smtyp=cur&fbclid=IwAR3_cp2ASvTJK7l4zeFA8epGd7yxjUuL2MZX9iG72ph_zRs0nuWaXzvoSt4 [accessed 2021-04-09]
34. Scripps Research invites public to join app-based DETECT study, leveraging wearable data to potentially flag onset of viral illnesses. Scripps Research. URL: https://www.scripps.edu/news-and-events/press-room/2020/20200325-detect-study-viral-illnesses.html [accessed 2021-04-09]

Abbreviations

CDC: Centers for Disease Control and Prevention
WHO: World Health Organization

Edited by C Basch; submitted 12.12.20; peer-reviewed by P Joseph, V Minichiello; comments to author 28.02.21; revised version received 16.03.21; accepted 25.03.21; published 22.04.21.

Please cite as:
Koyama S, Ueha R, Kondo K
Loss of Smell and Taste in Patients With Suspected COVID-19: Analyses of Patients' Reports on Social Media
J Med Internet Res 2021;23(4):e26459
URL: https://www.jmir.org/2021/4/e26459
doi:10.2196/26459
PMID:33788699

Original Paper

Google Trends for Pain Search Terms in the World's Most Populated Regions Before and After the First Recorded COVID-19 Case: Infodemiological Study

Istvan-Szilard Szilagyi[1], MSc, DMedSc; Torsten Ullrich[2,3], Dip Math, Dr techn; Kordula Lang-Illievich[1], MD, MSc; Christoph Klivinyi[1], MD; Gregor Alexander Schittek[1], MD; Holger Simonis[4], MD, DMedSc, MHBA; Helmar Bornemann-Cimenti[1], MD, DMedSc, MSc

[1]Department of Anesthesiology and Intensive Care Medicine, Medical University of Graz, Graz, Austria

[2]Institute of Computer Graphics and Knowledge Visualisation, Graz University of Technology, Graz, Austria

[3]Fraunhofer Austria Research GmbH, Graz, Austria

[4]Department of Anesthesiology, Perioperative and Intensice Care Medicine, University Hospital Salzburg, Salzburg, Austria

Corresponding Author:
Helmar Bornemann-Cimenti, MD, DMedSc, MSc
Department of Anesthesiology and Intensive Care Medicine
Medical University of Graz
Auenbruggerplatz 5/5
Graz, 8036
Austria
Phone: 1 316 385
Email: helmar.bornemann@medunigraz.at

Abstract

Background: Web-based analysis of search queries has become a very useful method in various academic fields for understanding timely and regional differences in the public interest in certain terms and concepts. Particularly in health and medical research, Google Trends has been increasingly used over the last decade.

Objective: This study aimed to assess the search activity of pain-related parameters on Google Trends from among the most populated regions worldwide over a 3-year period from before the report of the first confirmed COVID-19 cases in these regions (January 2018) until December 2020.

Methods: Search terms from the following regions were used for the analysis: India, China, Europe, the United States, Brazil, Pakistan, and Indonesia. In total, 24 expressions of pain location were assessed. Search terms were extracted using the local language of the respective country. Python scripts were used for data mining. All statistical calculations were performed through exploratory data analysis and nonparametric Mann–Whitney *U* tests.

Results: Although the overall search activity for pain-related terms increased, apart from pain entities such as headache, chest pain, and sore throat, we observed discordant search activity. Among the most populous regions, pain-related search parameters for shoulder, abdominal, and chest pain, headache, and toothache differed significantly before and after the first officially confirmed COVID-19 cases (for all, $P<.001$). In addition, we observed a heterogenous, marked increase or reduction in pain-related search parameters among the most populated regions.

Conclusions: As internet searches are a surrogate for public interest, we assume that our data are indicative of an increased incidence of pain after the onset of the COVID-19 pandemic. However, as these increased incidences vary across geographical and anatomical locations, our findings could potentially facilitate the development of specific strategies to support the most affected groups.

(J Med Internet Res 2021;23(4):e27214) doi:10.2196/27214

KEYWORDS

COVID-19; data mining; Google Trends; incidence; internet; interest; pain; research; trend

Introduction

Eysenbach [1] in 2002 defined Infodemiology as the "science of distribution and determinants of information in an electronic medium, specifically the Internet, or in a population, with the ultimate aim to inform public health and public policy." Since then, web-based sources have been suggested to be valuable in various academic fields for understanding timely and regional differences in the public interest in certain terms and concepts [2].

Especially in health and medical research, infodemiological studies using Google Trends have increased over the last decade [3]. Recently, several studies have used Google Trends data to elucidate the progression of the COVID-19 pandemic. For example, Husain et al [4] used Google Trends to track public interest in COVID-19 [4] and reported that in the United States, when containment and mitigation strategies were first implemented, public interest was very low. However, the search volume of COVID-19–related terms increased and correlated with the increase in positive findings on COVID-19 tests nationwide. Furthermore, Walker et al [5] correlated the search activity for the term "loss-of-smell," an early symptom of COVID-19, to the number of daily confirmed cases and COVID-19 deaths. Similarly, Jimenez et al [6] correlated searches for different symptoms of COVID-19 with daily incidences. Another study used Google Trends to identify mental health consequences related to physical distancing during the lockdown. Notably, Knipe et al [7] reported a reduction in search activity related to suicide and depression after the announcement of the pandemic.

The use of internet searches to gather information about patients experiencing pain has been insufficiently studied. Of 2 studies on populations of people with chronic pain, 1 reported that 24% of the population searched the internet for pain-related information, and the other reported that 39% of the population performed internet searches [8,9]. However, both studies are more than 10 years old. Considering the recent growth of the usage of online media, it is reasonable to suppose that the proportion of individuals searching for pain-related information on the internet increased during the last decade.

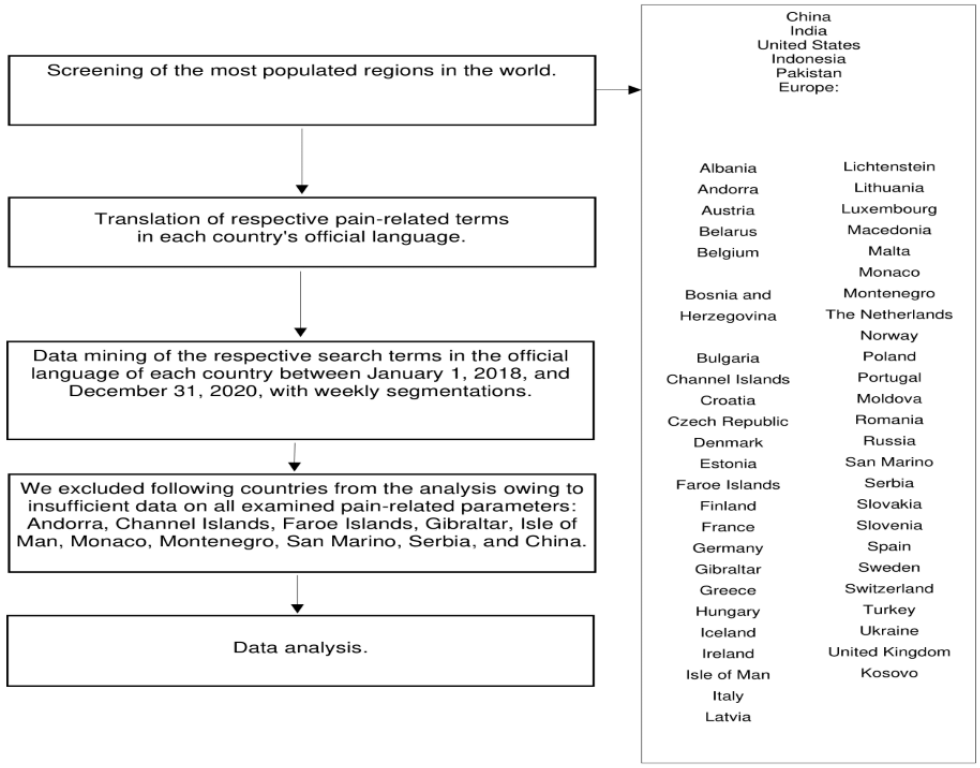
The COVID-19 pandemic has affected the health condition and quality of life of people with chronic pain in 2 ways. First, several changes have led to increased susceptibility to and a higher risk of pain exacerbation. Specifically, the availability of health care was significantly reduced and an atmosphere of fear and isolation was developed [10]. Previous experiences with epidemics have shown a latent, but long-term, deterioration of psychosocial problems [11]. Therefore, it has also been postulated that chronic pain, as a biopsychosocial phenomenon, exacerbates in response to the COVID-19 situation [12]. Second, acute stress can alleviate chronic pain [13,14]. Indeed, acute stress in response to disasters alleviates pain [15]. One could therefore hypothesize that the COVID-19 situation alleviates chronic pain, at least at the onset of the pandemic. To date, no studies have examined changes in the experience of chronic pain after the onset of the COVID-19 pandemic in large patient populations. To bridge this knowledge gap, search activity can be used as a surrogate for the public interest in pain. This does not necessarily reflect the incidence or the burden of pain in the general population; however, acute changes can be interpreted as a momentary shift of attention and thus highlight the importance of these symptoms.

This study aimed to assess the Google search activity of pain-related expressions in the most populated regions worldwide and to use them as markers of the social importance of pain before and during the first wave of the COVID-19 pandemic.

Methods

This infodemiological study was designed and carried out at Graz Medical University in cooperation with the Graz University of Technology and Fraunhofer Austria. The workflow of this study is presented in Figure 1. The temporary popularity of pain-related search terms was assessed using Google Trends [16]. Google Trends delivers data on web-based search queries for specified timeframes and countries. Google Trends offers 2 analysis modes: one with "terms" and the other with "topics." In contrast to the analysis with "terms," Google Trends analysis with "topics" corresponds to a fuzzy search. The difference is that "topics" include all search terms related to a particular aspect, whereas "terms" are specific, and the results will only show the relative volume of the corresponding term. Since different types of pain are related to one another—for example, neck pain and shoulder pain or dysmenorrhea and pelvic pain—a fuzzy search that includes related terms is not adequately discriminative. Furthermore, it is unclear which terms are summarized in a specific "topic." Hence, smaller, discriminative trend analysis with "terms" was preferred. Data are presented as relative popularity, normalized to the region and period and scaled from 0 to 100. Subsequently, the data were exported for further calculations [17]. The study followed the methodology framework developed by Mavragani and Ochoa [18].

Figure 1. Workflow for data retrieval and processing pipeline in this study.



The search terms used in this study were based on a recent study of global internet searches associated with pain [19]. In total, 24 expression entities based on anatomical regions were used: “abdominal pain,” “back pain,” “breast pain,” “chest pain,” “dysmenorrhea,” “dyspareunia,” “ear pain,” “epigastric pain,” “eye pain,” “groin pain,” “headache,” “knee pain,” “low back pain,” “neck pain,” “odynophagia,” “pelvic pain,” “penile pain,” “podalgia,” “rectal pain,” “shoulder pain,” “sore throat,” “testicular pain,” “toothache,” and “wrist pain.” The results for 3 of these terms (“penile pain,” “podalgia,” and “rectal pain”) were excluded from further analysis because of insufficient data. Weekly data sets were downloaded from January 1, 2018, to December 31, 2020. Google Trends analysis was carried out in February 2021.

Our intention was to provide an overview of the development of the aforementioned pain-related search parameters in the most populated regions worldwide. The most populous areas worldwide were extracted from the recommendations of the United Nations and a web-based population database [20,21]. Europe, consisting of individual European countries, is defined as a single region in this study. Figure 1 shows all the selected regions and the individual European countries. We identified the following as the most populated regions: Europe, the United States, China, India, Pakistan, Brazil, and Indonesia. Search terms were extracted in the respective language of each country. The data set of Europe is accumulated; it consists of all available nationwide data sets of the European countries weighted by their population. Details including the complete list of countries, their population sizes, the corresponding weighting factors, and the date of the first officially reported COVID-19 case of the country were provided by the European Centre for Disease Prevention and Control. Using this list of European countries, we retrieved a data vector for each country and pain type in the respective language of each country. Thereafter, the individual vectors were weighted with the country’s population and accumulated to a European vector (Multimedia Appendix 1). Since the weighted summation with sum 1 is an affine map, the result is also normalized. Calculations with the European vector thus correspond to stratified sampling; that is, individual characteristics of the target group deemed relevant for the study are transferred to the sample in approximately the same proportion as those in the population.

Data from China could not be extracted adequately, and data extraction yielded numerous error parameters. Therefore, China had to be excluded from our analysis.

We assumed that the occurrence of biological and psychosocial effects of the pandemic correlated with the first appearance of the virus in a certain region. Accordingly, we defined the beginning of the “COVID-19 period” for each geographic region with the first officially confirmed case for each country on the basis of reports from the European Centre for Disease Prevention and Control [22]. The date of the first officially reported COVID-19 case for each country was used as the independent variable for our calculations.

For data extraction, we used Python scripts. The Google Trends analysis data were obtained with the software libraries pytrends and pandas [23,24] using Python. The number of COVID-19 cases were obtained from the European Centre for Disease Prevention and Control on February 9, 2021. Results from countries with erroneous data or insufficient cases and subsequently inconsistent data are excluded from the study (Figure 1).

IBM SPSS Statistics (version 26, IBM Corp), Microsoft Office 365, and LibreOffice 7.1 were used for the statistical

calculations. Explorative data analysis and nonparametric Mann–Whitney *U* tests were performed. The level of significance was not corrected for multiple comparisons. These corrections are necessary for multiple tests carried out with the same samples: the greater the number of hypotheses tested on 1 data set, the higher the probability that 1 of them is (incorrectly) assumed to be true. As the data were collected separately by country and pain type, the precondition for test corrections (all tests on 1 data set) was not met; that is, Bonferroni corrections or similar methods were not required and hence not applied.

Additionally, we performed a visual analysis to illustrate the difference between significant and nonsignificant changes in the time series data (Figures 2 and 3) and to plot a side-by-side comparison of the progression of pain-related search criteria with the so-called waves of COVID-19 outbreaks (Figure 4). This analysis has been performed for "headache," as this query term increased significantly in all geographic regions, and for "wrist pain," as this query term did not show any significant changes.

For both terms, the analysis involved a seasonal autoregressive model whose parameters have been determined using Google Trends vectors retrieved for 2018 and 2019. Using these models, weekly differences between the model values for 2018-2019 and 2020 were calculated and plotted (Figures 2 and 4).

Figure 2. Trends for headache.

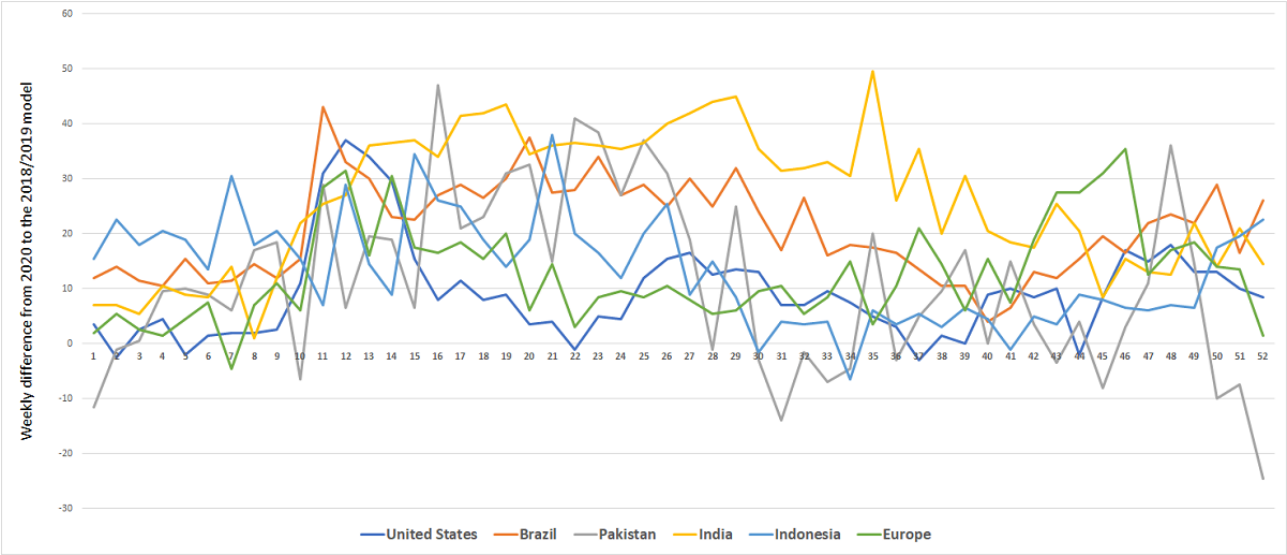


Figure 3. Trends for wrist pain.

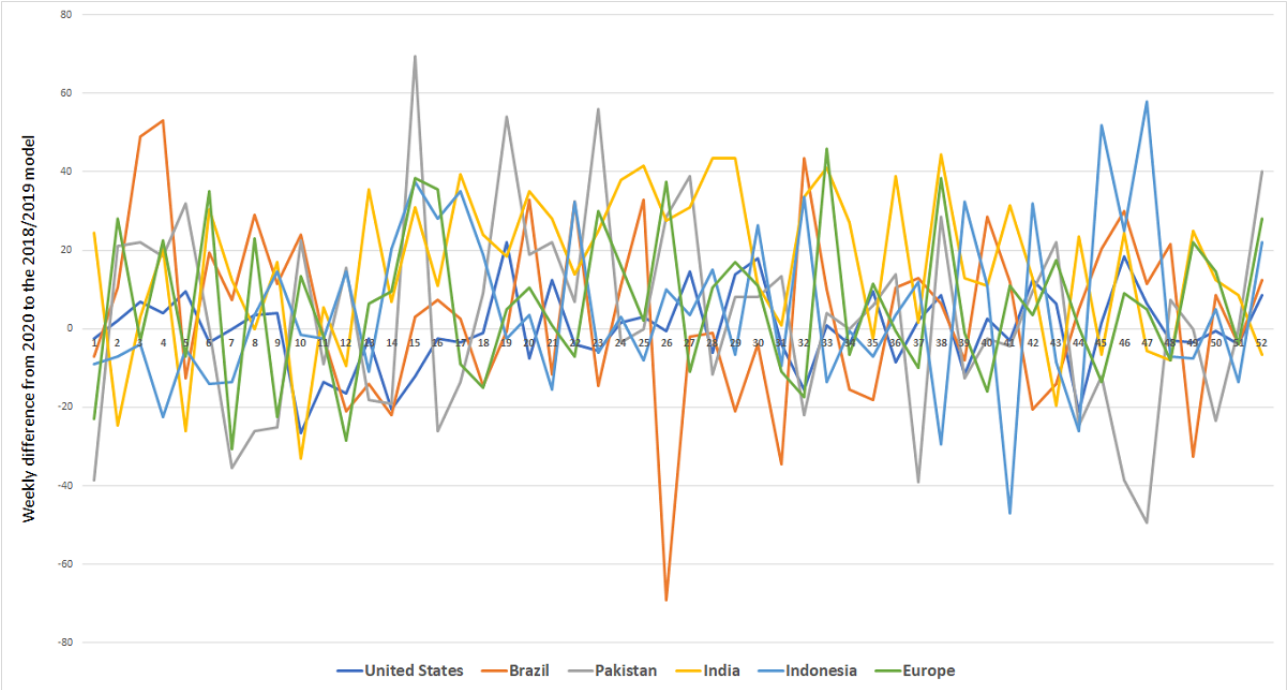


Figure 4. An overview of the 5 most populous countries (last row) and European countries in an arrangement that corresponds to the geography of Europe. Only countries for which sufficient data were available are included; data on China and Serbia were inadequate. Each chart consists of 4 rows, each used by a subchart, and 4 vertical columns, each covering 1 quarter of 2020.

Results

Our results demonstrate significant differences in pain-related search queries by comparing quantities before and after the first confirmed COVID-19 case. We observed a peak in the incidence of pain-related search parameters in March and April 2020 (Figure 5). Abdominal pain, dysmenorrhea, dyspareunia, groin pain, eye pain, knee pain, low back pain, and pelvic pain were the only pain types with a significantly decreased frequency of search relevance in some of the most populous geographic regions since the COVID-19 outbreak (Table 1). In contrast, the frequency of searches related to back pain, breast pain, chest pain, ear pain, headache, odynophagia, neck pain, shoulder pain, sore throat, testicular pain, toothache, and wrist pain

significantly increased after the first confirmed case of COVID-19 was reported (Table 1, Figure 4).

Figure 5. Summation trends of pain-related search parameters in the most populated regions worldwide (Europe, the United States, Brazil, Pakistan, India, and Indonesia).

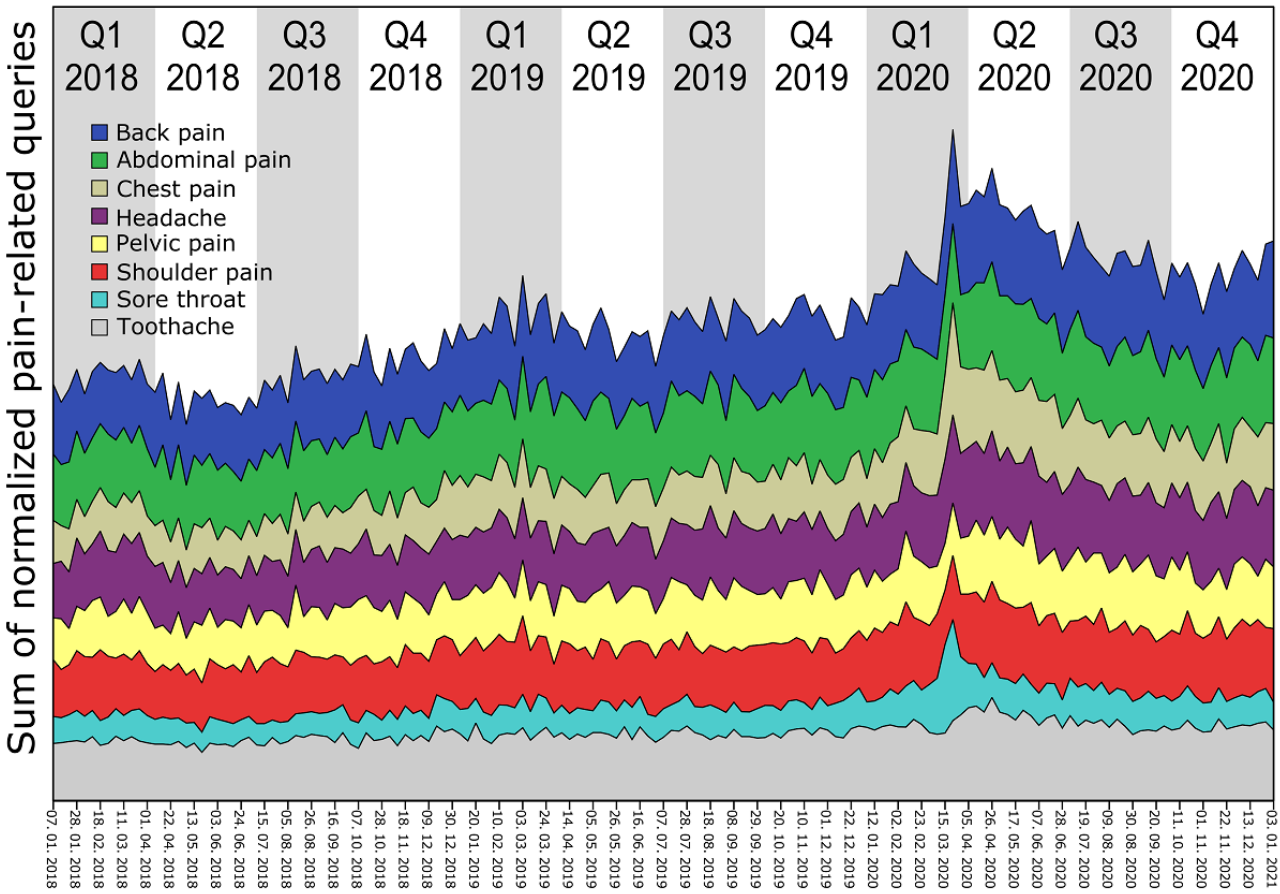


Table 1. Comparison of search trends between January 1, 2018, and December 31, 2020, for pain-related terms before and after the first officially confirmed COVID-19 case in in Europe, the United States, Brazil, Pakistan, India, and Indonesia (China was excluded from the data set owing to insufficient data).

Pain type	Europe (*P* value)	United States (*P* value)	Brazil (*P* value)	Pakistan (*P* value)	India (*P* value)	Indonesia (*P* value)	Overall (*P* value)
Back pain	<.001[a]	.09[b]	<.001[a]	.04[a]	<.001[a]	<.001[a]	<.001[a]
Abdominal pain	<.001[a]	<.001[c]	<.001[a]	.05[a]	<.001[a]	<.001[a]	<.001[a]
Breast pain	<.001[a]	.07[b]	<.001[a]	.10[b]	<.001[a]	<.001[a]	<.001[a]
Chest pain	<.001[a]	<.001[a]	<.001[a]	<.001[a]	<.001[a]	<.001[a]	<.001[a]
Dysmenorrhea	.15[b]	.35[b]	.09[b]	.13[b]	<.001[a]	.009[c]	.12[b]
Dyspareunia	.12[b]	<.001[c]	.31[b]	.22[b]	.20[b]	.86[b]	.28[b]
Ear pain	.01[a]	.84>[b]	.06[b]	.01[a]	<.001[a]	.003[a]	<.001[a]
Epigastric pain	.45[b]	.61[b]	<.001[a]	.31[b]	.11[b]	N/A[d]	<.001[a]
Groin pain	.64[b]	<.001[c]	<.010[a]	.84[b]	<.001[a]	.03[c]	.73[b]
Eye pain	<.001[a]	.47[b]	<.001[a]	.02[a]	.05[a]	.03[c]	<.001[a]
Headache	<.001[a]	<.001[a]	<.001[a]	.03[a]	<.001[a]	<.001[a]	<.001[a]
Knee pain	.43[b]	<.001[c]	.42[b]	.23[b]	.04[a]	<.001[a]	.09[b]
Low back pain	.01[a]	<.001[c]	<.001[a]	.06[b]	.23[b]	<.001[c]	.77[b]
Odynophagia	<.001[a]	.45[b]	<.001[a]	.05[b]	.22[b]	.40[b]	.01[a]
Neck pain	<.001[a]	.44[b]	.03[a]	.08[b]	<.001[a]	<.001[a]	<.001[a]
Pelvic pain	.01[a]	<.001[c]	<.001[a]	.03[a]	<.001[a]	<.001[a]	<.001[a]
Shoulder pain	.01[a]	.40[b]	<.001[a]	<.001[a]	<.001[a]	<.001[a]	<.001[a]
Sore throat	<.001[a]	.05[a]	<.001[a]	<.001[a]	<.001[a]	.23[b]	<.001[a]
Testicular pain	.42[b]	.16[b]	.10[b]	.45[b]	<.001[a]	<.001[a]	<.001[a]
Toothache	<.001[a]	.31[b]	<.001[a]	<.001[a]	<.001[a]	<.001[a]	<.001[a]
Wrist pain	.08[b]	.94[b]	.11[b]	.77[b]	<.001[a]	.04[a]	<.001[a]

[a]Relevance of the pain-related search parameter increased significantly after the first confirmed COVID-19 case was reported.

[b]No significant changes.

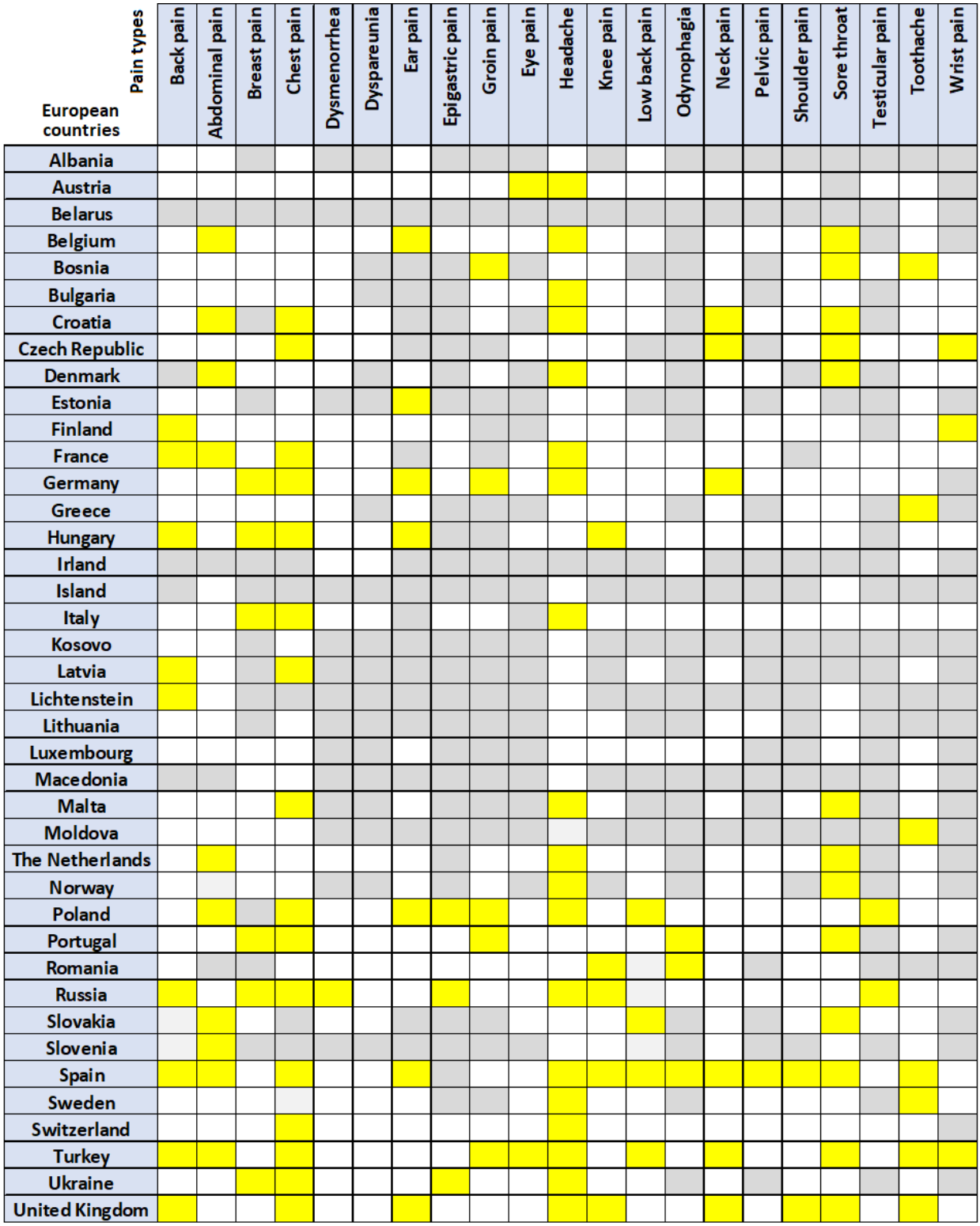
[c]Relevance of pain-related search parameter decreased significantly after the first confirmed COVID-19 case was reported.

[d]N/A: not applicable; no data available/inadequate number of cases.

Our data show that in Europe, Brazil, Pakistan, and India, all pain entities listed above increased significantly or remained unchanged after the first confirmed COVID-19 case was reported. In contrast, in the United States, we observed an inhomogeneous trend, with the frequency of some pain-related search terms decreasing significantly after the first confirmed COVID-19 case was reported, whereas that of the others increased or remained unchanged (Table 1).

A comparison among individual European countries revealed that the most frequently observed significant differences among countries were noted for headache, chest pain, sore throat, abdominal pain, and back pain since the COVID-19 outbreak. With respect to individual European countries, Spain displayed the most frequent significant pre– and post–COVID-19 discrepancies in the frequency of pain-related search queries (Figures 3 and 6).

Figure 6. Summation trends of pain-related search parameters in European countries. Yellow=relevance of pain-related search parameters changed significantly after the report of the first confirmed COVID-19 case, white=no significant changes, and gray=no data available or inadequate data.



The analysis of the difference plots for the term "headache" revealed that the weekly differences between the 2020 and the 2018-2019 models were more often significantly positive than negative; in other words, the number of queries for "headache" increased. As the Google Trends data are normalized with a range of 0 to 100, the 2018-2019 model was normalized accordingly (Figure 2).

If the search frequency of a term did not increase significantly, the differences would be expected to be as often positive as negative. Consequently, the difference plots would resemble random noise; for example, the search term "wrist pain" displayed no significant change in most geographic regions and its plots illustrate random noise around 0.

Furthermore, on assessing country-specific developments in pain-related search parameters, which have significantly

increased in relevance since the COVID-19 pandemic outbreak, it becomes apparent that the relevance of pain-related search parameters increased in almost all countries before the peak of the first wave of the COVID-19 pandemic. “Headache” was consistently high in relevance, especially in Belgium, Brazil, Germany, Indonesia, Italy, Russia, Turkey, and the United States. The relevance of “chest pain” and “sore throat” seemed to resonate with the first wave of the COVID-19 pandemic, as seen in several countries including Belgium, Croatia, France, Italy, the Netherlands, and the United Kingdom. In contrast, the frequency of “abdominal pain” appeared highly dynamic compared to that of COVID-19 cases (Figure 4).

Discussion

Principal Findings

Our results indicate that for most pain entities, the frequency of pain-related search queries significantly increased after the official report of the first confirmed COVID-19 case (Table 1). In particular, search queries for “chest pain” and “headache” significantly increased in all of the world's most populous regions included in this study.

On assessing the trends for the most populated regions, we found that the incidence of pain-related search criteria increased for back pain, breast pain, chest pain, ear pain, headache, neck pain, shoulder pain, sore throat, and toothache after the COVID-19 outbreak. With the exception of the United States and Indonesia, we observed a significant increase in pain-related search parameters for almost all the pain types (Table 1).

Part of our findings of increased numbers of searches of pain-related terms may reflect symptoms attributed to COVID-19. For example, sore throat [25] and headache [26] are frequent symptoms of COVID-19 and have been often discussed as signs of COVID-19 in the news media [27]. Consequently, an increased interest in this search term was expected. This may also apply to chest pain. Similar to sore throats and headaches, “chest tightness” was a term that was frequently used in the news media in connection with COVID-19 [27]. However, since chest pain is also a key symptom of major cardiac events, our data should be interpreted with caution as other studies have reported a drastic increase in heart attacks and cardiovascular deaths during the first wave of the COVID-19 pandemic [28]. This is postulated to be associated with delays in seeking help. Compared with other pain-related terms, the increase in searches for “toothache” was delayed but persisted at a higher level. The changes in search behavior for toothache may be explained by the fact that many patients had no opportunity or were too afraid to visit a dentist during the lockdown [29]. The number of dental visits decreased to less than one-fifth that before the onset of the COVID-19 pandemic [30]. Therefore, painful dental conditions may have remained untreated or may have developed over time. The variation in the search pattern of eye pain in response to the COVID-19 pandemic has been previously reported [31]. However, it remains unclear if this can be attributed to a direct effect of the virus or is the effect of changes in lifestyle during the lockdown (eg, more time spent in closed rooms and increased screen time). The same argument could hold true for neck pain; as internet usage for private purposes, remote working, and education increased, one can assume that problems with posture also became more prevalent [32].

It remains unclear why the search frequency for groin and knee pain decreased in certain populous regions before to after the onset of the COVID-19 pandemic. Each of these pain entities was associated with physical activity [33,34]. Owing to the potential consequences of a lockdown, decreased physical activity can be anticipated as a possible effect.

In Pakistan, India, Brazil, and Indonesia, the frequency of most pain-related keywords increased after the first confirmed COVID-19 case. Furthermore, Europe displayed a predominant increase in search queries across almost all pain types. On further inspection of individual European countries, we found that similar to the overall picture, that COVID-19–related symptoms dominate the increase in searches, such as “headache,” “sore throat,” or “chest pain” (Figure 6). On the contrary, in the United States the frequencies for most pain-related searches decreased.

Owing to our study design, we can only speculate the reasons for this geographically heterogeneous pattern. We assume that except for those symptoms (chest pain, headache, neck pain, sore throat, and toothache) that showed the same direction of change in all geographic regions, there are no direct or indirect biological effects of COVID-19 on pain. Thus, the reasons for the observed regional differences are most likely attributed to psychosocial dimensions. Differences in social systems and cultures are evident in various fields of health [35]. Especially in the more developed regions, the overall satisfaction with the health care system is generally high [36,37]. One could speculate that these differences contribute to the impact of the COVID-19 pandemic on people with chronic pain on an individual level.

A special situation was observed in the United States, in that we observed a reduction in the frequency of most search terms. This may be explained by the fact that the exposure to COVID-19 was time-shifted for different areas in the United States. Only the third wave of the pandemic encompassed the country simultaneously and its entirety.

Limitations

This study has limitations, which need to be addressed. Some of these limitations are related to our study design. It is evident that analyzing online searches may only reflect the behavior of internet users, thus potentially excluding relevant groups; for example, older or uneducated individuals [38]. Other demographic groups are potentially overrepresented. Younger individuals and women displayed a higher prevalence in health-related information searches on the internet [39].

Similarly, one can only speculate the reasons for the individual searches. Generally, internet searches are considered a surrogate for public interest. This does not necessarily imply that searches are carried out by patients experiencing a certain symptom. In our case, relatives, advocates, or researchers may also search for pain-related terms.

Figure 4 shows high fluctuations in the Google Trends data set. According to Google, only a sample is used for the trend

analysis and not the complete data set of all search queries [40]. Google does not disclose which sampling strategy is used. However, Figure 4 suggests that the population size has an influence on the sampling rate considering the number of search queries. The lower the sampling rate (ie, lesser data on which the data calculation is based), the greater the inaccuracies that are reflected in the fluctuations in the graphs [41].

Some limitations are specific to our study. For instance, our study only uses Google data and therefore represents the search behavior of only Google users. However, as Google has a nearly 70% worldwide market share, it is representative of the majority of internet users [42]. In our study, we excluded China as we did not obtain adequate data from this origin. The use of other sources including Baidu could have provided more details; however, as these data would not be comparable to Google Trends data, we decided against this approach.

We have summarized the data of European countries. In doing so, the exact timepoint of the "first official case" was not defined at the national level but rather on the continental level in in Europe. It can be assumed that the inaccuracy of these statistics for the temporal delimitation for the local onset of the pandemic is of the same order of magnitude as for the individual country data sets provided by Google Trends. The problem also exists in these data sets as outbreaks of the pandemic did not occur homogenously (eg, in all US states or in all provinces of China) simultaneously.

The definition of search terms is crucial when extracting information from search databases [38]. Our keywords represent anatomical locations instead of certain diagnoses (eg, fibromyalgia and migraine). This approach was previously introduced by Kamiński et al [19]. As we have not been aware of another already established search strategy for pain-related keywords in an infodemiological study, we decided to follow the method of Kamiński et al.

Conclusions

Our results indicate considerable changes in internet search behavior related to pain-related keywords for the most populated regions worldwide. There are many possible reasons for these changes in internet search behavior. As expected, we found that certain search terms that are closely related to COVID-19 symptoms are increasing, such as the pain entities of headache, chest pain, or sore throat. Our study describes the analysis of the trends in pain-related search parameters on the Google search network, as developed over a 2-year period.

In summary, apart from COVID-19–related pain symptoms, the frequency of search parameters related to pain across the most populated regions worldwide was observed to change over time before and after the onset of the COVID-19 pandemic. As internet searches are a surrogate for public interest, we assume that our data are indicative of an increased incidence of pain after the onset of the COVID-19 pandemic. However, as these increased incidences vary across both geographical and anatomical locations, our data could potentially facilitate the development of specific strategies to support the most affected groups.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary table.
[DOCX File , 19 KB - jmir_v23i4e27214_app1.docx]

References

1. Eysenbach G. Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the Internet. J Med Internet Res 2009;11(1):e11 [FREE Full text] [doi: 10.2196/jmir.1157] [Medline: 19329408]
2. Mavragani A, Ochoa G, Tsagarakis KP. Assessing the Methods, Tools, and Statistical Approaches in Google Trends Research: Systematic Review. J Med Internet Res 2018 Nov 06;20(11):e270 [FREE Full text] [doi: 10.2196/jmir.9366] [Medline: 30401664]
3. Nuti SV, Wayda B, Ranasinghe I, Wang S, Dreyer RP, Chen SI, et al. The use of google trends in health care research: a systematic review. PLoS One 2014 Oct;9(10):e109583 [FREE Full text] [doi: 10.1371/journal.pone.0109583] [Medline: 25337815]
4. Husain I, Briggs B, Lefebvre C, Cline DM, Stopyra JP, O'Brien MC, et al. Fluctuation of Public Interest in COVID-19 in the United States: Retrospective Analysis of Google Trends Search Data. JMIR Public Health Surveill 2020 Jul 17;6(3):e19969 [FREE Full text] [doi: 10.2196/19969] [Medline: 32501806]
5. Walker A, Hopkins C, Surda P. The use of google trends to investigate the loss of smell related searches during COVID-19 outbreak. Int Forum Allergy Rhinol 2020 Apr 11;10(7):839-847. [doi: 10.1002/alr.22580] [Medline: 32279437]
6. Jimenez A, Estevez-Reboredo R, Santed M, Ramos V. COVID-19 Symptom-Related Google Searches and Local COVID-19 Incidence in Spain: Correlational Study. J Med Internet Res 2020 Dec 18;22(12):e23518 [FREE Full text] [doi: 10.2196/23518] [Medline: 33156803]

7. Knipe D, Evans H, Marchant A, Gunnell D, John A. Mapping population mental health concerns related to COVID-19 and the consequences of physical distancing: a Google trends analysis. Wellcome Open Res 2020;5:82 [FREE Full text] [doi: 10.12688/wellcomeopenres.15870.2] [Medline: 32671230]
8. de Boer MJ, Versteegen GJ, van Wijhe M. Patients' use of the Internet for pain-related medical information. Patient Educ Couns 2007 Sep;68(1):86-97. [doi: 10.1016/j.pec.2007.05.012] [Medline: 17590563]
9. Corcoran TB, Haigh F, Seabrook A, Schug SA. A survey of patients' use of the internet for chronic pain-related information. Pain Med 2010 Apr;11(4):512-517. [doi: 10.1111/j.1526-4637.2010.00817.x] [Medline: 20202143]
10. Yasri S, Wiwanitkit V. Pain Management During the COVID-19 Pandemic. Pain Med 2020 Sep 01;21(9):2008 [FREE Full text] [doi: 10.1093/pm/pnaa200] [Medline: 32585008]
11. Mak IWC, Chu CM, Pan PC, Yiu MGC, Chan VL. Long-term psychiatric morbidities among SARS survivors. Gen Hosp Psychiatry 2009;31(4):318-326 [FREE Full text] [doi: 10.1016/j.genhosppsych.2009.03.001] [Medline: 19555791]
12. Clauw DJ, Häuser W, Cohen SP, Fitzcharles M. Considering the potential for an increase in chronic pain after the COVID-19 pandemic. Pain 2020 Aug;161(8):1694-1697 [FREE Full text] [doi: 10.1097/j.pain.0000000000001950] [Medline: 32701829]
13. Vachon-Presseau E, Martel M, Roy M, Caron E, Albouy G, Marin M, et al. Acute stress contributes to individual differences in pain and pain-related brain activity in healthy and chronic pain patients. J Neurosci 2013 Apr 17;33(16):6826-6833 [FREE Full text] [doi: 10.1523/JNEUROSCI.4584-12.2013] [Medline: 23595741]
14. Ferdousi M, Finn D. Stress-induced modulation of pain: Role of the endogenous opioid system. In: O'Mara S, editor. Progress in Brain Research. Cambridge, MA: Elsevier; 2018:121-177.
15. Koopman C, Classen C, Cardeña E, Spiegel D. When disaster strikes, acute stress disorder may follow. J Trauma Stress 1995 Jan;8(1):29-46. [doi: 10.1007/BF02105405] [Medline: 7712057]
16. Explore what the world is searching. Google Trends. URL: https://trends.google.com/trends/?geo=US [accessed 2021-02-18]
17. FAQ about Google Trends data. Google Trends. URL: https://support.google.com/trends/answer/4365533?hl=en [accessed 2020-10-28]
18. Mavragani A, Ochoa G. Google Trends in Infodemiology and Infoveillance: Methodology Framework. JMIR Public Health Surveill 2019 May 29;5(2):e13439 [FREE Full text] [doi: 10.2196/13439] [Medline: 31144671]
19. Kamiński M, Łoniewski I, Marlicz W. "Dr. Google, I am in Pain"-Global Internet Searches Associated with Pain: A Retrospective Analysis of Google Trends Data. Int J Environ Res Public Health 2020 Feb 04;17(3):954 [FREE Full text] [doi: 10.3390/ijerph17030954] [Medline: 32033087]
20. Population of the world and countries. Country Meters. URL: https://countrymeters.info [accessed 2021-02-18]
21. World Population Prospects 2019. United Nations Department of Economic and Social Affairs. URL: https://population.un.org/wpp/ [accessed 2021-02-18]
22. COVID-19 pandemic. European Centre for Disease Prevention and Control. URL: https://www.ecdc.europa.eu/en [accessed 2021-02-18]
23. Pseudo API for Google Trends. GitHub. URL: https://github.com/GeneralMills/pytrends [accessed 2021-02-18]
24. Pandas. URL: https://pandas.pydata.org [accessed 2021-02-18]
25. Micarelli A, Granito I, Carlino P, Micarelli B, Alessandrini M. Self-perceived general and ear-nose-throat symptoms related to the COVID-19 outbreak: a survey study during quarantine in Italy. J Int Med Res 2020 Oct;48(10):300060520961276 [FREE Full text] [doi: 10.1177/0300060520961276] [Medline: 33081538]
26. Poncet-Megemont L, Paris P, Tronchere A, Salazard J, Pereira B, Dallel R, et al. High Prevalence of Headaches During Covid-19 Infection: A Retrospective Cohort Study. Headache 2020 Nov;60(10):2578-2582 [FREE Full text] [doi: 10.1111/head.13923] [Medline: 32757419]
27. Jeon J, Baruah G, Sarabadani S, Palanica A. Identification of Risk Factors and Symptoms of COVID-19: Analysis of Biomedical Literature and Social Media Data. J Med Internet Res 2020 Oct 02;22(10):e20509 [FREE Full text] [doi: 10.2196/20509] [Medline: 32936770]
28. Wu J, Mamas MA, Mohamed MO, Kwok CS, Roebuck C, Humberstone B, et al. Place and causes of acute cardiovascular mortality during the COVID-19 pandemic. Heart 2021 Jan;107(2):113-119 [FREE Full text] [doi: 10.1136/heartjnl-2020-317912] [Medline: 32988988]
29. Aquilanti L, Gallegati S, Temperini V, Ferrante L, Skrami E, Procaccini M, et al. Italian Response to Coronavirus Pandemic in Dental Care Access: The DeCADE Study. Int J Environ Res Public Health 2020 Sep 24;17(19):6977 [FREE Full text] [doi: 10.3390/ijerph17196977] [Medline: 32987661]
30. Yu J, Zhang T, Zhao D, Haapasalo M, Shen Y. Characteristics of Endodontic Emergencies during Coronavirus Disease 2019 Outbreak in Wuhan. J Endod 2020 Jun;46(6):730-735. [doi: 10.1016/j.joen.2020.04.001] [Medline: 32360053]
31. Deiner M, Seitzman G, McLeod S, Chodosh J, Hwang D, Lietman T, et al. Ocular Signs of COVID-19 Suggested by Internet Search Term Patterns Worldwide. Ophthalmology 2021 Jan;128(1):167-169 [FREE Full text] [doi: 10.1016/j.ophtha.2020.06.026] [Medline: 32562704]
32. Feldmann A, Gasser O, Lichtblau F, Pujol E, Poese I, Dietzel C, et al. The Lockdown Effect: Implications of the COVID-19 Pandemic on Internet Traffic. In: IMC '20: Proceedings of the ACM Internet Measurement Conference. 2020 Oct Presented at: IMC '20: ACM Internet Measurement Conference; October 2020; Virtual Event p. 1-18 URL: https://dl.acm.org/doi/10.1145/3419394.3423658 [doi: 10.1145/3419394.3423658]

33. Pålsson A, Kostogiannis I, Lindvall H, Ageberg E. Hip-related groin pain, patient characteristics and patient-reported outcomes in patients referred to tertiary care due to longstanding hip and groin pain: a cross-sectional study. BMC Musculoskelet Disord 2019 Sep 14;20(1):432 [FREE Full text] [doi: 10.1186/s12891-019-2794-7] [Medline: 31521142]
34. Song J, Chang A, Chang R, Lee J, Pinto D, Hawker G, et al. Relationship of knee pain to time in moderate and light physical activities: Data from Osteoarthritis Initiative. Semin Arthritis Rheum 2018 Apr;47(5):683-688 [FREE Full text] [doi: 10.1016/j.semarthrit.2017.10.005] [Medline: 29103557]
35. Napier AD, Ancarno C, Butler B, Calabrese J, Chater A, Chatterjee H, et al. Culture and health. The Lancet 2014 Nov 15;384(9954):1607-1639 [FREE Full text] [doi: 10.1016/S0140-6736(14)61603-2]
36. Stepurko T, Pavlova M, Groot W. Overall satisfaction of health care users with the quality of and access to health care services: a cross-sectional study in six Central and Eastern European countries. BMC Health Serv Res 2016 Aug 02;16(a):342 [FREE Full text] [doi: 10.1186/s12913-016-1585-1] [Medline: 27485751]
37. Papanicolas I, Cylus J, Smith PC. An analysis of survey data from eleven countries finds that 'satisfaction' with health system performance means many things. Health Aff (Millwood) 2013 Apr;32(4):734-742. [doi: 10.1377/hlthaff.2012.1338] [Medline: 23569053]
38. Arora VS, McKee M, Stuckler D. Google Trends: Opportunities and limitations in health and health policy research. Health Policy 2019 Mar;123(3):338-341. [doi: 10.1016/j.healthpol.2019.01.001] [Medline: 30660346]
39. Kontos E, Blake KD, Chou WS, Prestin A. Predictors of eHealth usage: insights on the digital divide from the Health Information National Trends Survey 2012. J Med Internet Res 2014 Jul;16(7):e172 [FREE Full text] [doi: 10.2196/jmir.3117] [Medline: 25048379]
40. Inc. FAQ about Google Trends data. Google Trends. URL: https://support.google.com/trends/answer/4365533?hl=en [accessed 2021-04-14]
41. Steegmans J. The Pearls and Perils of Google Trends: A Housing Market Application. IDEAS. 2019. URL: https://ideas.repec.org/p/use/tkiwps/1911.html [accessed 2021-04-14]
42. Search Engine Market Share. Net MarketShare. URL: https://netmarketshare.com/search-engine-market-share.aspx [accessed 2021-03-27]

Edited by C Basch; submitted 17.01.21; peer-reviewed by M K., A Mavragani; comments to author 31.01.21; revised version received 04.03.21; accepted 10.04.21; published 22.04.21.

Please cite as:
Szilagyi IS, Ullrich T, Lang-Illievich K, Klivinyi C, Schittek GA, Simonis H, Bornemann-Cimenti H
Google Trends for Pain Search Terms in the World's Most Populated Regions Before and After the First Recorded COVID-19 Case: Infodemiological Study
J Med Internet Res 2021;23(4):e27214
URL: https://www.jmir.org/2021/4/e27214
doi:10.2196/27214
PMID:33844638

Original Paper

Communicating Scientific Uncertainty About the COVID-19 Pandemic: Online Experimental Study of an Uncertainty-Normalizing Strategy

Paul K J Han[1], MD, MA, MPH; Elizabeth Scharnetzki[1], DPhil; Aaron M Scherer[2], DPhil; Alistair Thorpe[3], DPhil; Christine Lary[1], DPhil; Leo B Waterston[1], MA; Angela Fagerlin[3,4], DPhil; Nathan F Dieckmann[5], DPhil

[1]Center for Outcomes Research and Evaluation, Maine Medical Center Research Institute, Portland, ME, United States

[2]Division of General Internal Medicine, University of Iowa Carver College of Medicine, Iowa City, IA, United States

[3]Department of Population Health Sciences, University of Utah, Salt Lake City, UT, United States

[4]Salt Lake City VA Center for Informatics Decision Enhancement and Surveillance, Salt Lake City, UT, United States

[5]School of Nursing, Oregon Health and Science University, Portland, OR, United States

Corresponding Author:
Paul K J Han, MD, MA, MPH
Center for Outcomes Research and Evaluation
Maine Medical Center Research Institute
509 Forest Avenue
Portland, ME, 04101
United States
Phone: 1 207 661 7619
Email: hanp@mmc.org

Abstract

Background: Communicating scientific uncertainty about public health threats such as COVID-19 is an ethically desirable task endorsed by expert guidelines on crisis communication. However, the communication of scientific uncertainty is challenging because of its potential to promote *ambiguity aversion*—a well-described syndrome of negative psychological responses consisting of heightened risk perceptions, emotional distress, and decision avoidance. Communication strategies that can inform the public about scientific uncertainty while mitigating ambiguity aversion are a critical unmet need.

Objective: This study aimed to evaluate whether an "uncertainty-normalizing" communication strategy—aimed at reinforcing the expected nature of scientific uncertainty about the COVID-19 pandemic—can reduce ambiguity aversion, and to compare its effectiveness to conventional public communication strategies aimed at promoting hope and prosocial values.

Methods: In an online factorial experiment conducted from May to June 2020, a national sample of 1497 US adults read one of five versions of an informational message describing the nature, transmission, prevention, and treatment of COVID-19; the versions varied in level of expressed scientific uncertainty and supplemental focus (ie, uncertainty-normalizing, hope-promoting, and prosocial). Participants then completed measures of cognitive, emotional, and behavioral manifestations of ambiguity aversion (ie, perceived likelihood of getting COVID-19, COVID-19 worry, and intentions for COVID-19 risk-reducing behaviors and vaccination). Analyses assessed (1) the extent to which communicating uncertainty produced ambiguity-averse psychological responses; (2) the comparative effectiveness of uncertainty-normalizing, hope-promoting, and prosocial communication strategies in reducing ambiguity-averse responses; and (3) potential moderators of the effects of alternative uncertainty communication strategies.

Results: The communication of scientific uncertainty about the COVID-19 pandemic increased perceived likelihood of getting COVID-19 and worry about COVID-19, consistent with ambiguity aversion. However, it did not affect intentions for risk-reducing behaviors or vaccination. The uncertainty-normalizing strategy reduced these aversive effects of communicating scientific uncertainty, resulting in levels of both perceived likelihood of getting COVID-19 and worry about COVID-19 that did not differ from the control message that did not communicate uncertainty. In contrast, the hope-promoting and prosocial strategies did not decrease ambiguity-averse responses to scientific uncertainty. Age and political affiliation, respectively, moderated the effects of uncertainty communication strategies on intentions for COVID-19 risk-reducing behaviors and worry about COVID-19.

Conclusions: Communicating scientific uncertainty about the COVID-19 pandemic produces ambiguity-averse cognitive and emotional, but not behavioral, responses among the general public, and an uncertainty-normalizing communication strategy reduces these responses. Normalizing uncertainty may be an effective strategy for mitigating ambiguity aversion in crisis communication efforts. More research is needed to test uncertainty-normalizing communication strategies and to elucidate the factors that moderate their effectiveness.

(J Med Internet Res 2021;23(4):e27832) doi:10.2196/27832

KEYWORDS

uncertainty; communication; ambiguity; vaccination; COVID-19

Introduction

Public health crises such as the COVID-19 pandemic pose difficult communication challenges, due in large part to the substantial scientific uncertainty surrounding the nature and management of all new and emerging health threats [1]. This uncertainty is a defining feature of public health crises [2,3] and is important to communicate in order to foster public accountability and trust [3-5]. The communication of scientific uncertainty in public health crises is also important because it promotes more realistic expectations about the benefits of risk-reducing actions and allows people to prepare for different potential outcomes. Uncertainty communication in crisis situations has thus been a central focus of expert guidance, such as the Crisis and Emergency Risk Communication (CERC) guidelines issued by the US Centers for Disease Control and Prevention. CERC guidelines recommend that communicators both acknowledge uncertainty by clarifying what is known, what is not known, and what is being done to reduce the uncertainty, and avoid promoting excess certainty about future outcomes that cannot be controlled [6].

The challenge, however, is that uncertainty can have negative psychological effects. As CERC guidelines also acknowledge, the communication of uncertainty can heighten perceptions of risk and promote fear, panic, anxiety, emotional distress, and feelings of hopelessness and helplessness, which can prevent people from taking action [6]. These aversive psychological responses have been empirically documented by a large body of research showing that uncertainty caused by a lack of reliability, credibility, or adequacy of risk information—features of information that constitute what decision theorists have termed *ambiguity* [7]—produces a set of cognitive, emotional, and behavioral responses [8-10]. These include heightened risk perceptions, pessimistic appraisals of risk-reducing actions, fear and anxiety, and avoidance of decision making. These responses, collectively known as *ambiguity aversion*, have been demonstrated in numerous decision-making settings, including health care [7,10,11].

These effects are not universal; individuals vary in their tolerance of ambiguity as well as uncertainty arising from other causes [10-12]. Furthermore, communicating uncertainty can increase, rather than decrease, individuals' confidence and trust in information when they expect such uncertainty to exist [13-15]. Nevertheless, the predominance of aversive responses to uncertainty for most individuals and situations makes the communication of scientific uncertainty in public health crises challenging [16]. Furthermore, although expert guidelines recommend adjunctive strategies, such as expressing empathy as a means of mitigating the negative psychological effects of uncertainty [5,6], empirical evidence for this or other strategies is lacking, and the optimal methods for communicating uncertainty in public health crises remain unknown [2]. Consequently, available empirical evidence suggests that in these situations scientific uncertainty is rarely communicated in a clear, explicit manner, either by experts or journalists [17-19].

One promising theory-based strategy, however, may be to normalize uncertainty; that is, to emphasize that existing uncertainty is an expected experience that does not indicate an unusual deficit in people's abilities. A leading theoretical account of ambiguity aversion, the *competence hypothesis*, suggests that ambiguity is aversive because it lowers people's perceptions of their own competence in decision making [20]. An extension of this account, the *comparative ignorance hypothesis*, posits that aversion to ambiguity about a given prospect is driven by an implicit comparison with a less ambiguous prospect or the state of mind of more knowledgeable individuals [21]. Winkler has posited that ambiguity aversion may also arise from an erroneous belief in the existence of a single ''true'' objective probability for individual events and a discomfort with not knowing this probability [22]. Together, these theories suggest that ambiguity aversion will be heightened if decision makers' perceived competence is decreased and their comparative ignorance increased (eg, when they are made aware that relevant risk information is unavailable to them but available to others). In contrast, ambiguity aversion will be diminished if perceived competence is increased and comparative ignorance decreased. Chow and colleagues obtained experimental evidence supporting these effects by showing that individual decision makers' ambiguity aversion—manifested by their reluctance to bet on an uncertain outcome—diminished when they were made aware that the risks at hand are unknown not only to them but to all individuals; that is, they are *unknowable* [20,23,24].

Normalizing uncertainty as an expected state, therefore, may be a potentially effective strategy for reducing negative psychological responses to the communication of scientific uncertainty in public health crises. The overarching objective of this study was to evaluate this possibility in the real-life context of the COVID-19 pandemic. In a previous experimental study of public responses to uncertainty about a hypothetical viral pandemic [25], we found that an uncertainty-normalizing strategy did not reduce ambiguity-averse cognitive, emotional, and behavioral responses (ie, heightened risk perceptions and worry and diminished vaccination intentions) to scientific

uncertainty about the pandemic. However, the generalizability of these findings was limited by the hypothetical nature of the study. In the current study, we addressed this limitation by evaluating the effects of an uncertainty-normalizing strategy in a real public health crisis. Its specific objective was to test whether normalizing uncertainty reduces ambiguity-averse responses, compared to commonly used strategies aimed at promoting either (1) hope or (2) prosocial values [26-28]. These alternative strategies have been broadly implemented in public information campaigns about COVID-19, and focus on mitigating hopelessness, helplessness, and stigmatization—important adverse responses to public health crises [3,5,6]. However, because these strategies do not directly target perceptions of, or responses to, uncertainty, they should be less effective in reducing ambiguity aversion than an uncertainty-normalizing strategy.

To evaluate this possibility, we conducted an online survey–based experiment comparing alternative approaches to communicating scientific uncertainty about multiple aspects of the COVID-19 pandemic, including its controllability, prognosis, and severity. The experiment tested the following hypotheses:

- Hypothesis 1 (H1). The communication of uncertainty about the COVID-19 pandemic will result in ambiguity-averse psychological responses—consisting of greater perceived likelihood of developing COVID-19, greater worry about COVID-19, and lower intentions for COVID-19 risk-reducing behaviors—compared to the noncommunication of uncertainty.
- Hypothesis 2 (H2). Ambiguity-averse responses to the communication of uncertainty about the COVID-19 pandemic will be reduced by uncertainty-normalizing language but not by either hope-promoting or prosocial language.

As an exploratory objective, we also evaluated the extent to which individual differences, including sociodemographic characteristics (ie, age, gender, and education), political affiliation, health literacy, trait-level risk aversion, trait-level ambiguity aversion, and dispositional optimism—all factors that might influence people's responses to medical uncertainty [11,12,25,29,30]—might moderate the effects of these different uncertainty communication strategies.

Methods

Study Design and Experimental Manipulation

The study was part of a larger online experiment, hosted by the internet survey vendor Qualtrics, designed to test different strategies, including language aimed at promoting hope and prosocial values, for communicating to the general public about the nature and prevention of COVID-19. This study focused specifically on strategies for communicating about uncertainty surrounding the COVID-19 pandemic. All alternative strategies were created by adding language to basic information on the nature, transmission, prevention, and treatment of COVID-19, reproduced from a public website produced by a government public health department [31]. This basic information contained no explicit communication of scientific uncertainty and served as the *control* strategy. Supplementing this basic information with additional language resulted in a total of five alternative uncertainty communication strategies, which constituted separate experimental conditions to which participants were randomly assigned: (1) control, (2) uncertainty, (3) uncertainty + uncertainty-normalizing, (4) uncertainty + hope-promoting, and (5) uncertainty + prosocial.

The *uncertainty* condition highlighted the existence of scientific uncertainty about the controllability, prognosis, and severity of the COVID-19 pandemic. The *uncertainty + uncertainty-normalizing* condition combined expressed uncertainty with language emphasizing the unknowability of these various aspects of COVID-19 and the expected nature of scientific uncertainty. The *uncertainty + hope-promoting* condition combined expressed uncertainty with language conveying optimism about future advances in knowledge and control over the pandemic. The *uncertainty + prosocial* condition combined expressed uncertainty with language encouraging awareness of obligations to other community members and concern for the collective good. The alternative uncertainty communication strategies varied in length from 940 to 1273 words; the full text of all strategies is presented in Multimedia Appendix 1.

Study Population and Recruitment

The study population consisted of a national sample of adult members (aged ≥18 years) of the US public belonging to a voluntary opt-in web survey panel professionally managed by the internet survey vendor Qualtrics. Panel members have experience and interest in completing online surveys for marketing purposes, for which they are provided modest monetary incentives. Qualtrics maintains sociodemographic and geographic data on panel members, which provides the capacity to target recruitment to prespecified quotas in order to achieve a sociodemographically diverse study sample. This study employed quotas aimed at obtaining a balanced distribution by age, gender, race, geographic region of the United States, education level (ie, ≥20% high school diploma or less), and income (ie, ≥50% annual income of US $50,000 or less), and to exclude participants who reported a current or prior diagnosis of COVID-19. To ensure data quality, we excluded participants who gave logically inconsistent responses to two screener questions about participants' attitudes toward risk-reducing behaviors or whose survey completion time was below 12 minutes—the time cut point accounting for the majority of inconsistent responses in preliminary fielding of the study.

The study was approved by the MaineHealth Institutional Review Board. The survey was fielded from May 7 to June 11, 2020; during this time, the number of total coronavirus infections in the United States increased from >1.2 million to >1.6 million, and total deaths increased from >77,200 to >98,000 [32].

Measures

After reading their randomly assigned informational vignettes, participants completed a survey questionnaire consisting of the measures summarized below.

Outcome Variables

Perceived uncertainty about COVID-19 served as the manipulation check for the study and was assessed using a 6-item scale (α=.71) developed for this study (see Multimedia Appendix 2). This measure assessed participants' perceptions of uncertainty arising from various sources (ie, probability, ambiguity, and complexity) and pertaining to various issues (ie, controllability, prognosis, and severity of the COVID-19 pandemic) raised in the experimental vignettes. Example items include "There are conflicting estimates of how long the COVID-19 pandemic will past." Likert scale response options ranged from 1 (*strongly disagree*) to 7 (*strongly agree*).

Perceived likelihood of getting COVID-19 was assessed with a single item used in prior studies [25,33,34]: "How likely does it feel that you will get COVID-19 within the next month?" Likert scale response options ranged from 1 (*not at all*) to 7 (*very*).

Worry about COVID-19 was assessed with a single item used in prior studies [25,33,34]: "How worried are you about getting COVID-19 within the next month?" Likert scale response options ranged from 1 (*not at all*) to 7 (*very*).

Intentions for COVID-19 risk-reducing behaviors was assessed by measuring participants' willingness to follow 14 recommended COVID-19 risk-reducing behaviors (eg, handwashing, avoiding social gatherings, and wearing masks) (see Multimedia Appendix 2). Likert scale response options ranged from 0 (*I am not planning to follow this guideline at all*) to 100 (*I am planning to follow this guideline fully*). Participants' responses were averaged to create a composite score (α=.95).

Intentions for vaccination was assessed with a single item used in prior studies [25,33,34]: "If a vaccine becomes available for COVID-19, how likely would you be to get vaccinated against COVID-19?" Likert scale response options ranged from 1 (*definitely would not get vaccinated*) to 7 (*definitely would get vaccinated*).

Covariates and Potential Moderators

Sociodemographic characteristics included age (ie, <30, 30-39, 40-49, 50-59, 60-69, and ≥70 years), gender, race, and political affiliation (ie, Democrat, independent or other, and Republican).

Subjective health literacy was assessed using an abbreviated, single-item version of a validated health literacy screening measure [35]: "How often do you have someone (like a family member, friend, hospital/clinic worker, or caregiver) help you read instructions, pamphlets, or other written health materials from your doctor or pharmacy?" Likert scale response options ranged from 1 (*never*) to 5 (*always*).

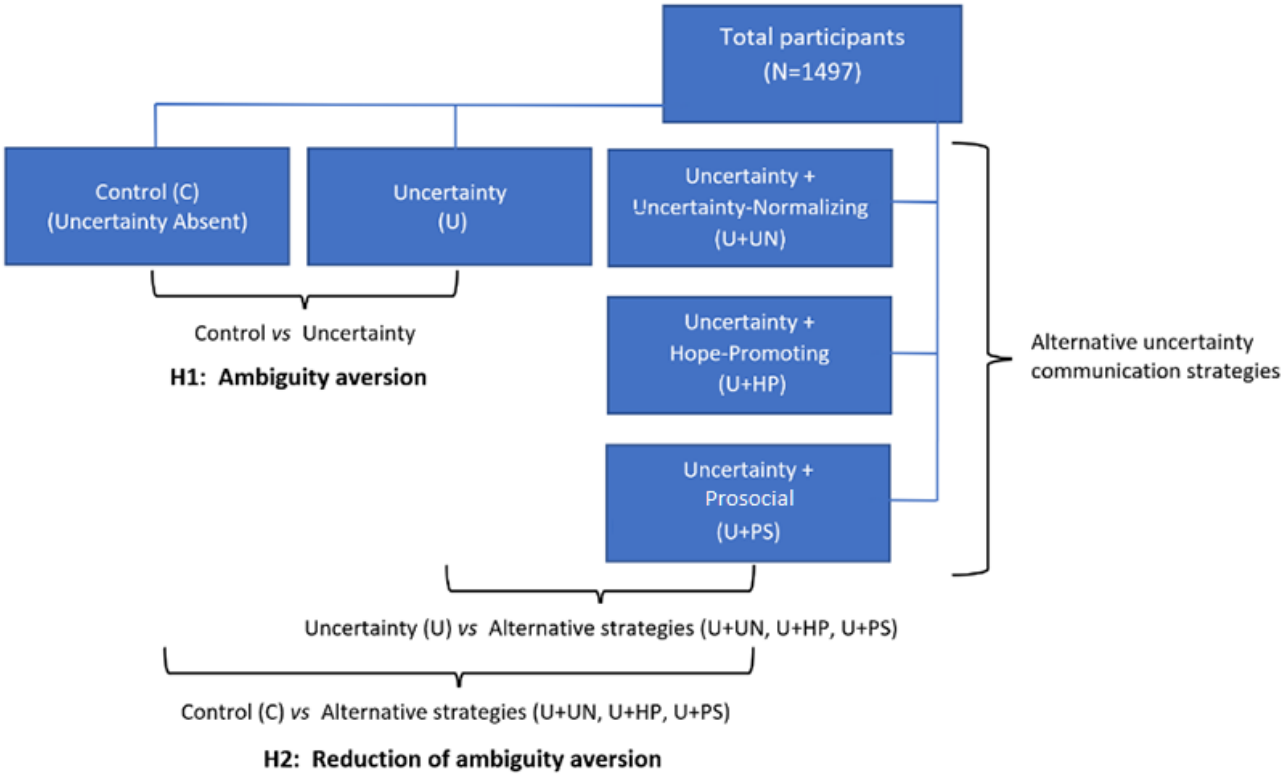
Data Analysis

To compare the effectiveness of alternative uncertainty communication strategies in reducing cognitive, emotional, and behavioral manifestations of ambiguity aversion, we fit analysis of variance models with perceived likelihood of getting COVID-19, worry about COVID-19, and intentions for both risk-reducing behaviors and vaccination as dependent variables and communication strategy as the independent variable. For each dependent variable, we used prespecified contrasts to assess the following: (1) the extent to which the communication of uncertainty produced ambiguity-averse responses, compared to the noncommunication of uncertainty (H1), and (2) the extent to which the three alternative uncertainty communication strategies (ie, uncertainty-normalizing, hope-promoting, prosocial) reduced ambiguity-averse responses (H2) (see Figure 1).

H1 was assessed by the contrast between the *uncertainty* and *control* conditions, while H2 was assessed by contrasts between each of the alternative uncertainty communication strategies and both the *uncertainty* and *control* conditions. For each contrast, we estimated the effect size by calculating Cohen *d*, which represents the standardized mean difference between two groups [36].

To explore potential moderating effects of sociodemographic characteristics (ie, age, gender, race, and political affiliation) and subjective health literacy, we fit separate models with perceived risk of COVID-19, worry about COVID-19, and intentions for both risk-reducing behaviors and vaccination as dependent variables and communication strategy as the independent variable; we entered relevant interaction terms one at a time. All analyses were conducted using SPSS, version 27.0 (IBM Corp).

Figure 1. Study design. Alternative uncertainty communication strategies and between-group comparisons. H1: Hypothesis 1; H2: Hypothesis 2.



Results

Overview

In our primary conditions, we received data from 1524 respondents. We excluded 2 respondents who gave inconsistent responses and another 25 individuals who reported current or previous COVID-19 illness, leaving a final sample of 1497 respondents (see Table 1). Data were assumed to be missing at random; thus, we utilized a listwise deletion strategy for participants with missing data on any of the outcome measures.

On average, participants took 27.82 (SD 34.59) minutes to complete the study. There were no significant between-group differences in time of completion of the experimental task ($F_{7,2386}$=0.479; P=.85), suggesting that the cognitive effort required by the task was similar across conditions.

Table 1. Sample population characteristics.

Characteristic	Value (N=1497), n (%)
Age (years)	
<30	306 (20.4)
30-39	229 (15.3)
40-49	209 (14.0)
50-59	183 (12.2)
60-69	227 (15.2)
≥70	343 (22.9)
Gender	
Male	743 (49.6)
Female	748 (50.0)
Other or prefer not to say	6 (0.4)
Race	
White	1003 (67.0)
Black or African American	178 (11.9)
Asian	147 (9.8)
Multiracial or other	169 (11.3)
Education	
Less than high school	264 (17.6)
High school graduate	242 (16.2)
Some college or trade school	396 (26.5)
College graduate or higher	595 (39.7)
Income (US $)	
0-24,999	394 (26.3)
25,000-49,999	371 (24.8)
50,000-99,999	370 (24.7)
100,000-149,999	235 (15.7)
≥150,000	127 (8.5)
Political affiliation	
Democrat	540 (36.1)
Republican	438 (29.3)
Independent or other[a]	519 (34.7)

[a]Includes other third party or no party affiliation.

Manipulation Check

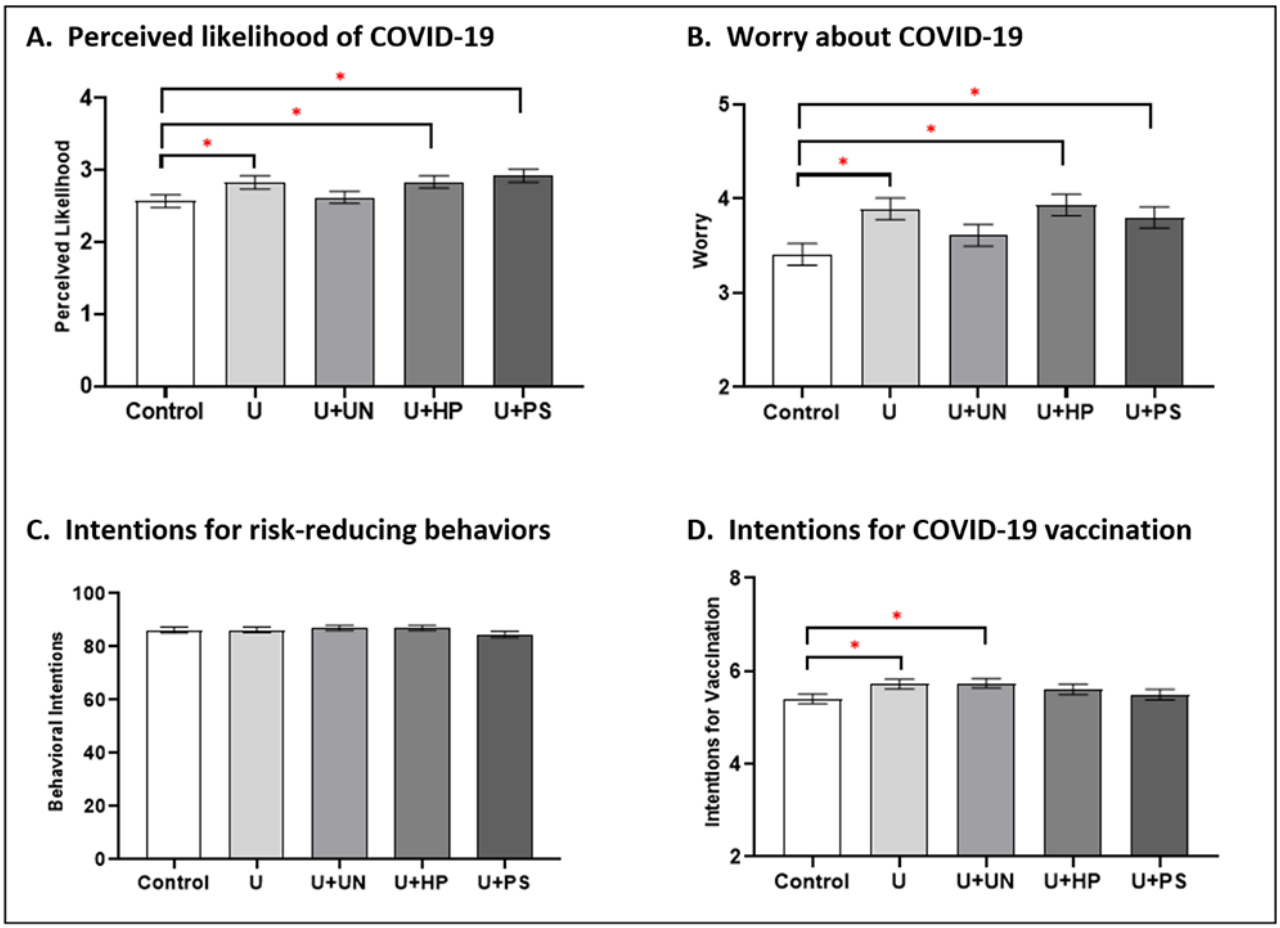
Supporting the intended effect of the experimental manipulation, perceived uncertainty about COVID-19 was significantly higher in all experimental conditions containing uncertainty ($F_{4,1492}$=3.52; η^2=0.009; *P*=.007)—that is, *uncertainty* (*d*=–0.28; *P*=.001), *uncertainty + uncertainty-normalizing* (*d*=–0.23; *P*=.006), *uncertainty + hope-promoting* (*d*=–0.20; *P*=.01), and *uncertainty + prosocial* (*d*=–0.22; *P*=.007)—compared to the *control* condition containing no uncertainty. These differences were in the small effect-size range.

Perceived Likelihood of Getting COVID-19

Consistent with an ambiguity-averse cognitive response to the communication of uncertainty (H1), perceived likelihood of getting COVID-19 was significantly higher in the *uncertainty* condition than in the *control* condition ($F_{4,1492}$=2.95; η^2=0.008; *P*=.02) (see Figure 2A). Supporting the effectiveness of the uncertainty-normalizing strategy in reducing ambiguity aversion (H2), perceived likelihood of getting COVID-19 was not significantly different for the *uncertainty + uncertainty-normalizing* condition compared to the *control* condition (*d*=–0.04; *P*=.66). However, this ambiguity aversion–reducing effect was not seen for the *hope-promoting*

(*d*=–0.18; *P*=.03) or *prosocial* (*d*=–0.23; *P*=.005) communication strategies; perceived likelihood remained significantly higher for these conditions than for the *control* condition.

Figure 2. Effects of uncertainty and uncertainty communication strategies on cognitive, emotional, and behavioral manifestations of ambiguity aversion. Asterisks indicate statistically significant pairwise differences (*P*<.05); error bars indicate standard error. U: uncertainty; U+HP: uncertainty + hope-promoting; U+PS: uncertainty + prosocial; U+UN: uncertainty + uncertainty-normalizing.



Worry About COVID-19

Consistent with an ambiguity-averse emotional response to the communication of uncertainty (H1), worry about COVID-19 was significantly higher in the *uncertainty* condition than in the *control* condition ($F_{4,1492}$=3.65; η^2=0.01; *P*=.006) (see Figure 2B). Supporting the effectiveness of the uncertainty-normalizing strategy in reducing ambiguity aversion (H2), worry was not significantly different for the *uncertainty + uncertainty-normalizing* condition compared to the *control* condition (*d*=–0.10; *P*=.21). However, this ambiguity aversion–reducing effect was not seen for the *hope-promoting* (*d*=–0.27; *P*=.001) or *prosocial* (*d*=–0.20; *P*=.02) communication strategies; worry remained significantly higher for these conditions than for the *control* condition.

Intentions for COVID-19 Risk-Reducing Behaviors and Vaccination

Inconsistent with an ambiguity-averse behavioral response to uncertainty, intentions regarding COVID-19 risk-reducing behaviors (η^2=0.002; *P*=.49) and vaccination (η^2=0.005; *P*=.14) showed no significant differences between any of the experimental conditions (see Figure 2C and D). However, prespecified contrasts revealed higher vaccination intentions in both the *uncertainty* (*d*=–0.17; *P*=.04) and the *uncertainty + uncertainty-normalizing* (*d*=–0.18; *P*=.03) conditions compared to the *control* condition, suggesting that the communication of uncertainty itself motivated vaccination intentions and that the addition of uncertainty-normalizing language preserved this motivation (see Figure 2D).

Moderating Effects

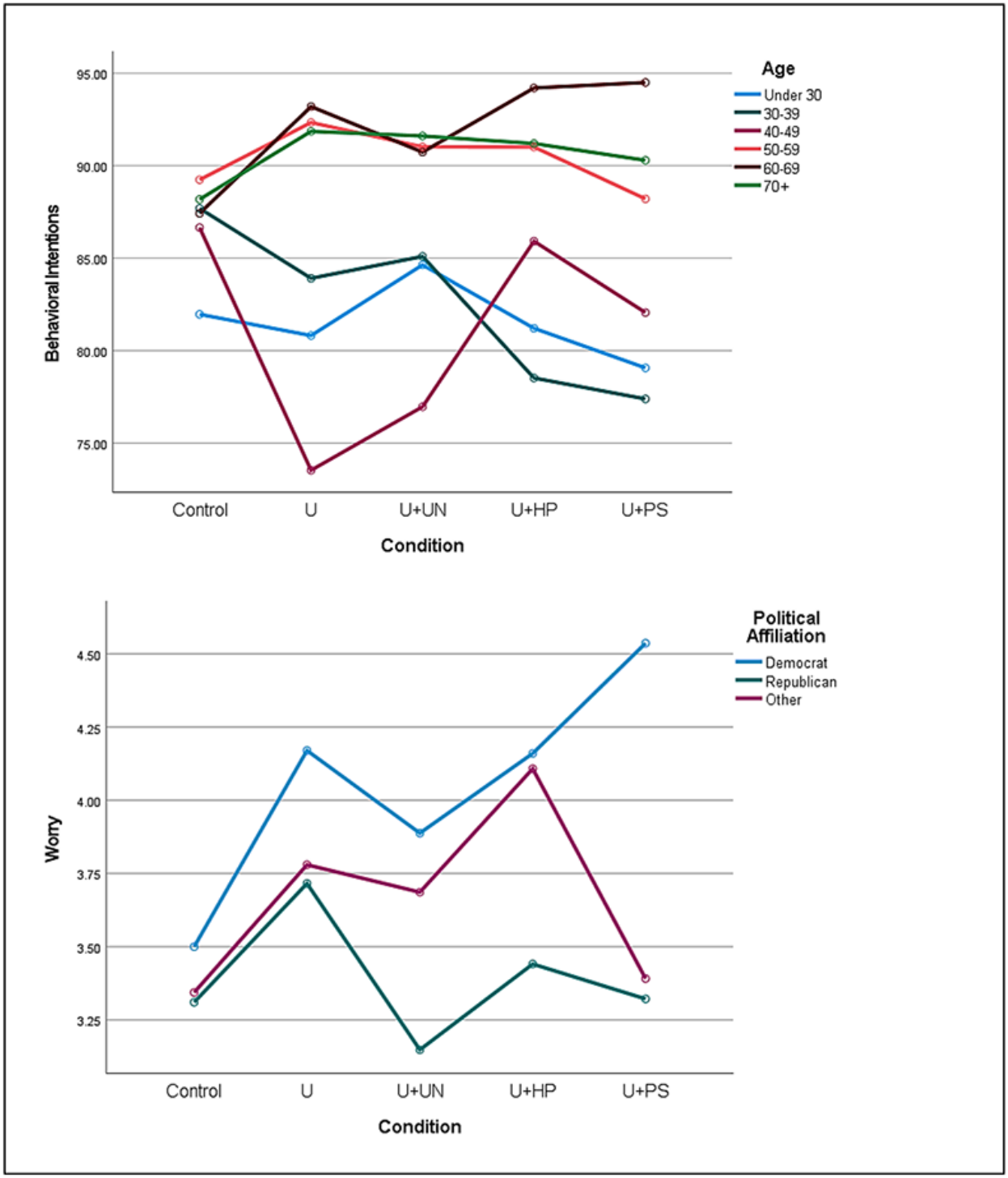
Two factors, age and political affiliation, were found to moderate the effects of uncertainty communication strategy on different ambiguity-averse responses to the communication of uncertainty. Age moderated the effect of communication strategy on intentions for COVID-19 risk-reducing behaviors ($F_{20,1445}$=1.86; $\eta\rho^2$=0.025; *P*=.01), such that older participants (ie, aged 50 years and older) generally reported higher intentions in all of the supplementary uncertainty communication conditions compared to the *control* condition, while younger participants (ie, less than 50 years of age) generally reported lower intentions (see Figure 3, top plot).

Political affiliation showed a weaker interaction with uncertainty communication strategy on worry about COVID-19 ($\eta\rho^2$=0.010; *P*=.06), such that self-reported Republicans had lower worry in

the *uncertainty + uncertainty-normalizing* condition compared to the *control* condition, while self-reported Democrats and independents had higher worry (see Figure 3, bottom plot). In other words, the uncertainty-normalizing strategy reduced ambiguity aversion to a greater extent for Republicans than for Democrats. Democrats also had higher worry in the *prosocial* condition compared to the *control* condition, while Republicans and independents had lower worry.

No significant moderating effects were noted for other sociodemographic factors or subjective health literacy.

Figure 3. Moderators of the effects of uncertainty communication strategy on manifestations of ambiguity aversion: age and political affiliation. U: uncertainty; U+HP: uncertainty + hope-promoting; U+PS: uncertainty + prosocial; U+UN: uncertainty + uncertainty-normalizing.



Discussion

This experimental study evaluated the comparative effectiveness of different communication strategies in reducing ambiguity-averse cognitive, emotional, and behavioral responses to uncertainty in information about the COVID-19 pandemic. We believe its findings have several implications for future efforts to understand and improve the communication of uncertainty in public health crises.

Consistent with predictions, a strategy aimed at normalizing uncertainty as an expected state of affairs was effective in reducing at least some aversive psychological responses to the communication of uncertainty, whereas widely used alternative strategies aimed at promoting hope and prosocial values had no such effect. A major barrier to open, explicit communication of the uncertainties that inevitably exist during public health crises is a real concern about exacerbating both the perception of vulnerability as well as feelings of fear and panic among the general public [2,3,5,6]. Our findings suggest, however, that language aimed at normalizing these uncertainties can reduce aversive cognitive and emotional responses to them. When uncertainty-normalizing language was added to a message that communicated scientific uncertainty about the COVID-19 pandemic, levels of COVID-19 risk perceptions and worry did not differ from those produced by a message that did not communicate scientific uncertainty. In other words, uncertainty-normalizing language neutralized ambiguity aversion. The overall size of this effect was relatively small, and although uncertainty-normalizing language resulted in lower levels of COVID-19 risk perceptions and worry than those produced by a message that communicated uncertainty alone, this difference was not statistically significant. Nevertheless, even small effects in reducing aversive psychological responses to uncertainty may be beneficial in large-scale efforts to communicate with the general public about health crises. If our findings can be replicated and validated, they suggest a promising new approach to inoculating people against the vulnerability and fear that typically accompany the communication of uncertainty in these situations.

Contrary to predictions, uncertainty-normalizing language had no effect on intentions for COVID-19 risk-reducing behaviors or vaccination. Notably, however, the communication of uncertainty itself also had no effect; it neither decreased nor increased behavioral intentions. In other words, ambiguity aversion in this study was manifest cognitively and emotionally, but not behaviorally. This pattern may be attributable to several factors. Potential negative effects of scientific uncertainty about COVID-19 on intentions for risk-reducing behaviors may have been attenuated by the legally mandated nature of several of these behaviors (eg, mandatory quarantines and regulations requiring social distancing and use of masks). Furthermore, uncertainties about the controllability, prognosis, and severity of the COVID-19 pandemic may have mixed, opposing effects on behavioral intentions. They may decrease intentions by fostering skepticism about the benefits of risk-reducing behaviors, thereby promoting a tendency toward inaction, which is consistent with ambiguity aversion. At the same time, these different uncertainties may also increase intentions by promoting fear about the consequences of avoiding risk-reducing behaviors, thereby promoting a tendency toward action, which is consistent with ambiguity tolerance. More research is needed to understand the factors that moderate people's behavioral responses to different uncertainties and favor either inaction or action.

Our study sheds light on at least some of these factors. Age moderated the effect of communication strategy on intentions for COVID-19 risk-reducing behaviors. For adults over 50 years of age, all active uncertainty communication strategies resulted in higher behavioral intentions compared to the control (ie, no uncertainty) strategy, while for younger adults, all active uncertainty communication strategies resulted in lower intentions. This moderating effect may be attributable to several factors. Older adults have been identified as being at higher risk for complications of COVID-19 and may, thus, be more motivated to take action in the face of uncertainty. Political affiliation also appeared to partially moderate the effect of communication strategy on worry about COVID-19; the uncertainty-normalizing strategy was more effective in reducing ambiguity aversion for Republicans than for Democrats. This moderating effect is intriguing, and its causes are unclear. Political party affiliation has been shown to influence COVID-19 risk perceptions and intentions for risk-reducing behaviors; Republicans generally demonstrate lower risk perceptions and behavioral intentions than Democrats [37-39]. The differential worry-reducing effect of uncertainty normalization for Republicans versus Democrats suggests the existence of respectively opposing propensities toward either minimizing feelings of vulnerability in the face of uncertainty (for Republicans) or else maximizing them (for Democrats). Political affiliation may be a proxy for numerous factors, including ideologies, values, and worldviews, that may predispose people to more optimistic or pessimistic appraisals of uncertain threats [40]. More research is needed to elucidate how these and other factors produce differential responses to different uncertainty communication strategies and to identify other important moderators.

Our study had several limitations that qualify its findings. The sample consisted of web survey panel members who, by virtue of their willingness to participate regularly in market research and other studies, may not be representative of the general population. However, the sample was large and both geographically and sociodemographically diverse, providing support for the external validity of its findings. Nevertheless, more research is needed to assess the reproducibility of our findings and their generalizability to other populations. The alternative uncertainty communication strategies tested in this study varied modestly in length; however, the absence of significant between-group differences in study completion time suggests that our findings are not attributable solely to differences in cognitive burden or effort. The uncertainty-normalizing language tested in this study was also novel and unvalidated, and it may have contained unintended hope-promoting or prosocial messages. We believe the significant between-group differences observed in our study argue against this possibility; however, further research is needed to ascertain the precision and efficacy of our uncertainty-normalizing language in conveying the normal, expected nature of uncertainty. We also conducted multiple exploratory analyses to identify potential moderators of the effects of different communication strategies; the two significant interactions identified could thus have resulted from chance, and further research is needed to confirm these findings. Finally, some key constructs (ie, perceived uncertainty about COVID-19 and intentions for COVID-19 risk-reducing behaviors) were assessed using new measures that have yet to be validated. Other constructs were assessed using existing single-item measures;

however, similar measures have been used in prior studies and shown to have predictive validity [25,30,41-43].

In spite of these limitations, this study yields important new insights on the nature and extent of aversion to ambiguity in the context of the COVID-19 pandemic and on a new and potentially effective uncertainty communication strategy that can minimize this aversion. It remains for future research to confirm our findings and to develop more effective strategies for communicating the unavoidable and irreducible scientific uncertainties that complicate all public health crises.

Acknowledgments

This work was supported by a Pilot Project Program Award to PH from the Northern New England Clinical & Translational Research Network (1U54GM115516-01). The funding agreement ensured the authors' independence in designing the study, interpreting the data, and in writing and publishing the report.

Authors' Contributions

PH conceived and designed the study, developed the experimental conditions and measures, conducted the experiments, analyzed the data, and drafted and revised the manuscript. ES, AS, AT, CL, LW, AF, and ND contributed to the development of the experimental conditions and measures, provided input on data analysis, and revised the manuscript. ES also programmed the survey and analyzed the data. All authors approved the final version.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Experimental conditions.
[DOCX File , 30 KB - jmir_v23i4e27832_app1.docx]

Multimedia Appendix 2
Measures developed in this study.
[DOCX File , 16 KB - jmir_v23i4e27832_app2.docx]

References

1. Freimuth VS. Order out of chaos: The self-organization of communication following the anthrax attacks. Health Commun 2006;20(2):141-148. [doi: 10.1207/s15327027hc2002_5] [Medline: 16965251]
2. Liu BF, Bartz L, Duke N. Communicating crisis uncertainty: A review of the knowledge gaps. Public Relat Rev 2016 Sep;42(3):479-487. [doi: 10.1016/j.pubrev.2016.03.003]
3. Seeger MW. Best practices in crisis communication: An expert panel process. J Appl Commun Res 2006 Aug;34(3):232-244. [doi: 10.1080/00909880600769944]
4. Berg JW. All for one and one for all: Informed consent and public health. Houst Law Rev 2012;50(1):1-40 [FREE Full text]
5. Covello VT. Best practices in public health risk and crisis communication. J Health Commun 2003;8 Suppl 1:5-8; discussion 148. [doi: 10.1080/713851971] [Medline: 14692565]
6. Crisis and Emergency Risk Communication (CERC) manual. US Centers for Disease Control and Prevention. 2019. URL: https://emergency.cdc.gov/cerc/manual/index.asp [accessed 2021-04-05]
7. Ellsberg D. Risk, ambiguity, and the savage axioms. Q J Econ 1961 Nov;75(4):643. [doi: 10.2307/1884324]
8. Machina MJ, Siniscalchi M. Ambiguity and ambiguity aversion. In: Machina MJ, Viscusi WK, editors. Handbook of the Economics of Risk and Uncertainty. Oxford, UK: North-Holland; 2014:729-807.
9. Golman R, Gurney N, Loewenstein G. Information gaps for risk and ambiguity. Psychol Rev 2021 Jan;128(1):86-103. [doi: 10.1037/rev0000252] [Medline: 32757571]
10. Camerer C, Weber M. Recent developments in modeling preferences: Uncertainty and ambiguity. J Risk Uncertain 1992 Oct;5(4):325-370. [doi: 10.1007/bf00122575]
11. Han PKJ, Reeve BB, Moser RP, Klein WMP. Aversion to ambiguity regarding medical tests and treatments: Measurement, prevalence, and relationship to sociodemographic factors. J Health Commun 2009 Sep;14(6):556-572 [FREE Full text] [doi: 10.1080/10810730903089630] [Medline: 19731127]
12. Hillen MA, Gutheil CM, Strout TD, Smets EM, Han PK. Tolerance of uncertainty: Conceptual analysis, integrative model, and implications for healthcare. Soc Sci Med 2017 May;180:62-75 [FREE Full text] [doi: 10.1016/j.socscimed.2017.03.024] [Medline: 28324792]
13. Du N, Budescu DV, Shelly MK, Omer TC. The appeal of vague financial forecasts. Organ Behav Hum Decis Process 2011 Mar;114(2):179-189. [doi: 10.1016/j.obhdp.2010.10.005]

14. Morss R, Demuth J, Lazo J. Communicating uncertainty in weather forecasts: A survey of the US public. Weather Forecast 2008;23(5):974-991. [doi: 10.1175/2008waf2007088.1]
15. Wegwarth O, Wagner G, Spies C, Hertwig R. Assessment of German public attitudes toward health communications with varying degrees of scientific uncertainty regarding COVID-19. JAMA Netw Open 2020 Dec 01;3(12):e2032335 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.32335] [Medline: 33301021]
16. Frewer L. The public and effective risk communication. Toxicol Lett 2004 Apr 01;149(1-3):391-397. [doi: 10.1016/j.toxlet.2003.12.049] [Medline: 15093286]
17. Goodall C, Sabo J, Cline R, Egbert N. Threat, efficacy, and uncertainty in the first 5 months of national print and electronic news coverage of the H1N1 virus. J Health Commun 2012;17(3):338-355. [doi: 10.1080/10810730.2011.626499] [Medline: 22188164]
18. Singer E. A question of accuracy: How journalists and scientists report research on hazards. J Commun 1990 Dec;40(4):102-116 [FREE Full text] [doi: 10.1111/j.1460-2466.1990.tb02284.x]
19. Pellechia MG. Trends in science coverage: A content analysis of three US newspapers. Public Underst Sci 2016 Dec 21;6(1):49-68. [doi: 10.1088/0963-6625/6/1/004]
20. Heath C, Tversky A. Preference and belief: Ambiguity and competence in choice under uncertainty. J Risk Uncertain 1991 Jan;4(1):5-28. [doi: 10.1007/BF00057884]
21. Fox CR, Tversky A. Ambiguity aversion and comparative ignorance. Q J Econ 1995 Aug;110(3):585-603 [FREE Full text] [doi: 10.2307/2946693]
22. Winkler RL. Ambiguity, probability, preference, and decision analysis. J Risk Uncertain 1991 Jul;4(3):285-297. [doi: 10.1007/BF00114158]
23. Chow CC, Sarin RK. Comparative ignorance and the Ellsberg Paradox. J Risk Uncertain 2001;22:129-139. [doi: 10.1023/A:1011157509006]
24. Chow CC, Sarin RK. Known, unknown, and unknowable uncertainties. Theory Decis 2002;52:127-138. [doi: 10.1023/A:1015544715608]
25. Han P, Zikmund-Fisher B, Duarte C, Knaus M, Black A, Scherer A, et al. Communication of scientific uncertainty about a novel pandemic health threat: Ambiguity aversion and its mechanisms. J Health Commun 2018;23(5):435-444 [FREE Full text] [doi: 10.1080/10810730.2018.1461961] [Medline: 29648962]
26. Keshavarz S, Coventry K, Fleming P. Relative deprivation and hope: Predictors of risk behavior. J Gambl Stud 2020 Dec 16:1 [FREE Full text] [doi: 10.1007/s10899-020-09989-4] [Medline: 33326060]
27. Heffner J, Vives M, FeldmanHall O. Emotional responses to prosocial messages increase willingness to self-isolate during the COVID-19 pandemic. Pers Individ Dif 2021 Feb 15;170:110420 [FREE Full text] [doi: 10.1016/j.paid.2020.110420] [Medline: 33082614]
28. Jung H, Albarracín D. Concerns for others increases the likelihood of vaccination against influenza and COVID-19 more in sparsely rather than densely populated areas. Proc Natl Acad Sci U S A 2021 Jan 05;118(1):1-8 [FREE Full text] [doi: 10.1073/pnas.2007538118] [Medline: 33443183]
29. Strout T, Hillen M, Gutheil C, Anderson E, Hutchinson R, Ward H, et al. Tolerance of uncertainty: A systematic review of health and healthcare-related outcomes. Patient Educ Couns 2018 Sep;101(9):1518-1537. [doi: 10.1016/j.pec.2018.03.030] [Medline: 29655876]
30. Han P, Klein W, Lehman T, Killam B, Massett H, Freedman A. Communication of uncertainty regarding individualized cancer risk estimates: Effects and influential factors. Med Decis Making 2011;31(2):354-366 [FREE Full text] [doi: 10.1177/0272989X10371830] [Medline: 20671211]
31. COVID-19: General information. Maine Center for Disease Control & Prevention. 2020. URL: https://www.maine.gov/dhhs/mecdc/infectious-disease/epi/airborne/coronavirus/general-information.shtml [accessed 2020-07-01]
32. Coronavirus world map: Tracking the global outbreak. The New York Times. 2020. URL: https://www.nytimes.com/interactive/2020/world/coronavirus-maps.html [accessed 2020-06-01]
33. Zikmund-Fisher BJ, Scherer AM, Knaus M, Das E, Fagerlin A. Discussion of average versus extreme case severity in pandemic risk communications. Emerg Infect Dis 2017 Apr;23(4):706-708 [FREE Full text] [doi: 10.3201/eid2304.161600] [Medline: 28322691]
34. Fagerlin A, Valley TS, Scherer AM, Knaus M, Das E, Zikmund-Fisher BJ. Communicating infectious disease prevalence through graphics: Results from an international survey. Vaccine 2017 Jul 13;35(32):4041-4047 [FREE Full text] [doi: 10.1016/j.vaccine.2017.05.048] [Medline: 28647168]
35. Chew LD, Griffin JM, Partin MR, Noorbaloochi S, Grill JP, Snyder A, et al. Validation of screening questions for limited health literacy in a large VA outpatient population. J Gen Intern Med 2008 May;23(5):561-566 [FREE Full text] [doi: 10.1007/s11606-008-0520-5] [Medline: 18335281]
36. Cohen J. Statistical Power Analysis for the Behavioral Sciences. 2nd edition. Mahwah, NJ: Lawrence Erlbaum Associates; 1988.
37. Bruine de Bruin W, Saw H, Goldman D. Political polarization in US residents' COVID-19 risk perceptions, policy preferences, and protective behaviors. J Risk Uncertain 2020 Nov 18:1-18 [FREE Full text] [doi: 10.1007/s11166-020-09336-3] [Medline: 33223612]

38. van Holm EJ, Monaghan J, Shahar DC, Messina J, Surprenant C. The impact of political ideology on concern and behavior during COVID-19. SSRN 2020 Apr:1. [doi: 10.2139/ssrn.3573224]
39. Kushner Gadarian S, Goodman SW, Pepinsky TB. Partisanship, health behavior, and policy attitudes in the early stages of the COVID-19 pandemic. SSRN 2020 Mar:1. [doi: 10.2139/ssrn.3562796]
40. Jost JT. Ideological asymmetries and the essence of political psychology. Polit Psychol 2017 Mar 15;38(2):167-208. [doi: 10.1111/pops.12407]
41. Diefenbach M, Miller S, Daly M. Specific worry about breast cancer predicts mammography use in women at risk for breast and ovarian cancer. Health Psychol 1999 Sep;18(5):532-536. [doi: 10.1037//0278-6133.18.5.532] [Medline: 10519469]
42. Lipkus I, Iden D, Terrenoire J, Feaganes J. Relationships among breast cancer concern, risk perceptions, and interest in genetic testing for breast cancer susceptibility among African-American women with and without a family history of breast cancer. Cancer Epidemiol Biomarkers Prev 1999 Jun;8(6):533-539 [FREE Full text] [Medline: 10385144]
43. Han P, Klein W, Killam B, Lehman T, Massett H, Freedman A. Representing randomness in the communication of individualized cancer risk estimates: Effects on cancer risk perceptions, worry, and subjective uncertainty about risk. Patient Educ Couns 2012 Jan;86(1):106-113 [FREE Full text] [doi: 10.1016/j.pec.2011.01.033] [Medline: 21377311]

Abbreviations

CERC: Crisis and Emergency Risk Communication
H1: Hypothesis 1
H2: Hypothesis 2

Edited by C Basch; submitted 08.02.21; peer-reviewed by J Wijesundara, L Grant; comments to author 21.03.21; revised version received 21.03.21; accepted 21.03.21; published 22.04.21.

Please cite as:
Han PKJ, Scharnetzki E, Scherer AM, Thorpe A, Lary C, Waterston LB, Fagerlin A, Dieckmann NF
Communicating Scientific Uncertainty About the COVID-19 Pandemic: Online Experimental Study of an Uncertainty-Normalizing Strategy
J Med Internet Res 2021;23(4):e27832
URL: https://www.jmir.org/2021/4/e27832
doi:10.2196/27832
PMID:33769947

Original Paper

Machine Learning–Based Prediction of Growth in Confirmed COVID-19 Infection Cases in 114 Countries Using Metrics of Nonpharmaceutical Interventions and Cultural Dimensions: Model Development and Validation

Arnold YS Yeung[1,2], MSc; Francois Roewer-Despres[1,2], MSc; Laura Rosella[3], MHSc, PhD; Frank Rudzicz[1,2,4], MEng, PhD

[1]Department of Computer Science, University of Toronto, Toronto, ON, Canada
[2]Vector Institute for Artificial Intelligence, Toronto, ON, Canada
[3]Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada
[4]Unity Health Toronto, Toronto, ON, Canada

Corresponding Author:
Arnold YS Yeung, MSc
Department of Computer Science
University of Toronto
27 King's College Cir
Toronto, ON, M5S 3H7
Canada
Phone: 1 416 978 2011
Email: arnoldyeung@cs.toronto.edu

Abstract

Background: National governments worldwide have implemented nonpharmaceutical interventions to control the COVID-19 pandemic and mitigate its effects.

Objective: The aim of this study was to investigate the prediction of future daily national confirmed COVID-19 infection growth—the percentage change in total cumulative cases—across 14 days for 114 countries using nonpharmaceutical intervention metrics and cultural dimension metrics, which are indicative of specific national sociocultural norms.

Methods: We combined the Oxford COVID-19 Government Response Tracker data set, Hofstede cultural dimensions, and daily reported COVID-19 infection case numbers to train and evaluate five non–time series machine learning models in predicting confirmed infection growth. We used three validation methods—in-distribution, out-of-distribution, and country-based cross-validation—for the evaluation, each of which was applicable to a different use case of the models.

Results: Our results demonstrate high R^2 values between the labels and predictions for the in-distribution method (0.959) and moderate R^2 values for the out-of-distribution and country-based cross-validation methods (0.513 and 0.574, respectively) using random forest and adaptive boosting (AdaBoost) regression. Although these models may be used to predict confirmed infection growth, the differing accuracies obtained from the three tasks suggest a strong influence of the use case.

Conclusions: This work provides new considerations in using machine learning techniques with nonpharmaceutical interventions and cultural dimensions as metrics to predict the national growth of confirmed COVID-19 infections.

(J Med Internet Res 2021;23(4):e26628) doi:10.2196/26628

KEYWORDS

COVID-19; machine learning; nonpharmaceutical interventions; cultural dimensions; random forest; AdaBoost; forecast; informatics; epidemiology; artificial intelligence

Introduction

Background

In response to the COVID-19 pandemic, national governments have implemented nonpharmaceutical interventions (NPIs) to control and reduce the spread in their respective countries [1-5]. Indeed, early reports suggested the potential effectiveness of the implementation of NPIs to reduce the transmission of COVID-19 [2,4-8] and other infectious diseases [9-11]. Many epidemiological models that forecast future infection numbers have therefore suggested the role of NPIs in reducing infection rates [2,4,7,12], which can aid the implementation of national strategies and policy decision-making. Recent research incorporates publicly available data with machine learning for use cases such as reported infection case number forecasting [13-16]. Although these studies have used various features, such as existing infection statistics [13], weather [14], media and internet activity [15], and lockdown type [16], to predict infection case numbers, no study has yet examined the combination of NPIs and cultural dimensions in predicting infection growth. In this paper, we include the implementation of NPIs at the national level as features (ie, independent variables) in predicting the national growth of the number of confirmed infection cases. Based on recent studies that identify cultural dimensions as having influence in the effectiveness of NPIs [17-19], we also incorporate cultural dimensions as features. Prior work has focused on NPI variations in different regions of specific countries [2,5,6,20,21]. In contrast, our study involves 114 countries.

Various metrics may provide different perspectives and insights on the pandemic. In this study, we focus on one: confirmed infection growth (CIG), which we define as the 14-day growth in the cumulative number of reported infection cases. Other common metrics to measure the transmission rates of an infectious disease are the basic reproduction number, R_0, which measures the expected number of direct secondary infections generated by a single primary infection when the entire population is susceptible [3,22] and the effective reproduction number, R_t [2], which accounts for immunity within a specified population. Although such metrics are typically used by epidemiologists as measures of the transmission of an infectious disease, these metrics are dependent on estimation model structures and assumptions; therefore, they are application-specific and can potentially be misapplied [22]. Furthermore, the public may be less familiar with such metrics as opposed to more practical and observable metrics, such as the absolute or relative change in cumulative reported cases.

Related Work

Mathematical modelling of the transmission of infectious disease is a common method to simulate infection trajectories. A common technique for epidemics is the susceptible-infected-recovered (SIR) model, which separates the population into three subpopulations (susceptible, infected, and recovered) and iteratively models the interaction and shift between these subpopulations, which change throughout the epidemic [23,24]. Variations of this model have since been introduced to reflect other dynamics expected of the spread of infectious diseases [25-27]. These variations of the SIR model have also been applied to the ongoing COVID-19 pandemic [28-31].

The recent increase in data availability through advances in the internet and other data sources has enabled the inclusion of other factors in epidemiology modelling [32,33]. Since the early months of the COVID-19 pandemic, Johns Hopkins University has managed the COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE), which aggregates daily statistics of reported infection and mortality numbers across multiple countries [34]. Data sets related to governmental policies and NPIs have also been released publicly on the web. Notable COVID-19–related data sets include the Oxford COVID-19 Government Response Tracker (OxCGRT) [1], Complexity Science Hub COVID-19 Control Strategies List [35], CoronaNet [36], county-level socioeconomic data for predictive modeling of epidemiological effects (US-specific) [37], and CAN-NPI (Canada-specific) [20]. Additional COVID-19 data sets relate to social media activity [38-41], scientific publications [42-44], population mobility [45-48], and medical images [49-52]. In this work, we focus on the use of NPIs in the forecast of COVID-19 infection growth. Specifically, we selected the CSSE data set for infection statistics and the OxCGRT for NPI features due to their global comprehensiveness. Although features can be extracted from additional COVID-19 data sets in our models, we limited the scope of this study to COVID-19 NPI features.

Recent research has also linked the effect of cultural dimensions in responses to the COVID-19 pandemic. Studies suggest that cultural dimensions may affect individual and collective behavior [53-57] and the effectiveness of NPIs [17-19], and that cultural dimensions should be considered when implementing NPIs [17]. Although these studies identify the importance of cultural dimensions in controlling the COVID-19 pandemic, to our knowledge, this work is the first to complement cultural dimensions with NPIs to forecast future COVID-19 infection growth. We recognize that various cultural dimension models exist, such as the six Hofstede cultural dimensions [58], Global Leadership and Organizational Effectiveness (GLOBE) [59], and the Cultural Value Scale (CVSCALE) [60], and that each model has their advocates and criticisms [61]. In this work, we selected the 2015 edition of the Hofstede model [62] due to the relevance of its cultural dimensions in the mentioned studies [17-19,55-57].

Machine learning has been used in applications to combat the COVID-19 pandemic, such as in patient monitoring and genome sequencing [63-66]. Recent studies have also used various statistical and machine learning techniques for short-term forecasting of infection rates for the COVID-19 pandemic [13,15,16,30,33] using reported transmission and mortality statistics, population geographical movement data, and media activity. Pinter et al [13] combined multilayer perceptron with fuzzy inference to predict reported infection and mortality numbers in Hungary with only case number features from May to August 2020. Although reported infection and mortality case numbers aligned with their predictions for May 2020, comparison of the predictions with actual reported numbers from June to August 2020 suggest inaccuracies. Liu et al [15]

used internet and news activity predictors within a clustering machine learning model for reported COVID-19 case numbers within Chinese provinces. However, the predictors used within this work are heavily limited to Chinese populations (eg, Baidu search and mobility data, Chinese media sources), and they only predicted cases 2 days ahead. Malki et al [14] used weather, temperature, and humidity features as predictors for COVID-19 mortality rates in regressor machine learning models. Their results suggest that these predictors are relevant for COVID-19 mortality rate modelling. Similar to our work, Saba et al [16] implemented multiple machine learning models to forecast COVID-19 cases based on NPI implementation. However, their work differs in that it only includes lockdown type as an NPI feature (and does not consider cultural dimensions), the study is limited to 9 countries, and the reported case numbers are predicted instead of the change in case numbers. To our knowledge, no other studies have combined NPI and cultural dimension features to predict the growth of reported COVID-19 cases using machine learning. Furthermore, only this work forecasts COVID-19 growth as a measure of CIG (ie, 14-day growth in the cumulative number of reported cases at a national level) across 114 countries via three validation methods, each of which is applicable to a different use case of the model.

Description of the Study

Due to its direct inference from the number of reported cases, the CIG is a verifiable metric, and it may have a greater impact on the public perception of the magnitude of the COVID-19 pandemic than the actual transmission rate. In this work, CIG reflects the growth in the total number of reported cases within a country in 14 days relative to the total number of previously reported infections, including recoveries and mortalities. We selected 14 days as a suitable period for measuring the change in reported cases because of the expected incubation period of COVID-19. Researchers have found that 97.5% of reported patients with identifiable symptoms developed symptoms within 11.5 days, and 99% developed symptoms within 14 days [67]. We therefore propose the use of 14 days, or 2 weeks, as a suitable period to observe changes in reported case numbers occurring after the implementation of NPIs. A shorter period may lead to the misleading inclusion of reported infections that occurred prior to the implementation of an NPI. Results for a longer period may be misleading as well, given the higher likelihood of change in NPIs within this period that will not be accounted for during prediction. We propose that the CIG over 14 days is a suitable metric that enables inference of the effect of NPIs while being within a relevant period for short-term epidemiology forecasting. We emphasize that the reported number of infections may not necessarily be correlated with the actual transmission rate due to factors such as different testing criteria and varying accessibility in testing over time.

We deployed five machine learning models to predict the CIG for individual countries across 14 days. Explicitly, this value was the label (ie, dependent variable) we sought to predict. We used features (ie, independent variables) representing the implementation levels of NPIs and the cultural dimensions of each country. We obtained daily metrics for the implementation of NPIs at the national level from the OxCGRT data set [1]. Although different countries may implement similar NPIs, researchers have suggested that cross-cultural variations across populations lead to different perceptions and responses toward these NPIs [53,54,68]. We intended to capture any effects due to national cross-cultural differences by complementing the OxCGRT data set with national cultural norm values from the Hofstede cultural dimensions [58]. Our non–time series deep learning models predicted the expected future national CIG using both NPI implementation and cultural norm features. Although time series deep learning models (eg, recurrent neural networks or transformers) may also provide CIG predictions, these models generally require greater amounts of accurately labeled trajectory data and assume that past trajectory trends are readily available representatives of future trajectories. Instead, our non–time series models were trained on more granular data that did not necessarily need to be temporally concatenated into a trajectory. We also opted for less complex non–time series models due to indeterminacies in acquiring and verifying sufficient trajectory data, especially due to the lack of reliable data at the onset of the COVID-19 outbreak.

Our results suggest that non–time series machine learning models can predict future CIG according to multiple validation methods, depending on the user's application. Although we do not necessarily claim state-of-the-art performance for infection rate prediction given the rapidly growing amount of parallel work in this area, to the best of our knowledge, our work is the first to use machine learning techniques to predict the change in national cumulative numbers of reported COVID-19 infections by combining NPI implementation features with national cultural features.

Our implementation uses publicly available data retrieved from the internet and relies on the open-sourced Python libraries Pandas [69] and Scikit-Learn [70].

Methods

Data and Preprocessing

Candidate features at the national level were extracted from three data sets for input into our machine learning models: NPIs, cultural dimensions, and current confirmed COVID-19 case numbers.

OxCGRT provides daily level metrics of the NPIs implemented by countries [1]. This data set sorts NPIs into 17 categories, each with either an ordinal policy level metric ranging from 0 (not implemented) to 2, 3, or 4 (strictly enforced) or a continuous metric representing a monetary amount (eg, research funding). The value of each national NPI metric is assigned daily from data in publicly available sources by a team of Oxford University staff and students using the systematic format described in [1]. We limited our candidate features to the 13 ordinal policy categories and 4 computed indices, which represent the implementation of different policy types taken by governments, based on the implemented NPIs. This data set contains data starting from January 1, 2020.

To represent cultural differences across populations of different countries, the 2015 edition of the Hofstede cultural dimensions [62,71] was tagged to each country. Although these dimensions are rarely used in epidemiology studies, they have been used

frequently in international marketing studies and cross-cultural research as indicators of the cultural values of national populations [61,72]. Multiple studies have also linked cultural dimensions to health care–related behavior, such as antibiotic usage and body mass index [73-76]. Because the 2015 edition of this data set groups certain geographically neighboring countries together (eg, Ivory Coast, Burkina Faso, Ghana, etc, into Africa West), we tagged all subgroup countries with the dimension values of their group. Although we recognize that this approach is far from ideal and will likely lead to some degree of inaccurate approximation in these subgroup countries, we performed this preprocessing step to include those countries in our study. The dimension values for each country were constant across all samples. Six cultural dimensions were presented for each country or region [71]:

- Power distance index: the establishment of hierarchies in society and organizations and the extent to which lower hierarchical members accept inequality in power
- Individualism versus collectivism: the degree to which individuals are not integrated into societal groups, such as individual or immediate family (individualistic) versus extended families (collectivistic)
- Uncertainty avoidance: a society's tendency to avoid uncertainty and ambiguity through use of societal disapproval, behavioral rules, laws, etc
- Masculinity versus femininity: Societal preference toward assertiveness, competitiveness, and division in gender roles (masculinity) compared to caring, sympathy, and similarity in gender roles (femininity)
- Long-term versus short-term orientation: Societal values toward tradition, stability, and steadfastness (short-term) versus adaptability, perseverance, and pragmatism (long-term)
- Indulgence versus restraint: The degree of freedom available to individuals for fulfilling personal desires by social norms, such as free gratification (indulgence) versus controlled gratification (restraint)

We extracted the daily number of confirmed cases, n_t, for each country from the COVID-19 Data Repository by the CSSE at Johns Hopkins University [34]. We used a rolling average of the previous 5-day window to smooth fluctuations in n_t, which may be caused by various factors, such as inaccurate case reporting, no release of confirmed case numbers (eg, on weekends and holidays), and sudden infection outbreaks. We refer to the smoothed daily number of confirmed cases for date t as  .

We computed the CIG for a specified date, τ, as:

The CIG represents the expected number of new confirmed cases from date $\tau - 13$ to date τ as a percentage of the total number of confirmed infection cases up to date $\tau - 14$.

Our goal was to predict the CIG 14 days in advance (ie, $CIG_{\tau+14}$) given information from the current date τ for each country. Available candidate features included all ordinal policy metrics and the four computed indices from OxCGRT, the six cultural dimension values from the Hofstede model, the CIG of the current date CIG_τ, and the smoothed cumulative number of confirmed cases  , for a total of 25 candidate features. Neither the date nor any other temporal features were included.

We trimmed samples with fewer than 10 cumulative confirmed infection cases and with the highest 2.5% and the lowest 2.5% of $CIG_{\tau+14}$ to remove outliers in the data. Because the lowest 2.5% of $CIG_{\tau+14}$ were all 0.0%, we removed the samples with $CIG_{\tau+14}$=0.0% by ascending date.

Our data range from April 1 to September 30, 2020, inclusively. We excluded all countries from our combined data set that had missing feature values. In total, our combined data set and our experiments applied to 114 countries: Algeria, Angola, Argentina, Australia, Austria, Bahrain, Bangladesh, Belgium, Benin, Botswana, Brazil, Bulgaria, Burkina Faso, Burundi, Cameroon, Canada, Central African Republic, Chad, Chile, China, Colombia, Comoros, Croatia, Czech Republic, Denmark, Djibouti, Egypt, El Salvador, Eritrea, Estonia, Ethiopia, Finland, France, Gabon, Gambia, Germany, Ghana, Greece, Guinea, Hong Kong, Hungary, India, Indonesia, Iran, Iraq, Ireland, Italy, Japan, Jordan, Kenya, Kuwait, Latvia, Lebanon, Lesotho, Liberia, Libya, Lithuania, Luxembourg, Madagascar, Malawi, Malaysia, Mali, Mauritania, Mauritius, Mexico, Morocco, Mozambique, Namibia, Netherlands, New Zealand, Niger, Nigeria, Norway, Oman, Pakistan, Palestine, Peru, Philippines, Poland, Portugal, Qatar, Romania, Russia, Rwanda, Saudi Arabia, Senegal, Serbia, Seychelles, Sierra Leone, Singapore, Slovenia, Somalia, South Sudan, Spain, Sudan, Sweden, Switzerland, Syria, Taiwan, Tanzania, Thailand, Togo, Trinidad and Tobago, Tunisia, Turkey, Uganda, United Arab Emirates, United States, Uruguay, Venezuela, Vietnam, Yemen, Zambia, and Zimbabwe.

The mean, standard deviation, and range of each candidate feature value for the above countries are shown in Table 1.

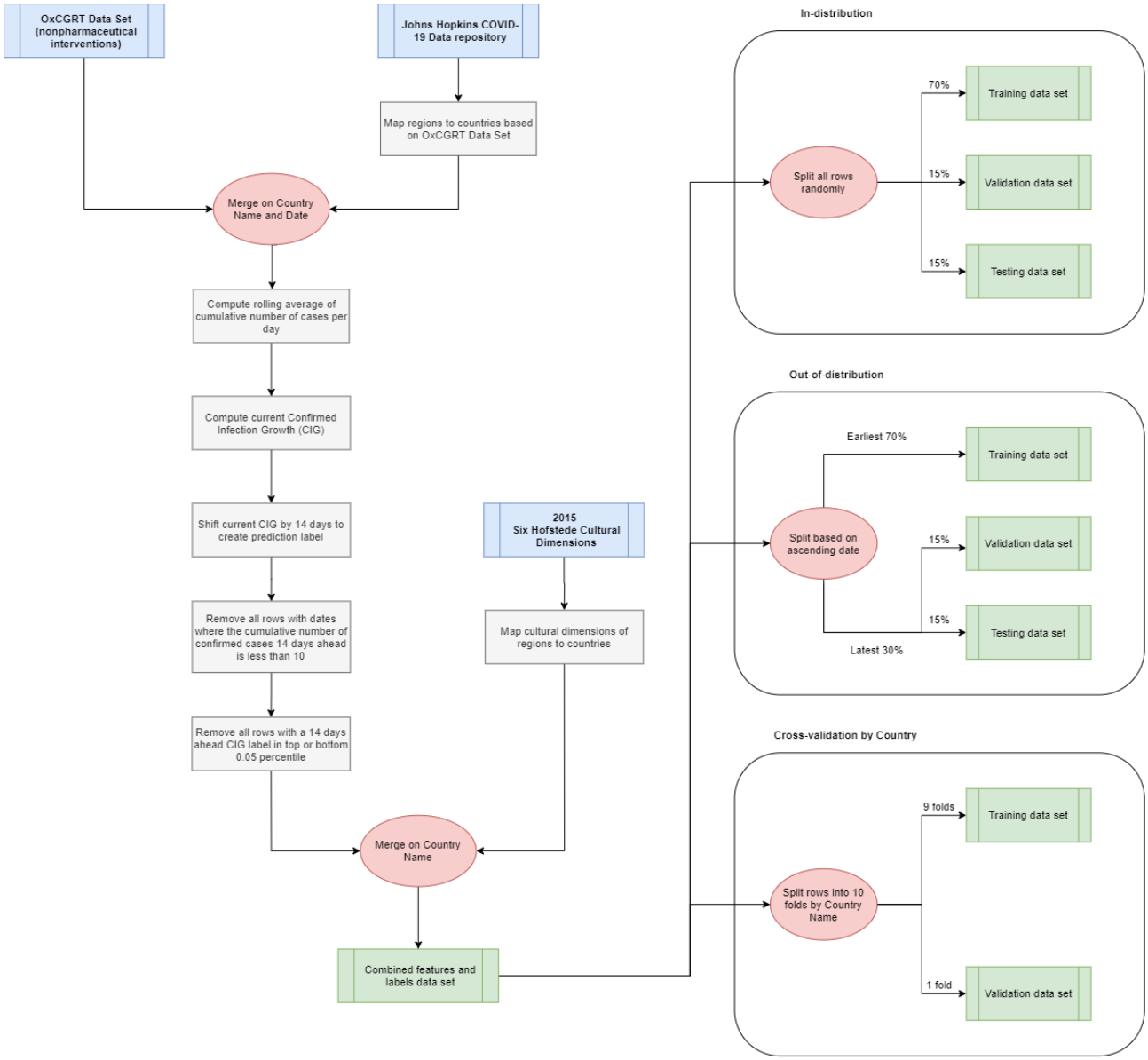
The data preprocessing procedure is shown in Figure 1.

Table 1. Statistical measurements of candidate feature values.

Candidate features	Mean (SD)	Range
Nonpharmaceutical interventions		
School closure	2.23 (1.01)	0.00 to 3.00
Workplace closure	1.67 (0.92)	0.00 to 3.00
Cancellation of public events	1.64 (0.65	0.00 to 2.00
Restrictions on gatherings	2.89 (1.27)	0.00 to 4.00
Closure of public transport	0.71 (0.77)	0.00 to 2.00
Stay-at-home requirements	1.17 (0.90)	0.00 to 2.00
Restrictions on internal movement	1.15 (0.88)	0.00 to 2.00
International travel controls	3.13 (1.00)	0.00 to 4.00
Income support	1.04 (0.79)	0.00 to 2.00
Debt/contract relief	1.23 (0.76)	0.00 to 2.00
Public information campaigns	1.97 (0.23)	0.00 to 2.00
Testing policy	1.84 (0.82)	0.00 to 2.00
Contact tracing	1.50 (0.64)	0.00 to 2.00
Stringency Index	63.02 (20.57)	0.00 to 100.00
Government Response Index	61.43 (15.03)	0.00 to 95.54
Containment Health Index	62.91 (16.50)	0.00 to 98.96
Economic Support Index	52.53 (28.93)	0.00 to 100.00
Current infection numbers		
Current cumulative number of confirmed cases:	113,302.24 (505,170.50)	4.00 to 7,155,220.00
CIG_{τ} [a]	0.85 (3.83)	–0.423 to 228.00
Hofstede cultural dimensions		
Power distance	66.74 (17.34)	11.00 to 104.00
Individualism	38.52 (18.71)	12.00 to 91.00
Masculinity	48.32 (14.06)	5.00 to 95.00
Uncertainty avoidance	64.17 (17.42)	8.00 to 112.00
Long-term orientation	35.36 (21.52)	3.52 to 92.95
Indulgence	46.88 (20.47)	0.00 to 100.00

[a]CIG_{τ}: confirmed infection growth on the current day.

Figure 1. Data preprocessing pipeline from the OxCGRT data set, Johns Hopkins COVID-19 Data Repository, and six Hofstede cultural dimensions to the training, validation, and test data sets for each validation method. OxCGRT: Oxford COVID-19 Government Response Tracker.



Feature Selection and Processing

We selected features to input into our machine learning models from our candidate feature pool using mutual information [77]. Mutual information is a measure of the dependency between an individual feature (ie, the independent variable) and the label (ie, the dependent variable), and it captures both linear and nonlinear dependencies. However, mutual information does not capture multivariate dependencies or indicate collinearity between features. To include both linear and nonlinear dependencies, features are selected if they achieve substantially nonzero mutual information (ie, greater than 0.10). Feature selection was conducted prior to training with the training set in all validation methods. Similar feature filtering and selection techniques have been used in other machine learning applications [70,78]. The candidate features considered for input and their respective mutual information are listed in Table 2 for the in-distribution and out-of-distribution validation methods. Mutual information was also computed for each of the ten folds of the cross-validation method.

All selected features were then normalized to the range [0,1] using standard min-max normalization.

Table 2. Mutual information of candidate features for the in-distribution and out-of-distribution validation methods. In the cross-validation method, the 10 folds have varying mutual information.

Candidate feature	Mutual information	
	In-distribution	Out-of-distribution
Nonpharmaceutical interventions		
School closure[a,b]	0.184	0.205
Workplace closure[b]	0.098	0.127
Cancellation of public events[b]	0.089	0.127
Restrictions on gatherings[a,b]	0.107	0.112
Closure of public transport[b]	0.094	0.124
Stay-at-home requirements[a,n]	0.139	0.163
Restrictions on internal movement[a,b]	0.126	0.146
International travel controls	0.099	0.099
Income support[b]	0.095	0.110
Debt/contract relief	0.043	0.053
Public information campaigns	0.020	0.023
Testing policy	0.056	0.064
Contact tracing	0.030	0.038
Stringency Index[a,b]	0.638	0.668
Government Response Index[a,b]	0.634	0.641
Containment Health Index[a,b]	0.621	0.655
Economic Support Index[a,b]	0.119	0.124
Current infection numbers		
Current cumulative number of confirmed cases: [a,b]	0.517	0.557
CIG_{τ} [a,b,c]	0.866	0.798
Hofstede cultural dimensions		
Power distance[a,b]	0.288	0.342
Individualism[a,b]	0.309	0.355
Masculinity[a,b]	0.310	0.372
Uncertainty avoidance[a,b]	0.314	0.370
Long-term orientation[a,b]	0.461	0.535
Indulgence[a,b]	0.456	0.529

[a]Selected feature for the in-distribution method.

[b]Selected feature for the out-of-distribution method.

[c]CIG_{τ}: confirmed infection growth on the current day.

Model Training and Validation

We trained the machine learning models by performing a grid search over the combinations of hyperparameters listed in Table 3 [70,79-82]. We optimized the models using the mean squared error (MSE) criterion and selected the model hyperparameters with the lowest mean absolute error (MAE) as the optimal configuration of the model. The MSE heavily penalizes large residual errors disproportionately, while the MAE provides an absolute mean of all residual errors [83]. The MAE of the training data acts as a measure of the goodness-of-fit of the model, while the MAE of the validation and testing data acts as a measure of the predictive performance [84].

Table 3. Machine learning models and hyperparameter combinations used in the grid search.

Model	Hyperparameters
Ridge regression	
α	0.00, 0.25, 0.50, 0.75, 1.00, 1.25
Decision tree regression	
Depth	5, 10, 15, 20, 25, 30
Minimum sample split	2, 5, 10
Minimum sample leaves	1, 2, 4, 8, 10
Random forest regression	
Depth	5, 10, 20, 25, 30
Estimators	3, 5, 10, 15, 20, 30, 50, 75, 100, 125, 150
Minimum sample split	2, 5, 10
Minimum sample leaves	1, 2, 4, 8, 10
AdaBoost[a] regression	
Weak learner	Decision tree (maximum depth: 2)
Estimators	3, 5, 10, 15, 20, 30, 50, 75, 100, 125, 150
Loss function	Linear
Learning rate	0.1, 0.5, 1.0
Support vector regression	
ε	0.00, 0.10, 0.20, 0.50
Kernel	Linear, radial, sigmoid

[a]AdaBoost: adaptive boosting.

To validate in-distribution and out-of-distribution, we split our samples into 70-15-15 training-validation-test sets. For cross-validation [85,86], we split our samples into 10 folds (ie, 90-10). These three methods of validation each represent a different definition of performance for the machine learning models.

In-Distribution Validation

We randomly split the samples into training, validation, and test sets. Consequently, the models were trained from samples distributed across the entire date range available in our data. This is critical, as it is generally expected that model performance is best when training and test data are drawn from the same distribution. Because the COVID-19 infection numbers naturally constitute a time series, this method ensures that validation and test samples are indeed from the same distribution as the training samples. Because the samples are disassociated from their dates and all other known temporal features, the prediction of the validation and test samples using the training samples is unordered. This method may be applicable to use cases in which the date-to-predict is expected to be in a similar distribution as the training samples, such as predicting $CIG_{\tau+14}$ when data up to the current date τ are available.

Out-of-Distribution Validation

Although the in-distribution method can ensure that the training, validation, and test data are all sampled from the same distribution, it may not necessarily be the most practical method. Generally, the goal of long-term infection rate forecasting is to anticipate future infection rates, and it should not be represented as an in-distribution task, where we trained it with data from near or later than the date-to-predict. Therefore, we also validated the performance of our models by training on the earliest 70% of the samples. The validation and test sets were then randomly split between the remaining 30% of the samples. This setup ensures that all training samples occurred earlier than the validation and testing samples and that no temporal features (known or hidden) were leaked. However, due to the changing environment related to COVID-19 infections (eg, the introduction of new NPIs, seasonal changes, new research), the validation and testing distributions are likely different from that of the training set. This method may be applicable for use cases in which the date-to-predict is in the far future and not all data up to 14 days prior to the date-to-predict are available.

Country-Based Cross-Validation

As a compromise between the above two methods, we also used a cross-validation method in which we split the available countries into 10 folds. The aim was to evaluate validation samples from the same date range as the training samples, but not the same country trajectory. That is, only data from countries not in the validation set are included in the training set. Although the samples from the training and validation sets are therefore sampled from different distributions (ie, different countries), we anticipate that features from the Hofstede cultural dimensions [58] may assist in identifying similar characteristics between countries, thus reducing the disparity between the training and validation distributions. This method may be applicable in

predicting the CIG of countries for which previous associated data is unavailable or unreliable.

Results

Feature Selection

For both the in-distribution and out-of-distribution training sets, we observed that most candidate features met our requirement of nonzero mutual information (≥0.10) (see Table 2).

In both training sets, the candidate features that did not meet the requirements were international travel control (0.099, 0.099), debt/contract relief (0.043, 0.053), public information campaigns (0.020, 0.023), testing policy (0.056, 0.064), and contact tracing (0.030, 0.038). Additional candidate features that did not meet the requirements for the in-distribution training set were workplace closure (0.098) and cancellation of public events (0.089). Overall, the in-distribution and out-of-distribution data sets contained 17 and 20 features, respectively.

CIG_{τ} had the highest mutual information out of all features, suggesting similarities between the feature CIG_{τ} and the label $CIG_{\tau+14}$. Further analysis showed a correlation of r=.309 between CIG_{τ} and $CIG_{\tau+14}$. This may be due to similar trends in the CIG when the implementation of NPIs is consistent within a 14-day period. We also observed that all candidate features for the six Hofstede cultural dimensions had higher mutual information than all individual NPI candidate features, aside from the aggregated indices. This finding suggests a high statistical relationship between each cultural dimension feature and the label we sought to predict. Although the cultural dimension values may not fully represent the cultural differences of each country (see Limitations), there is sufficient information between each cultural dimension feature and the label for them to be relevant predictors of the label.

Comparison of Machine Learning Models

Out of all the available configurations (ie, hyperparameter combinations) of each model, we selected the model configurations with the lowest validation errors and computed the test errors. The parameters for these selected models are listed in Table 4. The mean training, validation, and test errors are included in Table 5, Table 6, and Table 7, respectively, for the in-distribution, out-of-distribution, and cross-validation methods. We also include the median percent error [87], which is the percentage difference of the prediction $f(x^{(i)})$ and the label $y^{(i)}$ for each instance $\{x^{(i)}y^{(i)}\}$, computed as:

We observed that random forest regression had the lowest mean test error in the interpolation method (0.031) and adaptive boosting (AdaBoost) regression had the lowest mean test errors in the extrapolation and cross-validation methods (0.089 and 0.167, respectively) (see Table 5, Table 6, and Table 7). For all models aside from ridge regression, the in-distribution method had the lowest mean test errors and the lowest median percent error.

Table 4. Hyperparameters of the optimal configuration (lowest validation mean absolute error) for each model for each validation method.

Model	Validation method		
	In-distribution	Out-of-distribution	Cross-validation
Ridge regression			
α	0.00	0.25	0.00
Decision tree regression			
Depth	25	10	5
Minimum sample split	2	5	2
Minimum sample leaves	1	1	4
Random forest regression			
Depth	30	15	15
Estimators	150	10	125
Minimum sample split	2	2	2
Minimum sample leaves	1	10	10
AdaBoost[a] regression			
Estimators	5	5	3
Learning rate	0.1	1.0	0.1
Support vector regression			
ε	0.00	0.00	0.00
Kernel	Radial	Linear	Linear

[a]AdaBoost: adaptive boosting.

Table 5. Optimal MAE and median percent error values for the in-distribution validation method.

Model	Train MAE[a]	Validation MAE	Test MAE	Validation percent error	Test percent error
Ridge regression	0.270	0.269	0.259	1.58	0.60
Decision tree regression	0.001	0.041	0.039	1.00	0.00
Random forest regression[b]	0.012	0.033	0.031	1.01	1.01
AdaBoost[c] regression	0.162	0.166	0.155	1.31	1.24
Support vector regression	0.170	0.172	0.165	1.00	1.01

[a]MAE: mean absolute error.

[b]The model with the lowest test MAE.

[c]AdaBoost: adaptive boosting.

Table 6. Optimal MAE and median percent error values for the out-of-distribution validation method.

	Train MAE[a]	Validation MAE	Test MAE	Validation percent error	Test percent error
Ridge regression	0.296	0.240	0.247	2.26	1.22
Decision tree regression	0.117	0.109	0.114	1.15	0.12
Random forest regression	0.098	0.098	0.105	1.45	0.44
AdaBoost[b] regression[c]	0.207	0.081	0.089	1.40	0.39
Support vector regression	0.268	0.167	0.176	1.66	0.60

[a]MAE: mean absolute error.

[b]AdaBoost: adaptive boosting.

[c]The model with the lowest test MAE.

Table 7. Optimal MAE and median percent error values for the cross-validation method. Validation error is equivalent to test error for cross-validation.

Model	Train MAE[a]	Validation MAE	Validation percent error
Ridge regression	0.262	0.275	0.62
Decision tree regression	0.181	0.207	0.28
Random forest regression	0.073	0.175	0.40
AdaBoost[b] regression[c]	0.164	0.167	0.27
Support vector regression	0.230	0.240	0.03

[a]MAE: mean absolute error.

[b]AdaBoost: adaptive boosting.

[c]The model with the lowest test MAE.

Analysis of Best-Performing Models

Intercepts near 0.0 and slopes near 1.0 are the linear calibration measures that indicate a perfect calibration relationship between the predictions and the labels [84]. For the optimal models in all the validation methods, we observed slopes close to 1.0 and intercepts close to 0.0 (see Table 8). Due to the large sample sizes, statistical significance testing indicated that several slopes and intercepts are significantly different from 1.0 and 0.0, respectively. However, the small mean differences (standardized to the standard deviation, ie, the z score) indicate that these differences have no practical significance. High correlations (r>0.70) and moderate-to-high R^2 values (R^2>.50) [88,89] between the predictions and labels were observed in all three validation methods (see Figure 2, Figure 3, and Figure 4).

To assess the fine-grained model performance, we discretized both the true labels and model predictions into bins of size 0.5 for all three validation methods (see Figure 5, Figure 6, and Figure 7). Comparing the resulting empirical distributions, it can be seen that the resulting distributions are extremely similar in both the in-distribution and out-of-distribution methods. In the cross-validation method, the predictions skew slightly higher than the labels in the 0.0-1.0 range, showing a general tendency of the model to slightly overestimate the CIG within this range.

Further analysis shows that the performance of the models varies with the values of the labels. In both the in-distribution and cross-validation methods, the test MAE is lowest for samples with labels of 0.0 (see Table 9 and Table 10), followed by the label range of 0.0-0.5. In the out-of-distribution method, the test MAE is lowest for samples with labels from 0.0-0.5 (see Table 11). For all validation methods, the mean MAE and

median percent errors also increase with label bins greater than 1.0, showing a decrease in accuracy for a larger CIG.

Table 8. Linear calibration measures of the models with the lowest test mean absolute error for each validation method.

Measure	Validation method		
	In-distribution	Out-of-distribution	Cross-validation
Test sample size, n	2847	2811	19,669
Model	Random forest	AdaBoost[a]	AdaBoost
Correlation, *r*	0.979	0.716	0.758
Slope (SE)	1.037 (0.004)	0.986 (0.018)	0.968 (0.006)
Slope standardized mean difference (z score) from 1	0.176	–0.015	–0.039
Slope *P* value (mean of 1)	<.001	.43	<.001
Intercept (SE)	–0.013 (0.002)	–0.011 (0.004)	0.006 (0.003)
Intercept standardized mean difference (z score) from 0	–0.119	–0.044	0.014
Intercept *P* value (mean of 0)	<.001	.02	.06
R^2 value	0.959	0.513	0.574

[a]AdaBoost: adaptive boosting.

Figure 2. Calibration plot between the labels and predictions for the interpolation validation method, with the mean of each prediction bin of size 0.25.

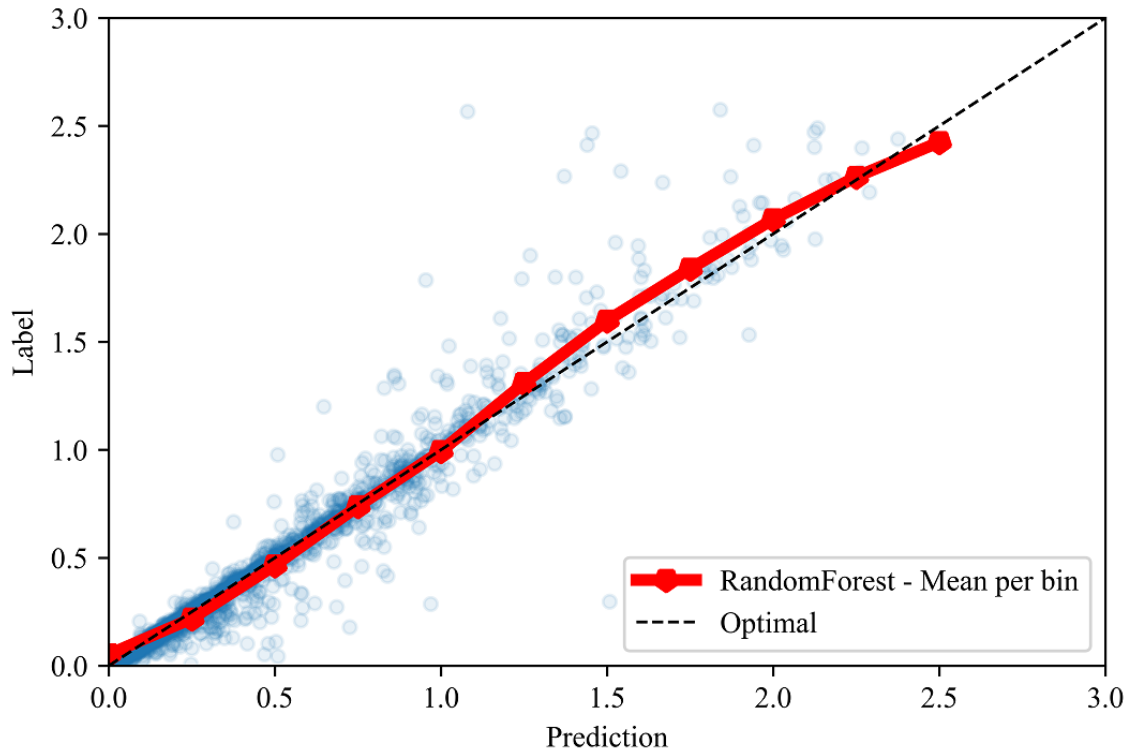


Figure 3. Calibration plot between the labels and predictions for the extrapolation validation method, with the mean of each prediction bin of size 0.25.

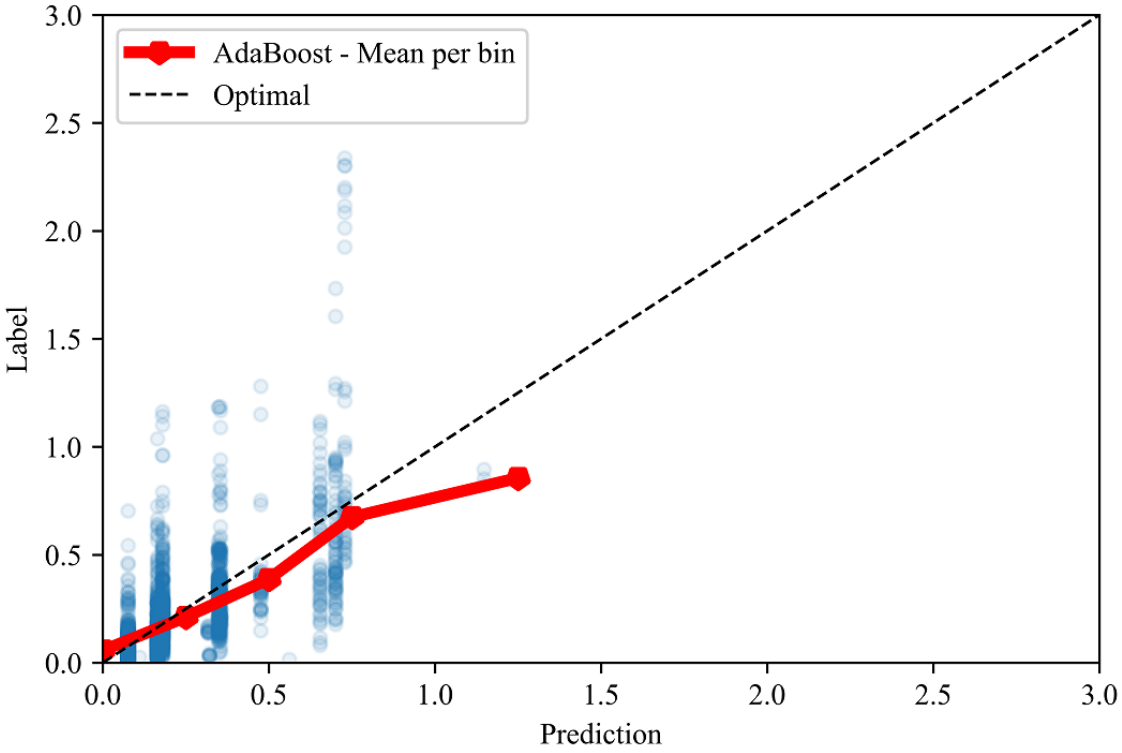


Figure 4. Calibration plot between the labels and predictions for the cross-validation method, with the mean of each prediction bin of size 0.25.

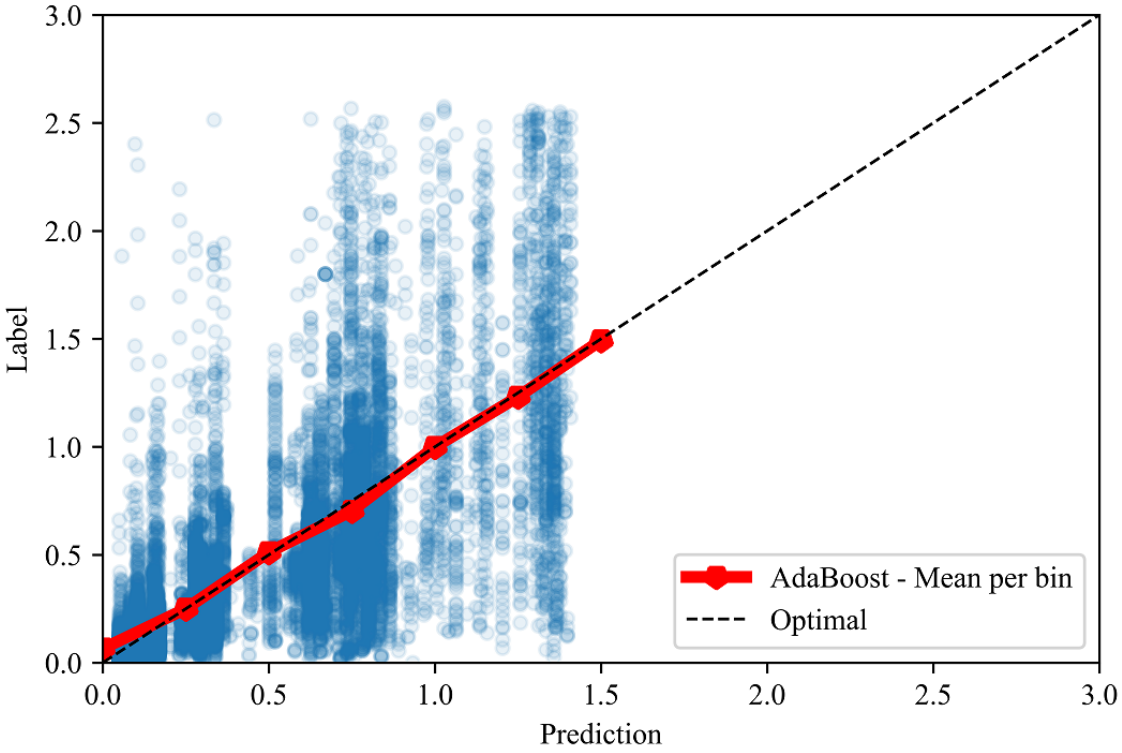


Figure 5. Distributions of the test labels (ie, true confirmed infection growth) and model predictions (n=2847) for the in-distribution method.

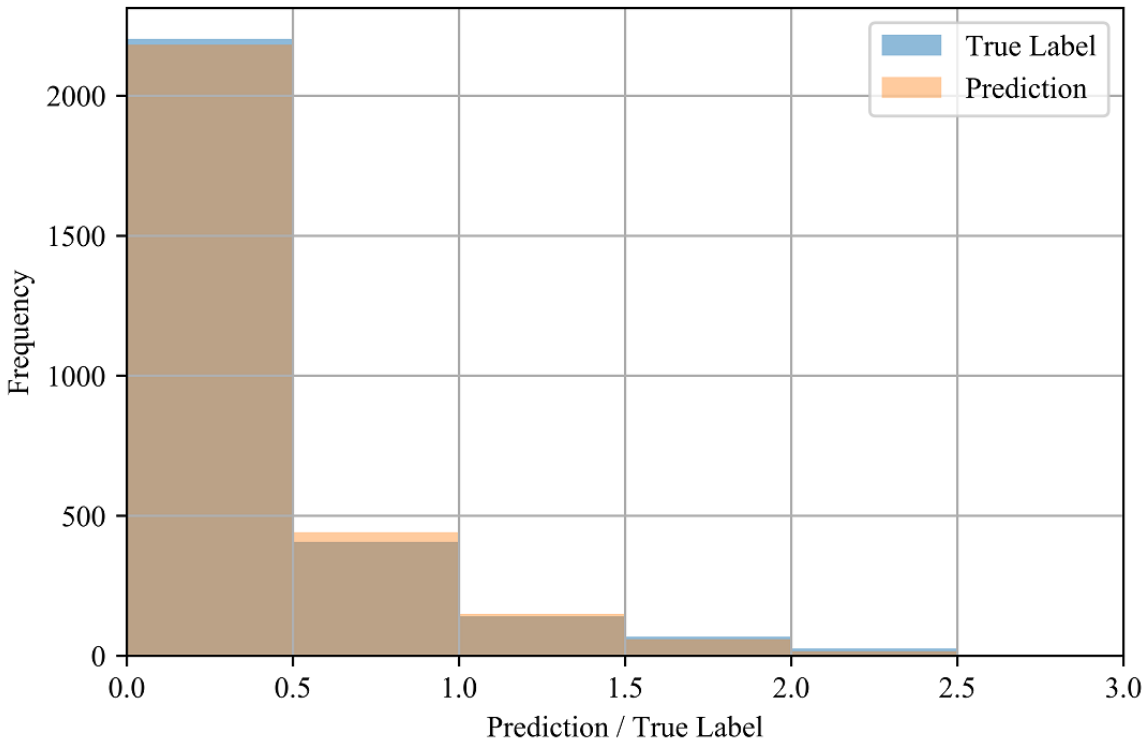


Figure 6. Distributions of the test labels (ie, true confirmed infection growth) and model predictions (n=2811) for the out-of-distribution method.

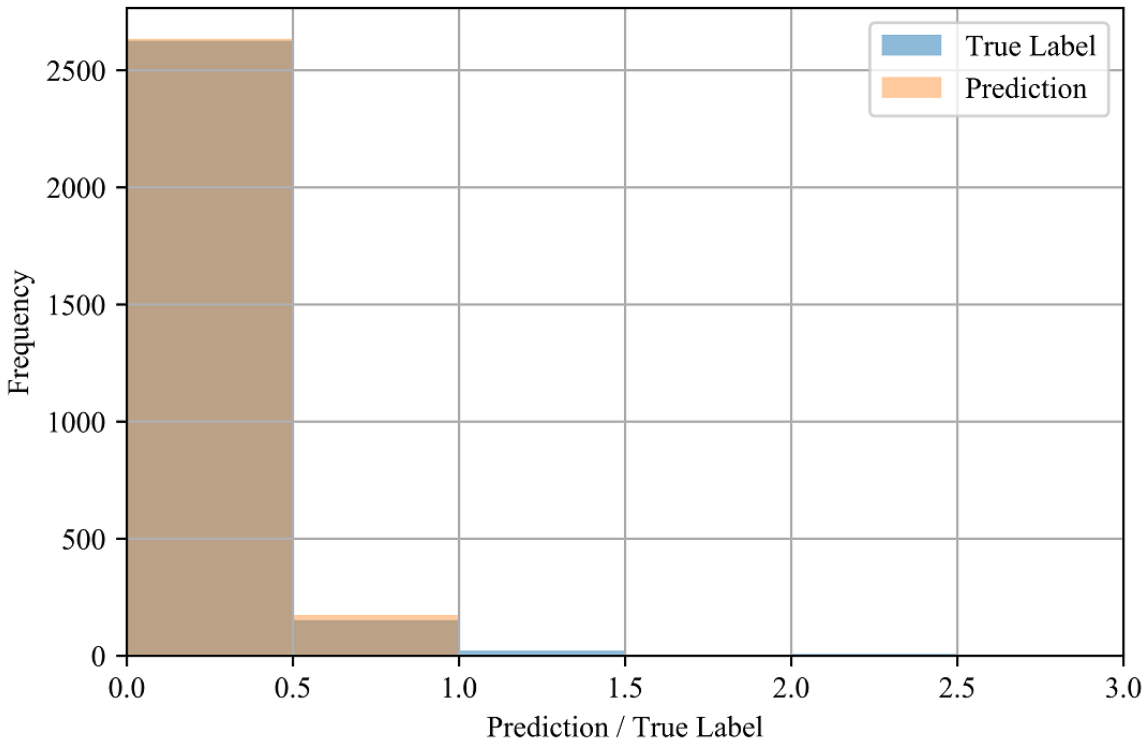


Figure 7. Distributions of the test labels (ie, true confirmed infection growth) and model predictions (n=19,669) for the cross-validation method.

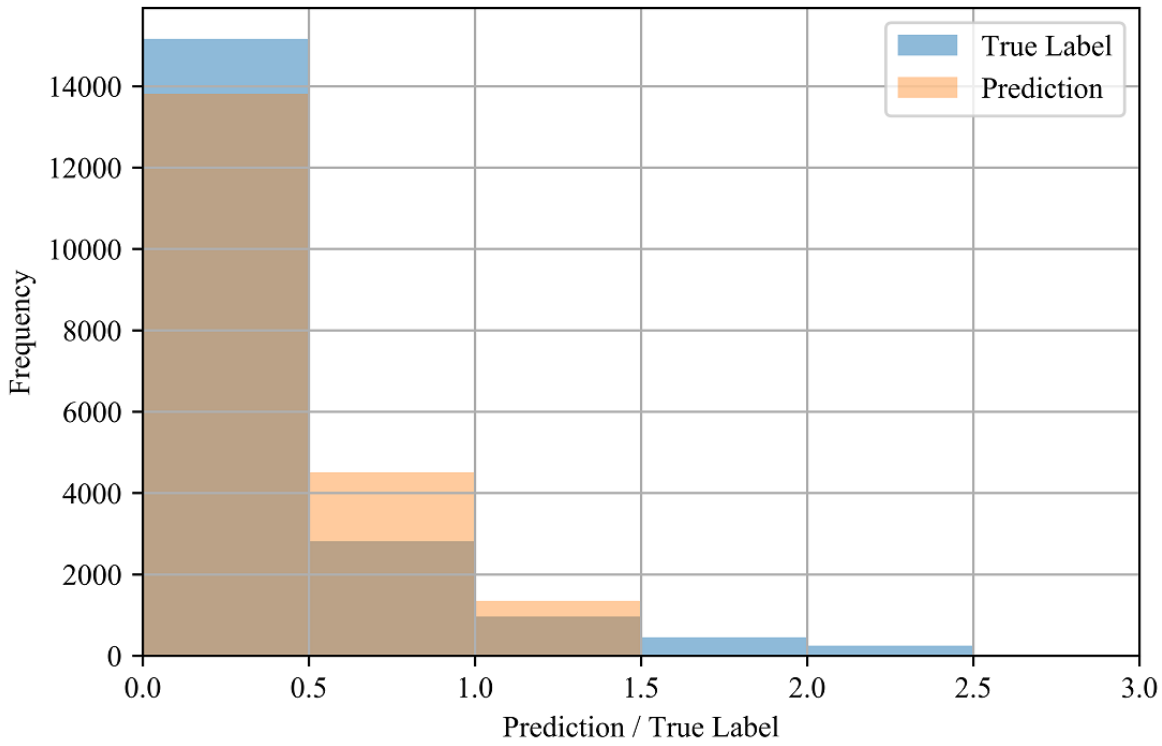


Table 9. Test errors and median percent errors of label bins of size 0.5 for the in-distribution validation method.

Upper threshold	Count	Test mean nonabsolute error (SD)	Test mean absolute error (SD)	Test percent error
0.0	20	0.000 (0.000)	0.000 (0.000)	N/A[a]
0.5	2183	0.011 (0.052)	0.017 (0.050)	0.01
1.0	408	0.003 (0.076)	0.047 (0.060)	0.00
1.5	140	–0.052 (0.139)	0.094 (0.115)	–0.02
2.0	68	–0.104 (0.205)	0.158 (0.167)	–0.04
2.5	26	–0.283 (0.309)	0.297 (0.294)	–0.08
3.0	2	–1.108 (0.470)	1.108 (0.470)	–0.43

[a]N/A: not applicable.

Table 10. Test errors and median percent errors of label bins of size 0.5 for the cross-validation method.

Upper threshold	Count	Test mean nonabsolute error (SD)	Test mean absolute error (SD)	Test percent error
0.0	114	–0.059 (0.086)	0.059 (0.086)	N/A[a]
0.5	15,056	–0.073 (0.174)	0.109 (0.153)	0.493
1.0	2815	–0.010 (0.282)	0.217 (0.181)	–0.006
1.5	960	0.333 (0.337)	0.393 (0.265)	–0.299
2.0	451	0.719 (0.370)	0.719 (0.370)	–0.391
2.5	246	1.141 (0.321)	1.141 (0.321)	–0.459
3.0	27	1.362 (0.266)	1.225 (0.266)	–0.486

[a]N/A: not applicable.

Table 11. Test errors and median percent errors of label bins of size 0.5 for the out-of-distribution validation method.

Upper threshold	Count	Test mean nonabsolute error (SD)	Test mean absolute error (SD)	Test percent error
0.0	19	0.076 (0.000)	0.076 (0.000)	N/A[a]
0.5	2607	0.034 (0.096)	0.071 (0.074)	0.44
1.0	152	–0.161 (0.228)	0.225 (0.164)	–0.25
1.5	22	–0.648 (0.222)	0.648 (0.222)	–0.52
2.0	3	–1.044 (0.147)	1.044 (0.147)	–0.60
2.5	8	–1.464 (0.116)	1.464 (0.116)	–0.67

[a]N/A: not applicable.

Discussion

Principal Results

Our results suggest that traditional, non–time series machine learning models can predict future CIG to an appreciable degree of accuracy, as suggested by the moderate-to-high R^2 values (R^2>0.50) and strong linear calibration relationships (r>0.70) [88,89] between the labels and predictions in all the validation methods.

A comparison of our results for all the validation methods suggests differences in the predictive performance of the machine learning models across the varying use cases. The in-distribution method has the highest R^2 value and the lowest test mean error and median percent error; this is to be expected, as the test samples were obtained from the same distribution as the training samples. Intuitively, although the samples in the in-distribution method are unordered (ie, no temporal features are included), the availability of samples across the entire temporal range in the training set enables the validation and test samples to interpolate between these training samples.

The out-of-distribution method achieved a higher test mean error and a lower R^2 value than the in-distribution method. This is expected, as the evolving COVID-19 infection trajectories observed in most countries give distributions of training samples from earlier dates that may differ greatly from those of validation and test samples from later dates (ie, data shift), which machine learning models are often ill-equipped to handle [90].

Conversely, although the cross-validation method contained the training and validation sets within the same date range, the cross-validation method also separated countries across these sets (ie, the 10 folds) such that no country had samples in both the training and validation sets. This difference led to higher test mean errors and median percent errors than the other two methods and a similar R^2 value to that of the out-of-distribution method, suggesting that including training samples from the same country as the validation samples is more important than ensuring temporal overlap. We speculate that this result occurs because the unique cultural dimensions per country may potentially act as categorical rather than continuous features for each country. In such cases, the cultural dimensions observed in the training set would be considered irrelevant to the cultural dimensions within the validation set.

The performance also varied depending on the value of the label (see Table 9, Table 10, and Table 11), which may be due to the imbalanced frequency of the training samples. That is, the rareness of samples with higher CIG compared to lower CIG in the training set may be the cause of their comparatively poor performance.

In Figure 3 and Figure 4, we also show constraints of the trained AdaBoost regression models. The discretization of the prediction values may be due to the low number of estimators used in the lowest mean test error configuration, as shown in Table 4. The low number of estimators in these configurations may also restrict the predictions to a maximum of 1.5 selected to the relatively low number of samples with labels greater than 1.5 (see Figure 6 and Figure 7). The label ranges with the most samples are selected over underrepresented ranges as candidates for prediction values in the discretized AdaBoost regression models. Although additional estimators in the AdaBoost regression models may result in less discrete prediction values, they may also cause over-fitting by increasing the complexity of the models.

Limitations

First, the scores in the OxCGRT and Hofstede cultural dimensions data sets are imprecise. NPI enforcement levels and definitions may vary even between countries with the same scores, while countries sharing similar cultural dimension scores may have unobserved differences in terms of cultural practices due to low representation of their cultures with only six dimensions. Although the Hofstede model is convenient for the goal of our work, it does not identify intracountry cultural differences. Furthermore, distinct countries may be grouped within specific geographical regions (eg, Africa West). We also acknowledge that there are trade-offs between different cultural models and different definitions of culture [61]. We encourage further exploration of appropriate cultural dimensions in addition to the Hofstede model, such as GLOBE [59] and CVSCALE [60]. Second, by predicting the CIG 14 days in advance of the current date, the models do not account for information regarding changes in NPIs between the current date and the date-to-predict. Third, the CIG is a measure of the change in the cumulative number of confirmed infections and may not necessarily be correlated with the change in the daily number of confirmed infections or the actual transmission rate of COVID-19. For example, differences in testing and reporting policies of different jurisdictions (eg, prioritizing high-risk patients, performing more tests per capita, and obfuscating test

results) may lead to a misleading representation of the infection growth.

Conclusion

In this study, we trained five non–time series machine learning models to predict the CIG 14 days into the future using NPI features extracted from the OxCGRT data set [1] and cultural norm features extracted from the Hofstede cultural dimensions [58]. Together, these features enabled the prediction of near-future CIG in multiple machine learning models. Specifically, we observed that random forest regression and AdaBoost regression resulted in the most accurate predictions out of the five evaluated machine learning models.

We observed differences in the predictive performance of the machine learning models across the three validation methods; the highest accuracy was achieved with the in-distribution method and the lowest with the cross-validation method. These differences in performance suggest that the models have varying levels of accuracy depending on the use case. Specifically, predictions are expected to have higher accuracies when existing data from the same country in nearby dates are available (ie, in-distribution method). This enables applications such as predicting the CIG over the upcoming 14 days from the current date. The decrease in accuracy when data from nearby dates are unavailable (ie, the out-of-distribution method) suggests weaker performance in predicting the CIG over 14 days for relatively distanced future dates. We observed the greatest decrease in performance when data from the same country were unavailable (ie, the cross-validation method). However, with all validation methods, we observed appreciable calibration measures between the predictions and labels of the test set.

This study adds to the rapidly growing body of work related to predicting COVID-19 infection rates by introducing an approach that incorporates routinely available data on NPIs and cultural dimensions. Importantly, this study emphasizes the utility of NPIs and cultural dimensions for predicting country-level growth of confirmed infections of COVID-19, which to date have been limited in existing forecasting models. Our findings offer a new direction for the broader inclusion of these types of measures, which are also relevant for other infectious diseases, using non–time series machine learning models. Our experiments also provide insight into validation methods for different applications of the models. As the availability of this data increases and the nature of the data continues to evolve, we expect that models such as these will produce accurate and generalizable results that can be used to guide pandemic planning and other infectious disease control efforts.

Acknowledgments

FR is supported by a Canadian Institute for Advanced Research Chair in Artificial Intelligence.

Conflicts of Interest

None declared.

References

1. Hale T, Petherick A, Phillips T, Webster S. Variation in government responses to COVID-19. Blavatnik School of Government Working Paper. 2020. URL: https://www.bsg.ox.ac.uk/sites/default/files/2020-12/BSG-WP-2020-032-v10.pdf [accessed 2021-04-16]
2. Flaxman S, Mishra S, Gandy A, Unwin H, Coupland H, Mellan T. Estimating the number of infections and the impact of non-pharmaceutical interventions on COVID-19 in European countries: technical description update. ArXiv. Preprint posted online on April 23, 2020 [FREE Full text]
3. Hens N, Vranck P, Molenberghs G. The COVID-19 epidemic, its mortality, and the role of non-pharmaceutical interventions. Eur Heart J Acute Cardiovasc Care 2020 Apr;9(3):204-208. [doi: 10.1177/2048872620924922] [Medline: 32352314]
4. Brauner J, Sharma M, Mindermann S, Stephenson A, Gavenčiak T, Johnston D, et al. The effectiveness and perceived burden of nonpharmaceutical interventions against COVID-19 transmission: a modelling study with 41 countries. medRxiv. Preprint posted online on June 02, 2020. [doi: 10.1101/2020.05.28.20116129]
5. Pan A, Liu L, Wang C, Guo H, Hao X, Wang Q, et al. Association of public health interventions with the epidemiology of the COVID-19 outbreak in Wuhan, China. JAMA 2020 May 19;323(19):1915-1923 [FREE Full text] [doi: 10.1001/jama.2020.6130] [Medline: 32275295]
6. Chang SL, Harding N, Zachreson C, Cliff OM, Prokopenko M. Modelling transmission and control of the COVID-19 pandemic in Australia. ArXiv. Preprint posted online on March 23, 2020 [FREE Full text]
7. Davies N, Kucharski A, Eggo R, Gimma A, Edmunds W, Jombart T, et al. Effects of non-pharmaceutical interventions on COVID-19 cases, deaths, and demand for hospital services in the UK: a modelling study. Lancet Public Health 2020 Jul;5(7):e375-e385. [doi: 10.1016/S2468-2667(20)30133-X]
8. Cowling B, Ali S, Ng T, Tsang T, Li J, Fong M, et al. Impact assessment of non-pharmaceutical interventions against coronavirus disease 2019 and influenza in Hong Kong: an observational study. Lancet Public Health 2020 May;5(5):e279-e288. [doi: 10.1016/s2468-2667(20)30090-6]
9. Aledort J, Lurie N, Wasserman J, Bozzette S. Non-pharmaceutical public health interventions for pandemic influenza: an evaluation of the evidence base. BMC Public Health 2007 Aug 15;7(1):a. [doi: 10.1186/1471-2458-7-208]

10. Merler S, Ajelli M, Fumanelli L, Gomes M, Piontti A, Rossi L, et al. Spatiotemporal spread of the 2014 outbreak of Ebola virus disease in Liberia and the effectiveness of non-pharmaceutical interventions: a computational modelling analysis. Lancet Infect Dis 2015 Feb;15(2):204-211. [doi: 10.1016/s1473-3099(14)71074-6]
11. Cowling BJ, Fung ROP, Cheng CKY, Fang VJ, Chan KH, Seto WH, et al. Preliminary findings of a randomized trial of non-pharmaceutical interventions to prevent influenza transmission in households. PLoS One 2008 May 07;3(5):e2101 [FREE Full text] [doi: 10.1371/journal.pone.0002101] [Medline: 18461182]
12. Ferguson N, Laydon D, Nedjati-Gilani G, Imai N, Ainslie K, Baguelin M. Report 9: Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand. Imperial College London. 2020 Mar 16. URL: https://www.imperial.ac.uk/media/imperial-college/medicine/sph/ide/gida-fellowships/Imperial-College-COVID19-NPI-modelling-16-03-2020.pdf [accessed 2021-04-16]
13. Pinter G, Felde I, Mosavi A, Ghamisi P, Gloaguen R. COVID-19 pandemic prediction for Hungary; a hybrid machine learning approach. Mathematics 2020 Jun 02;8(6):890. [doi: 10.3390/math8060890]
14. Malki Z, Atlam E, Hassanien A, Dagnew G, Elhosseini M, Gad I. Association between weather data and COVID-19 pandemic predicting mortality rate: machine learning approaches. Chaos Solitons Fractals 2020 Sep;138:110137 [FREE Full text] [doi: 10.1016/j.chaos.2020.110137] [Medline: 32834583]
15. Liu D, Clemente L, Poirier C, Ding X, Chinazzi M, Davis J. A machine learning methodology for real-time forecasting of the 2019-2020 COVID-19 outbreak using Internet searches, news alerts, and estimates from mechanistic models. ArXiv. Preprint posted online on April 8, 2020 [FREE Full text]
16. Saba T, Abunadi I, Shahzad MN, Khan AR. Machine learning techniques to detect and forecast the daily total COVID-19 infected and deaths cases under different lockdown types. Microsc Res Tech 2021 Feb 01 [FREE Full text] [doi: 10.1002/jemt.23702] [Medline: 33522669]
17. Ibanez A, Sisodia G. The role of culture on 2020 SARS-CoV-2 country deaths: a pandemic management based on cultural dimensions. GeoJournal 2020 Sep 30:1-17 [FREE Full text] [doi: 10.1007/s10708-020-10306-0] [Medline: 33020679]
18. Huynh TLD. Does culture matter social distancing under the COVID-19 pandemic? Saf Sci 2020 Oct;130:104872. [doi: 10.1016/j.ssci.2020.104872] [Medline: 32550745]
19. Wang Y. Government policies, national culture and social distancing during the first wave of the COVID-19 pandemic: International evidence. Saf Sci 2021 Mar;135:105138. [doi: 10.1016/j.ssci.2020.105138]
20. McCoy L, Smith J, Anchuri K, Berry I, Pineda J, Harish V, COVID-19 Canada Open Data Working Group: Non-Pharmaceutical Interventions, et al. CAN-NPI: a curated open dataset of Canadian non-pharmaceutical interventions in response to the global COVID-19 pandemic. medRxiv. Preprint posted online on April 22, 2020. [doi: 10.1101/2020.04.17.20068460]
21. Lai S, Ruktanonchai N, Zhou L, Prosper O, Luo W, Floyd J, et al. Effect of non-pharmaceutical interventions for containing the COVID-19 outbreak in China. medRxiv. Preprint posted online on March 06, 2020 [FREE Full text] [doi: 10.1101/2020.03.03.20029843] [Medline: 32511601]
22. Delamater PL, Street EJ, Leslie TF, Yang YT, Jacobsen KH. Complexity of the basic reproduction number (R). Emerg Infect Dis 2019 Jan;25(1):1-4 [FREE Full text] [doi: 10.3201/eid2501.171901] [Medline: 30560777]
23. Bjørnstad ON, Finkenstädt BF, Grenfell BT. Dynamics of measles epidemics: estimating scaling of transmission rates using a time series SIR model. Ecological Monographs 2002 May;72(2):169-184. [doi: 10.1890/0012-9615(2002)072[0169:domees]2.0.co;2]
24. Kermack WO, McKendrick AG. A contribution to the mathematical theory of epidemics. Proc R Soc Lond A 1997 Aug 01;115(772):700-721. [doi: 10.1098/rspa.1927.0118]
25. Nåsell I. The quasi-stationary distribution of the closed endemic sis model. Adv Appl Probab 2016 Jul 1;28(03):895-932. [doi: 10.1017/s0001867800046541]
26. Li MY, Muldowney JS. Global stability for the SEIR model in epidemiology. Math Biosci 1995 Feb;125(2):155-164. [doi: 10.1016/0025-5564(95)92756-5]
27. Abrams D, Grant P. Testing the social identity relative deprivation (SIRD) model of social change: the political rise of Scottish nationalism. Br J Soc Psychol 2012 Dec;51(4):674-689. [doi: 10.1111/j.2044-8309.2011.02032.x] [Medline: 21649666]
28. Chen Y, Lu P, Chang C, Liu T. A time-dependent SIR model for COVID-19 with undetectable infected persons. IEEE Trans Netw Sci Eng 2020 Oct 1;7(4):3279-3294. [doi: 10.1109/tnse.2020.3024723]
29. Calafiore G, Novara C, Possieri C. A modified SIR model for the COVID-19 contagion in Italy. 2020 Presented at: 59th IEEE Conference on Decision and Control (CDC); December 14-18, 2020; Jeju Island, South Korea. [doi: 10.1109/cdc42340.2020.9304142]
30. Yang Z, Zeng Z, Wang K, Wong S, Liang W, Zanin M, et al. Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions. J Thorac Dis 2020 Mar;12(3):165-174. [doi: 10.21037/jtd.2020.02.64] [Medline: 32274081]
31. Fernández-Villaverde J, Jones C. Estimating and simulating a SIRD model of COVID-19 for many countries, states, and cities. NBER Working Paper Series. 2020 May. URL: https://www.nber.org/system/files/working_papers/w27128/w27128.pdf [accessed 2021-04-16]

32. Shuja J, Alanazi E, Alasmary W, Alashaikh A. COVID-19 open source data sets: a comprehensive survey. Appl Intell 2020 Sep 21;51(3):1296-1325. [doi: 10.1007/s10489-020-01862-6]
33. Mohamadou Y, Halidou A, Kapen PT. A review of mathematical modeling, artificial intelligence and datasets used in the study, prediction and management of COVID-19. Appl Intell 2020 Jul 06;50(11):3913-3925. [doi: 10.1007/s10489-020-01770-9]
34. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. Lancet Infect Dis 2020 May;20(5):533-534. [doi: 10.1016/s1473-3099(20)30120-1]
35. Desvars-Larrive A, Dervic E, Haug N, Niederkrotenthaler T, Chen J, Di Natale A, et al. A structured open dataset of government interventions in response to COVID-19. Sci Data 2020 Aug 27;7(1):285 [FREE Full text] [doi: 10.1038/s41597-020-00609-9] [Medline: 32855430]
36. Cheng C, Barceló J, Hartnett AS, Kubinec R, Messerschmidt L. COVID-19 government response event dataset (CoronaNet v.1.0). Nat Hum Behav 2020 Jul 23;4(7):756-768. [doi: 10.1038/s41562-020-0909-7] [Medline: 32576982]
37. Killeen B, Wu J, Shah K, Zapaishchykova A, Nikutta P, Tamhane A, et al. A county-level dataset for informing the United States' response to COVID-19. ArXiv. Preprint posted online on April 1, 2020 [FREE Full text]
38. Zarei K, Farahbakhsh R, Crespi N, Tyson G. A first Instagram dataset on COVID-19. ArXiv. Preprint posted online on April 25, 2020 [FREE Full text]
39. Haouari F, Hasanain M, Suwaileh R, Elsayed T. arcCOV-19: The first Arabic COVID-19 Twitter dataset with propagation networks. ArXiv. Preprint posted online on April 13, 2020 [FREE Full text]
40. Qazi U, Imran M, Ofli F. GeoCoV19: a dataset of hundreds of millions of multilingual COVID-19 tweets with location information. SIGSPATIAL Spec 2020 Jun 05;12(1):6-15. [doi: 10.1145/3404820.3404823]
41. Memon S, Carley K. Characterizing COVID-19 misinformation communities using a novel Twitter dataset. ArXiv. Preprint posted online on August 3, 2020 [FREE Full text]
42. Wang L, Lo K, Chandrasekhar Y, Reas R, Yang J, Eide D, et al. CORD-19: the COVID-19 Open Research Dataset. ArXiv. Preprint posted online on April 22, 2020 [FREE Full text]
43. Chen Q, Allot A, Lu Z. LitCovid: an open database of COVID-19 literature. Nucleic Acids Res 2021 Jan 08;49(D1):D1534-D1540 [FREE Full text] [doi: 10.1093/nar/gkaa952] [Medline: 33166392]
44. Guo X, Mirzaalian H, Sabir E, Jaiswal A, Abd-Almageed W. CORD19STS: COVID-19 semantic textual similarity dataset. ArXiv. Preprint posted online on July 5, 2020 [FREE Full text]
45. Pepe E, Bajardi P, Gauvin L, Privitera F, Lake B, Cattuto C, et al. COVID-19 outbreak response, a dataset to assess mobility changes in Italy following national lockdown. Sci Data 2020 Jul 08;7(1):230 [FREE Full text] [doi: 10.1038/s41597-020-00575-2] [Medline: 32641758]
46. Ribeiro-Dantas MDC, Alves G, Gomes RB, Bezerra LC, Lima L, Silva I. Dataset for country profile and mobility analysis in the assessment of the COVID-19 pandemic. Data in Brief 2020 Aug;31:105698. [doi: 10.1016/j.dib.2020.105698]
47. Barbieri DM, Lou B, Passavanti M, Hui C, Lessa DA, Maharaj B, et al. A survey dataset to evaluate the changes in mobility and transportation due to COVID-19 travel restrictions in Australia, Brazil, China, Ghana, India, Iran, Italy, Norway, South Africa, United States. Data Brief 2020 Dec;33:106459 [FREE Full text] [doi: 10.1016/j.dib.2020.106459] [Medline: 33163599]
48. Kang Y, Gao S, Liang Y, Li M, Rao J, Kruse J. Multiscale dynamic human mobility flow dataset in the U.S. during the COVID-19 epidemic. Sci Data 2020 Nov 12;7(1):390 [FREE Full text] [doi: 10.1038/s41597-020-00734-5] [Medline: 33184280]
49. Afshar P, Heidarian S, Enshaei N, Naderkhani F, Rafiee M, Oikonomou A, et al. COVID-CT-MD: COVID-19 computed tomography (CT) scan dataset applicable in machine learning and deep learning. ArXiv. Preprint posted online on September 28, 2020 [FREE Full text]
50. Cohen J, Morrison P, Dao L, Roth K, Duong T, Ghassemi M. Covid-19 image data collection: prospective predictions are the future. ArXiv. Preprint posted online on June 22, 2020 [FREE Full text]
51. Yang X, He X, Zhao J, Zhang Y, Zhang S, Xie P. COVID-CT-Dataset: a CT image dataset about COVID-19. ArXiv. Preprint posted online on March 30, 2020 [FREE Full text]
52. Born J, Brändle G, Cossio M, Disdier M, Goulet J, Roulin J, et al. POCOVID-Net: automatic detection of COVID-19 from a new lung ultrasound imaging dataset (POCUS). ArXiv. Preprint posted online on April 25, 2020 [FREE Full text]
53. Bavel JJV, Baicker K, Boggio P, Capraro V, Cichocka A, Cikara M, et al. Using social and behavioural science to support COVID-19 pandemic response. Nat Hum Behav 2020 May;4(5):460-471. [doi: 10.1038/s41562-020-0884-z] [Medline: 32355299]
54. Dryhurst S, Schneider CR, Kerr J, Freeman ALJ, Recchia G, van der Bles AM, et al. Risk perceptions of COVID-19 around the world. J Risk Res 2020 May 05;23(7-8):994-1006. [doi: 10.1080/13669877.2020.1758193]
55. Guan Y, Deng H, Zhou X. Understanding the impact of the COVID-19 pandemic on career development: insights from cultural psychology. J Vocat Behav 2020 Jun;119:103438 [FREE Full text] [doi: 10.1016/j.jvb.2020.103438] [Medline: 32382162]
56. Furlong Y, Finnie T. Culture counts: the diverse effects of culture and society on mental health amidst COVID-19 outbreak in Australia. Ir J Psychol Med 2020 Sep;37(3):237-242 [FREE Full text] [doi: 10.1017/ipm.2020.37] [Medline: 32406358]

57. Ashraf B. Stock markets' reaction to COVID-19: Cases or fatalities? Res Int Bus Finance 2020 Dec;54:101249. [doi: 10.1016/j.ribaf.2020.101249]
58. Hofstede G, Bond M. Hofstede's Culture Dimensions. J Cross Cult Psychol 2016 Jul 27;15(4):417-433. [doi: 10.1177/0022002184015004003]
59. House R, Hanges P, Ruiz-Quintanilla S, Dorfman P, Javidan M, Dickson M. Cultural influences on leadership and organizations: Project Globe. In: Advances in Global Leadership. Bingley, UK: Emerald Group Publishing Ltd; 1999:171.
60. Yoo B, Donthu N, Lenartowicz T. Measuring Hofstede's five dimensions of cultural values at the individual level: development and validation of CVSCALE. J Int Consum Mark 2011:193. [doi: 10.1080/08961530.2011.578059]
61. Dahl S. Intercultural research: the current state of knowledge. SSRN Journal 2004 Feb 02. [doi: 10.2139/ssrn.658202]
62. Hofstede G. Dimension data matrix. GeertHoftstede. 2015. URL: https://geerthofstede.com/research-and-vsm/dimension-data-matrix/ [accessed 2021-04-16]
63. Alimadadi A, Aryal S, Manandhar I, Munroe PB, Joe B, Cheng X. Artificial intelligence and machine learning to fight COVID-19. Physiol Genomics 2020 Apr 01;52(4):200-202 [FREE Full text] [doi: 10.1152/physiolgenomics.00029.2020] [Medline: 32216577]
64. Yan L, Zhang H, Xiao Y, Wang M, Sun C, Tang X, et al. Prediction of criticality in patients with severe Covid-19 infection using three clinical features: a machine learning-based prognostic model with clinical data in Wuhan. medRxiv. Preprint posted online on March 03, 2020. [doi: 10.1101/2020.02.27.20028027]
65. Randhawa G, Soltysiak M, El Roz H, de Souza C, Hill K, Kari L. Machine learning using intrinsic genomic signatures for rapid classification of novel pathogens: COVID-19 case study. PLoS ONE 2020 Apr 24;15(4):e0232391. [doi: 10.1371/journal.pone.0232391]
66. Harmon SA, Sanford TH, Xu S, Turkbey EB, Roth H, Xu Z, et al. Artificial intelligence for the detection of COVID-19 pneumonia on chest CT using multinational datasets. Nat Commun 2020 Aug 14;11(1):4080 [FREE Full text] [doi: 10.1038/s41467-020-17971-2] [Medline: 32796848]
67. Lauer SA, Grantz KH, Bi Q, Jones FK, Zheng Q, Meredith HR, et al. The Incubation Period of Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed Cases: Estimation and Application. Ann Intern Med 2020 May 05;172(9):577-582. [doi: 10.7326/m20-0504]
68. Zhu N, Lu H, Chang L. Debate: facing uncertainty with(out) a sense of control - cultural influence on adolescents' response to the COVID-19 pandemic. Child Adolesc Ment Health 2020 Sep;25(3):173-174 [FREE Full text] [doi: 10.1111/camh.12408] [Medline: 32681578]
69. McKinney W. pandas: a foundational Python library for data analysis and statistics. 2011 Presented at: SC11: Python for High Performance and Scientific Computing; November 18, 2011; Seattle, WA URL: https://www.dlr.de/sc/portaldata/15/resources/dokumente/pyhpc2011/submissions/pyhpc2011_submission_9.pdf
70. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O. Scikit-learn: machine learning in Python. J Mach Learn Res 2011:2825 [FREE Full text]
71. Hofstede G. Dimensionalizing cultures: the Hofstede model in context. ORPC 2011 Dec 01;2(1). [doi: 10.9707/2307-0919.1014]
72. Soares AM, Farhangmehr M, Shoham A. Hofstede's dimensions of culture in international marketing studies. J Bus Res 2007 Mar;60(3):277-284. [doi: 10.1016/j.jbusres.2006.10.018]
73. Deschepper R, Grigoryan L, Lundborg CS, Hofstede G, Cohen J, Kelen GVD, et al. Are cultural dimensions relevant for explaining cross-national differences in antibiotic use in Europe? BMC Health Serv Res 2008 Jun 06;8(1):123 [FREE Full text] [doi: 10.1186/1472-6963-8-123] [Medline: 18538009]
74. Borg MA. National cultural dimensions as drivers of inappropriate ambulatory care consumption of antibiotics in Europe and their relevance to awareness campaigns. J Antimicrob Chemother 2012 Mar 26;67(3):763-767. [doi: 10.1093/jac/dkr541] [Medline: 22200725]
75. Borg M. Lowbury Lecture 2013. Cultural determinants of infection control behaviour: understanding drivers and implementing effective change. J Hosp Infect 2014 Mar;86(3):161-168. [doi: 10.1016/j.jhin.2013.12.006] [Medline: 24534705]
76. Masood M, Aggarwal A, Reidpath D. Effect of national culture on BMI: a multilevel analysis of 53 countries. BMC Public Health 2019 Sep 03;19(1):1212 [FREE Full text] [doi: 10.1186/s12889-019-7536-0] [Medline: 31481044]
77. Ross BC. Mutual information between discrete and continuous data sets. PLoS One 2014 Feb 19;9(2):e87357 [FREE Full text] [doi: 10.1371/journal.pone.0087357] [Medline: 24586270]
78. Pirbazari A, Chakravorty A, Rong C. Evaluating feature selection methods for short-term load forecasting. 2019 Presented at: 2019 IEEE International Conference on Big Data and Smart Computing (BigComp); February 27-March 2, 2019; Kyoto, Japan p. 1-8. [doi: 10.1109/bigcomp.2019.8679188]
79. Hoerl A, Kennard R. Ridge regression:biased estimation for nonorthogonal problems. Technometrics 1970 Feb;12(1):55-67. [doi: 10.1080/00401706.1970.10488634]
80. Breiman L. Random forests. Mach Learn 2001;45(1):32. [doi: 10.1023/A:1010933404324]
81. Freund Y, Schapire RE. A decision-theoretic generalization of on-line learning and an application to boosting. J Comput Syst Sci 1997 Aug;55(1):119-139. [doi: 10.1006/jcss.1997.1504]

82. Drucker H, Burges C, Kaufman L, Smola A, Vapnik V. Support vector regression machines. In: NIPS'96: Proceedings of the 9th International Conference on Neural Information Processing Systems. 1996 Dec Presented at: 9th International Conference on Neural Information Processing Systems; December 2-5, 1996; Denver, CO p. 155-161.
83. Willmott C, Matsuura K. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. Clim Res 2005;30:79-82. [doi: 10.3354/cr030079]
84. Steyerberg E, Vickers A, Cook N, Gerds T, Gonen M, Obuchowski N, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. Epidemiology 2010 Jan;21(1):128-138 [FREE Full text] [doi: 10.1097/EDE.0b013e3181c30fb2] [Medline: 20010215]
85. Schaffer C. Selecting a classification method by cross-validation. Mach Learn 1993 Oct;13(1):135-143. [doi: 10.1007/bf00993106]
86. Browne MW. Cross-validation methods. J Math Psychol 2000 Mar;44(1):108-132. [doi: 10.1006/jmps.1999.1279] [Medline: 10733860]
87. Ardabili S, Mosavi A, Ghamisi P, Ferdinand F, Varkonyi-Koczy A, Reuter U. COVID-19 outbreak prediction with machine learning. PsyArXiv. Preprint posted online on October 06, 2020. [doi: 10.31234/osf.io/5dyfc]
88. Mukaka M. Statistics corner: A guide to appropriate use of correlation coefficient in medical research. Malawi Med J 2012 Sep;24(3):69-71 [FREE Full text] [Medline: 23638278]
89. Moore DS, Kirkland S. The Basic Practice of Statistics. New York, NY: WH Freeman; 2007.
90. Quionero-Candela J, Sugiyama M, Schwaighofer A, Lawrence N. Dataset Shift in Machine Learning. Cambridge, MA: The MIT Press; 2009.

Abbreviations

CIG: confirmed infection growth
CSSE: Center for Systems Science and Engineering
CVSCALE: Cultural Value Scale
GLOBE: Global Leadership and Organizational Effectiveness
MAE: mean absolute error
MSE: mean squared error
NPI: nonpharmaceutical intervention
OxCGRT: Oxford COVID-19 Government Response Tracker
SIR: susceptible-infected-recovered

Edited by C Basch; submitted 18.12.20; peer-reviewed by S Rostam Niakan Kalhori, P Sarajlic, S Lalmuanawma; comments to author 12.02.21; revised version received 05.03.21; accepted 23.03.21; published 23.04.21.

Please cite as:
Yeung AYS, Roewer-Despres F, Rosella L, Rudzicz F
Machine Learning–Based Prediction of Growth in Confirmed COVID-19 Infection Cases in 114 Countries Using Metrics of Nonpharmaceutical Interventions and Cultural Dimensions: Model Development and Validation
J Med Internet Res 2021;23(4):e26628
URL: https://www.jmir.org/2021/4/e26628
doi:10.2196/26628
PMID:33844636

Original Paper

Patterns of Media Use, Strength of Belief in COVID-19 Conspiracy Theories, and the Prevention of COVID-19 From March to July 2020 in the United States: Survey Study

Daniel Romer[1], BA, DPhil; Kathleen Hall Jamieson[1], BA, DPhil

Annnenberg Public Policy Center, University of Pennsylvania, Philadelphia, PA, United States

Corresponding Author:
Daniel Romer, BA, DPhil
Annnenberg Public Policy Center
University of Pennsylvania
202 S 36th St
Philadelphia, PA, 19104
United States
Phone: 1 610 202 7315
Fax: 1 215 746 2074
Email: dan.romer@appc.upenn.edu

Abstract

Background: Holding conspiracy beliefs regarding the COVID-19 pandemic in the United States has been associated with reductions in both actions to prevent the spread of the infection (eg, mask wearing) and intentions to accept a vaccine when one becomes available. Patterns of media use have also been associated with acceptance of COVID-19 conspiracy beliefs. Here we ask whether the type of media on which a person relies increased, decreased, or had no additional effect on that person's COVID-19 conspiracy beliefs over a 4-month period.

Objective: We used panel data to explore whether use of conservative and social media in the United States, which were previously found to be positively related to holding conspiracy beliefs about the origins and prevention of COVID-19, were associated with a net increase in the strength of those beliefs from March to July of 2020. We also asked whether mainstream news sources, which were previously found to be negatively related to belief in pandemic-related conspiracies, were associated with a net decrease in the strength of such beliefs over the study period. Additionally, we asked whether subsequent changes in pandemic conspiracy beliefs related to the use of media were also related to subsequent mask wearing and vaccination intentions.

Methods: A survey that we conducted with a national US probability sample in March of 2020 and again in July with the same 840 respondents assessed belief in pandemic-related conspiracies, use of various types of media information sources, actions taken to prevent the spread of the disease and intentions to vaccinate, and various demographic characteristics. Change across the two waves was analyzed using path analytic techniques.

Results: We found that conservative media use predicted an increase in conspiracy beliefs (β=.17, 99% CI .10-.25) and that reliance on mainstream print predicted a decrease in their belief (β=–.08, 99% CI –.14 to –.02). Although many social media platforms reported downgrading or removing false or misleading content, ongoing use of such platforms by respondents predicted growth in conspiracy beliefs as well (β=.072, 99% CI .018-.123). Importantly, conspiracy belief changes related to media use between the two waves of the study were associated with the uptake of mask wearing and changes in vaccination intentions in July. Unlike other media, use of mainstream broadcast television predicted greater mask wearing (β=.17, 99% CI .09-.26) and vaccination intention (β=.08, 95% CI .02-.14), independent of conspiracy beliefs.

Conclusions: The findings point to the need for greater efforts on the part of commentators, reporters, and guests on conservative media to report verifiable information about the pandemic. The results also suggest that social media platforms need to be more aggressive in downgrading, blocking, and counteracting claims about COVID-19 vaccines, claims about mask wearing, and conspiracy beliefs that have been judged problematic by public health authorities.

(J Med Internet Res 2021;23(4):e25215) doi:10.2196/25215

KEYWORDS

COVID-19; conspiracy beliefs; social media; print news media; broadcast news media; conservative media; vaccination; mask wearing; belief; misinformation; infodemic; United States; intention; prevention

Introduction

At the outset of the COVID-19 pandemic in the United States, various conspiracy theories about the origins and prevention of COVID-19 began to circulate on social media and some conservative media outlets [1-4]. The study of conspiracies as explanations for major events achieved notice in Hofstadter's classic and influential 1966 volume *The Paranoid Style in American Politics* [5]. In his analysis, conspiracy beliefs presupposed a "vast, insidious, preternaturally effective international conspiratorial network designed to perpetrate acts of most fiendish character" [5]. Conspiracy beliefs, including those claiming the US government's responsibility for the assassination of John F Kennedy and the 9/11 terrorist attack [6,7], have been a focus of study in the political science literature, which regards such beliefs as ones "in which the ultimate cause of an event is believed to be due to a plot by multiple actors working together with a clear goal in mind, often unlawfully and in secret" [8]. In psychology, the focus has been on understanding what draws individuals to conspiracies [7,9], while in public health, the concern has been about their role in minimizing the likelihood of preventive behavior [10], and their creation of unfounded fears of interventions such as fluoridation and vaccination [11]. As with all conspiracy theories, it is difficult to determine the validity of those related to COVID-19 because the putative actors work in secret [12]. Although stigmatized as paranoid by some [5], such beliefs have a surprising ability to attract adherents [6], and their influence has increasingly been observed in response to COVID-19 public health recommendations, such as vaccination and social distancing [13,14].

A notable characteristic of conspiracy beliefs is the tendency for belief in any one to be associated with acceptance of others [2,15-17]. In the United States, three such beliefs prevalent early in the pandemic [1,2,4] concerned suspicions that the pandemic was the result of malign actions by either the Chinese government or the pharmaceutical industry or that some in the US government were exaggerating the danger of COVID-19 to undermine the president of the United States. A national probability sample of the US population in March and again in July of 2020 [18] found that belief in any one of the conspiracies was highly related to belief in the others and that those beliefs were stable over time. Furthermore, belief in a composite of the three conspiracies in March predicted unwillingness in July to obtain a vaccine for the virus should one become available. The beliefs also predicted a lower likelihood of reporting wearing a face mask outside the home when exposed to other people [18]. Although belief in pandemic conspiracies increased from March to July 2020, our earlier analysis did not identify potential sources of that increase or their possible effects on preventive behavior.

Previous research has found that both misinformation [19] and conspiracy beliefs are resistant to change [6,7,20,21] and that holding conspiracy beliefs related to COVID-19 is associated with lower levels of behaviors known to prevent its spread [13,22]. In addition, single cross-sectional studies have found a positive relationship between social media use and COVID-19 misinformation [1] and conspiracy beliefs [14,23,24] and found that mainstream media consumption is associated with greater rejection of them [25,26]. However, such cross-sectional data cannot determine whether persistent use of these sources is related to change in these beliefs across time or whether efforts undertaken between the two surveys by media outlets to decrease the amount of conspiracy content about COVID-19–related topics is associated with a decrease in these beliefs.

Using a longitudinal study design, we tested the possibility that exposure to different types of media sources might be responsible for change in COVID-19 conspiracy beliefs in the United States. Where our earlier work found that conspiracy beliefs were positively related to use of social and conservative media and negatively related to use of mainstream television and print [18], we sought to determine whether those media sources were also associated with subsequent change in the strength of conspiracy beliefs from March to July 2020.

In examining the role of the media, it is important to recognize that different types of outlets follow different norms when communicating information even in ordinary times. According to normative models of news reporting [27,28], journalists draw on reliable, predictable sources, such as government agencies and other accounts that can be independently verified through standard fact-checking procedures. In the case of conspiracy theories, such evidence is not available [29]. At the same time, their imperviousness to disconfirmation reduces the likelihood that mainstream news sources will feature them except to debunk them. However, commentators on conservative cable and talk radio, and those who post on social media are not bound by such conventions [27]. As one example, Tucker Carlson of Fox News noted that there was "a lot of speculation" that COVID-19 "is not a naturally occurring virus; that it was somehow created by the Chinese government" [30]. Rush Limbaugh alleged that "the coronavirus is being weaponized as yet another element to bring down Donald Trump" and that "it probably is a ChiCom laboratory experiment that is in the process of being weaponized" [31].

Misleading information percolated through social media as the pandemic unfolded [14,23]. On March 3, 2020, NewsGuard [32] raised an alarm due to the following finding:

> *Over the last 90 days, posts from the websites of the U.S. Centers for Disease Control and Prevention and the World Health Organization received 364,483 'engagements,' or likes, shares, and comments on social media. In that same period, 74 U.S. sites that NewsGuard found to have published coronavirus misinformation received a combined 52,053,542 'engagements'—more than 142 times the engagement of the two major public health institutions providing information about the outbreak.*

A video on a YouTube channel named the Next News Network was viewed nearly 7 million times before it was taken down by YouTube [33]. It claimed that the COVID-19 pandemic was the result of a deceptive plot that sought to impose "mandatory vaccines" on the public.

Unlike conservative cable and talk radio, the major social media platforms took active measures to interdict COVID-19–related misinformation and conspiracies by removing misleading and potentially harmful content, inserting warnings where appropriate, and featuring articles that refuted the widespread misinformation. In March 2020, Twitter announced that it would "prioritize removing content when it has a clear call to action that could directly pose a risk to people's health or well-being" [34] and in May announced that it would "put labels and warning messages on some tweets that contain disputed or misleading information related to Covid-19" [35]. In addition, in anticipation of the release of a COVID-19 conspiracy film, "Pinterest had its moderators run proactive searches for terms that might have been associated with the movie, deleting them to nip any problematic content in the bud" [36]. In early March, Facebook announced that it "was removing false claims and conspiracy theories flagged by global health organizations and the company is blocking people from running ads that try to exploit the fears of the public by pitching snake oil cures" [37]. In mid-April, Facebook began "showing messages in News Feed to people who have liked, reacted or commented on harmful misinformation about COVID-19 that we have since removed. These messages will connect people to COVID-19 myths debunked by the WHO including ones we've removed from our platform for leading to imminent physical harm" [38]. At the same time, it announced that it had "added a new section to [its] COVID-19 Information Center called Get the Facts. It includes fact-checked articles from [Facebook's] partners that debunk misinformation about the coronavirus." In August 2020, Facebook reported that it had removed "7 million posts pushing covid-19 misinformation from its main social media site and Instagram between April and June" and "put warning notes on 98 million covid-19 misinformation posts on Facebook" in the same period [39]. In December 2020, YouTube reported that it had removed more than 700,000 misleading COVID-19 videos to date [40].

Despite their efforts, the range of social media and the capacity of misinformation purveyors to repost interdicted content means that conspiracy theories about COVID-19 often gain substantial audiences before they are blocked or addressed [19,38,41,42]. For example, before YouTube removed a video asserting that the pandemic had been bioengineered, 570,000 subscribers to the website SGT Report had potentially been exposed to it [43]. Additionally, the 26-minute viral video, "Plandemic," which was also eventually removed in May 2020, nevertheless claimed that "vaccines kill millions, the flu vaccine contains the coronavirus, and that the virus was 'manipulated'" [44], and was viewed "more than eight million times on YouTube, Facebook, Twitter and Instagram, and had generated countless other posts," in a little over a week after its release [45]. If exposure to those media is partly responsible for the dissemination and credibility of conspiracy beliefs, we would expect that ongoing exposure would intensify the beliefs between March and July 2020. If the social media efforts to blunt the effects of the content were successful, we would expect that the beliefs reported in March by users of those media would not have intensified by July.

We also were interested to see whether mainstream media use predicted declines in conspiracy beliefs, as cross-sectional analyses of the relations between media sources and conspiracy beliefs in March found [1,18]. If the mainstream news media are a source of substantiated information rather than speculation about conspiracies, then one would expect that exposure to them would result in a decline in conspiracy beliefs regarding the pandemic. Such a finding would suggest that news media can play a role in reducing the strength of conspiracy beliefs, despite their resistance to refutation.

To further investigate the role of conspiracy beliefs on preventive behavior, we asked whether increased belief in COVID-19–related conspiracies from March to July 2020 was related to changes in either preventive behavior (mask wearing) or willingness to accept a COVID-19 vaccine when a safe and effective one becomes available. Furthermore, to the extent to which ongoing use of any media sources was associated with strengthened belief in the same pandemic conspiracies, we were also interested to see if patterns of media exposure might not only predict change in conspiracy beliefs, but also whether those changes were associated with changes in vaccination intentions or with uptake of mask wearing.

In summary, we explored whether use of different media that span the liberal to conservative political spectrum, including mainstream media, predicted change in the acceptance of conspiracy theories that were found to be prevalent at the start of the COVID-19 pandemic in the United States [1,2,4]. In particular, we tested our major hypothesis (H1) that use of mainstream news media such as broadcast television and newspapers that abide by journalistic norms of reporting would predict reduction in conspiracy belief. With regard to social media, which are not constrained by such norms and have been found to be associated with belief in conspiracy theories about the pandemic [14,23], we explored whether their use might have strengthened conspiracy beliefs or alternatively, consistent with efforts by major platforms to remove misinformation and conspiracies, have reduced the strength of those beliefs over time. Our prior analysis of the role of conservative media suggested our second major hypothesis (H2) that use of conservative media would strengthen conspiracy beliefs. Finally, to the extent that media use predicted continued growth or decline in conspiracy beliefs, we asked whether such changes were associated with changes in accepting a vaccine when one became available or with adopting the newly recommended preventive action of mask wearing.

Methods

Survey Sample

A sample of US residents (N=840) that was recruited from a national probability panel by Qualtrics completed two waves of an online survey as part of their participation in NORC (National Opinion Research Center) at the University of

Chicago's AmeriSpeak Panel [46]. The first survey was conducted in March 2020 and the second approximately four months later in July. We restricted our analysis to the respondents who completed both surveys (see [18]). The sample that remained at the second wave was very similar to that from the first wave and missingness was only slightly related to intention to vaccinate at wave 1. The study was deemed to be exempt from institutional review board review in as much as no personally identifiable information was retained from the survey firm.

NORC also provided demographic survey weights to enable projection to the US population according to age, gender, race/ethnicity, education, and Census Division based on the Current Population Survey of February 2020. Those weights were applied for descriptive purposes in describing belief in conspiracy theories, but all multivariate analyses were conducted with unweighted data and demographic differences controlled.

A power analysis conducted prior to the study indicated that a sample size of approximately 800 would enable us to detect standardized mediated relationships of .04 or greater at the 99% CI [18], which was regarded as sufficiently sensitive for direct relationships as well.

Survey Content

Conspiracy Beliefs

We assessed belief in three conspiracy theories circulating in social media and other venues at the time of the first wave [47], which were found to be accepted by at least 10% of the US population [1,2,4]:

1. "The pharmaceutical industry created the coronavirus to increase sales of its drugs and vaccines."
2. "The coronavirus was created by the Chinese government as a biological weapon."
3. "Some in the U.S. Centers for Disease Control and Prevention, also known as CDC, are exaggerating the danger posed by the coronavirus to damage the Trump presidency."

Belief in each conspiracy theory was registered on a 4-point scale ranging from "Definitely false" to Definitely true." Belief in conspiracies is distinguishable from endorsement of other forms of misinformation (eg, that taking vitamin C protects one from contracting COVID-19) that were also prevalent during the early phase of the pandemic in the United States [48]. We operationalized strengthening of conspiracy belief as a movement toward either "definitely" or "probably true" and weakening as movement toward "definitely false." These ratings were correlated within respondents with values ranging from .40 to .54 at both waves and high levels of reliability (α values=.71 and .74 and ω values=.67 and .74). The mean of these ratings increased from 1.75 (SD 0.85) to 1.90 (SD 1.08), $P<.001$, at the second wave. Belief in the conspiracies in March was inversely related to taking preventive action and accepting an eventual vaccine [18]. In addition, conspiracy belief in March prospectively predicted both action taken and change in vaccination intention assessed in July. In the analysis reported here, we determined whether change in conspiracy belief associated with media use was also related to prospective prediction of those outcomes.

Mask Wearing and Other Preventive Actions

We also developed an index encapsulating adoption of seven recommended actions to prevent the infection at wave 1 (eg, frequent hand washing) [18]. The index was the sum of the actions taken in the past few days (Yes or No). This index did not include mask wearing, which was not recommended until after the first wave was completed. At the second wave, we used this outcome as a measure of compliance with recommended action since it had become recognized as critical to halting the spread of the virus through the air as well as preventing contact between the hands and face [49]. At the second wave, 79% (n=668) of the sample reported wearing a mask every day they went to public places where they "might encounter other people." The two outcomes were positively correlated across the two waves (r=.23), indicating that the index at wave 1 was sensitive to willingness to comply with recommended behavior, and so we used the index at wave 1 to assess change in this tendency over time.

Vaccination Intentions

We also assessed intentions to accept a vaccine should one become available in the future with a 4-point scale going from "not at all likely" to "very likely." The proportion who reported either not at all likely or not very likely increased from 14.5% (n=121) at the first wave to 25.8% (n=215) at the second wave, $P<.001$.

Media Use

To determine whether media use predicted change across time in conspiracy beliefs, we assessed reliance on six sources of news and commentary in the media from across the political spectrum [11]. Our use of conservative, mainstream, and liberal media categories was consistent with classifications that are commonly used to categorize media in the United States [27,50-52].

For each source, respondents rated "How much information do you get from sources such as…" on a 6-point scale from "no information" to "a lot of information" (see Table 1 for each list of sources). The modal response was "no information" except for mainstream broadcast television news, for which the mode was the highest level of information. We refer to the use of major national newspapers and news services that have users both online and in traditional print as mainstream print. As seen in Table 1, the reported amount of information that respondents received from each source tended to decline from March to July and did so reliably for all but liberal news and aggregators. Nevertheless, individual use of each media category was highly stable across time, as evidenced by the correlations between the two time points. In addition, correlations between use of each media and belief in COVID-19 conspiracies ranged from a high of .37 for conservative media to –.28 for mainstream print.

Table 1. Use of six media (with examples of each) as sources of information in both March (T1) and July (T2) of 2020 and relations with political ideology (unweighted).

Media use	March 2020 (T1), mean (SD)	July 2020 (T2), mean (SD)	*t* value of paired difference, T1-T2	Probability of *t* value	Correlation over time	Correlation with T1 Conspiracy Belief Index
Mainstream television (eg, ABC News, CBS News, NBC News)	2.94 (1.77)	2.61 (1.73)	6.13	<.001	.62	–.18
Conservative media (eg, Fox News, Rush Limbaugh, Breitbart News, One America News, The Drudge Report)	1.49 (1.83)	1.39 (1.74)	2.19	.029	.71	.37
Liberal media (eg, MSNBC, Bill Maher, Huffington Post	1.35 (1.53)	1.31 (1.51)	0.69	.49	.57	–.10
Aggregators (eg, Google News, Yahoo News)	2.00 (1.63)	1.94 (1.59)	1.22	.22	.54	.06
Social media (eg, Facebook, Twitter, YouTube)	1.80 (1.71)	1.67 (1.63)	2.23	.026	.54	.16
Mainstream print (eg, Associated Press, New York Times, Washington Post, Wall Street Journal)	1.85 (1.75)	1.80 (1.71)	2.26	.026	.61	–.28

We controlled for the partisanship of the respondent as an alternative explanation for effects, using a measure of political ideology registered on a 5-point scale ranging from "very conservative" to "very liberal." As expected, self-identified liberals reported greater use of liberal media (r=.30), while conservatives reported greater use of conservative media (r=–.42).

We also assessed various demographic differences that could be associated with media exposure and therefore might explain changes between the two waves in relation to conspiracy beliefs and media use. The relations between these characteristics and belief in coronavirus conspiracies are described in the Results section. The text of the survey is located in Multimedia Appendix 1.

Analysis

We used the program Mplus [53] to test path models for the relation between media use in March 2020 and changes in conspiracy beliefs in July 2020. For any media use that predicted such change, we examined whether the change in conspiracy belief mediated changes in either preventive action (mask wearing) or intention to vaccinate. Less than 3% of the data was missing for any analysis, and Mplus imputed those scores using maximum likelihood estimation. We used bootstrap procedures with 1000 samples to construct 99% and 95% CIs for all tests of direct and mediated paths. We report standardized coefficients for all paths in the models that had CIs excluding zero. To avoid overfitting, paths with values that did not lie outside of at least 90% CIs were dropped from the models and fixed at a value of zero. We used standard measures of goodness of fit for all models [54].

Results

Overview

Table 2 shows the percentages of respondents at each wave who reported belief in the three conspiracy theories by various demographic characteristics and media use (weighted to national demographics). Overall, each belief received greater endorsement at the second wave, and this occurred across most of the demographic and media use characteristics. In total, belief that China had created the virus as a weapon was the most widely endorsed (38.1%), with belief that some within the Centers for Disease Control and Prevention (CDC) were using the pandemic to undermine the president was close behind (32.2%). The belief that the pharmaceutical industry was benefiting from the pandemic and possibly helped to create it was the least accepted (17.4%).

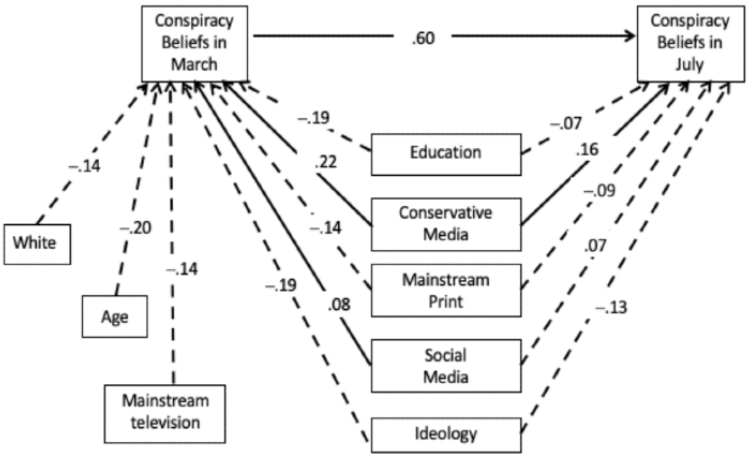
Table 2. Percentages of sample believing that each conspiracy belief was either "definitely true" or "probably true" by demographic characteristics and news use in March and July 2020 (N=840, weighted).

Characteristic	Percentage of sample	Pharmaceutical industry created the virus, %		Centers for Disease Control and Prevention wants to damage the Trump presidency, %		Chinese government created the virus, %	
		March	July	March	July	March	July
Gender							
Male	44.1	9.6	14.2	27.0[a]	38.2[a]	29.1	37.6[a]
Female	55.6	19.1	19.8	20.3[a]	27.0[a]	27.8	38.2[a]
Age							
18-29	20.7	26.9[a]	24.4[a]	22.7	26.9	35.1[a]	43.0[a]
30-44	23.8	21.9[a]	30.7[a]	30.1	34.9	29.8[a]	36.1[a]
45-59	24.5	10.9[a]	12.9[a]	25.1	36.9	30.3[a]	33.4[a]
≥60	31.0	4.3[a]	5.9[a]	17.1	30.1	22.3[a]	32.0[a]
Race/ethnicity							
White	74.7	9.7[a]	14.2[a]	22.3[a]	35.5[a]	25.5[a]	38.0
Black	13.9	36.3[a]	29.1[a]	31.4[a]	18.4[a]	43.8[a]	48.5[a]
Hispanic	15.4	29.1[a]	28.4[a]	28.6[a]	28.5[a]	33.8[a]	42.9[a]
Education							
High school or less	32.9	27.0[a]	29.0[a]	31.8[a]	42.3[a]	42.7[a]	58.1[a]
Some college	48.2	12.1[a]	14.0[a]	21.7[a]	30.7[a]	24.0[a]	32.8[a]
Postgraduate	19.0	6.9[a]	6.4[a]	12.1[a]	17.9[a]	12.7[a]	16.7[a]
Income (US $)							
<30,000	26.4	27.4[a]	22.9[a]	27.2	24.0[a]	37.0[a]	46.8[a]
30,000-85,000	41.1	13.7[a]	22.0[a]	21.5	36.9[a]	28.5[a]	39.9[a]
>85,000	32.4	5.6[a]	7.1[a]	22.5	33.1[a]	21.4[a]	28.5[a]
Political ideology							
Conservative	29.8	11.1[a]	17.1[a]	40.2[a]	63.1[a]	37.1[a]	54.8[a]
Neither	40.0	16.4[a]	17.2[a]	22.6[a]	23.6[a]	30.6[a]	39.9[a]
Liberal	30.2	16.2[a]	17.5[a]	8.0[a]	13.5[a]	17.4[a]	19.5[a]
News source							
Mainstream television	41.9	13.3	18.3[a]	15.1[a]	17.7[a]	24.7[a]	33.2[a]
Conservative	18.1	13.4	28.2[a]	33.5[a]	61.1[a]	51.6[a]	65.8[a]
Liberal	11.1	19.6	22.7[a]	11.0[a]	17.4[a]	21.8	34.1
Aggregators	20.9	24.2[a]	25.4[a]	25.0[a]	23.6[a]	32.2	43.0[a]
Social	19.6	19.8[a]	27.1[a]	28.6	32.7	41.8[a]	44.7
Mainstream print	21.0	15.0[a]	11.5[a]	13.2[a]	11.6[a]	18.4[a]	21.4[a]
Total	100	14.8	17.4	23.5	32.2	28.3	38.1

[a]Response distributions that are significantly different within each time period ($P<.05$) either across demographic groups or within each media use.

Figure 1 shows the results of the path model that tested our hypotheses regarding mainstream and conservative media as predictors of change in support of pandemic conspiracies and our exploration of social media as a continued influence. The model provided an excellent fit to the data, with a root mean square error of approximation (RMSEA) of .032, 90% CI .000-.068; comparative fit index (CFI)=.99; Tucker-Lewis Index (TLI)=.99, and standardized root mean square residual (SRMR)=.007. As seen in Figure 1, holding constant the relation between conspiracy beliefs across time, our first hypothesis (H1) regarding the effects of mainstream news was only partially supported, while H2 regarding the effects of conservative and social media was fully supported: both conservative (.159, 99% CI .089-.231) and social media use (.072, 99% CI .018-.123) predicted increased belief in the conspiracies. However, while mainstream television and print were inversely related to conspiracy belief in March, only mainstream print predicted a decline in conspiracy belief in July (–.090, 99% CI –.150 to –.029). No other media use predicted subsequent conspiracy belief.

Figure 1. Standardized predictors of conspiracy beliefs in March and again in July with all coefficients falling within 99% CIs, except for social media and conspiracy beliefs in March.

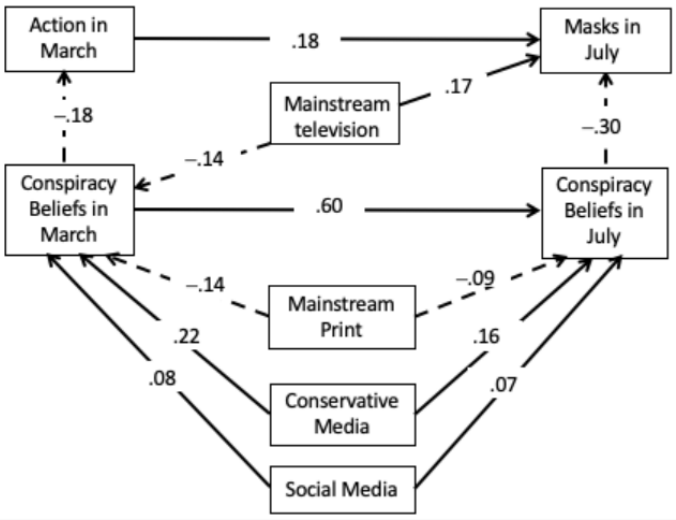


As seen in Figure 1, the media sources remained predictive despite controls for other factors, such as education and liberal political ideology, which predicted declines in conspiracy beliefs.

Relations With Mask Wearing in July

Figure 2 shows the results of the analysis for changes in action taken, with mask wearing specifically assessed at the second wave. This model also fit the data well, with an RMSEA of .036, 90% CI .021-.051; CFI=.98; TLI=.97; and SRMR=.019. Use of masks in July was inversely related to conspiracy belief in July (–.30, 99% CI –.38 to –.22), controlling for demographic differences and political ideology (not shown in the figure). Interestingly, use of mainstream television in March was a positive predictor of mask wearing in July (.17, 99% CI .09-.26), but this relation was direct and not mediated by conspiracy belief in July. This was in contrast with conservative media use in March, which indirectly predicted lower mask wearing in July with an overall relation of –.093, 99% CI –.132 to –.061. Importantly, about half of this mediated relation (–.047/–.093=51%) was attributable to change in conspiracy beliefs (–.047, 99% CI –.076 to –.024). The rest of the relation went through the carryover of conspiracy belief from March to July (–.039, 99% CI –.061 to –.021) and the carryover of preventive action in March to mask wearing in July (–.007, 99% CI –.015 to –.003). Thus, these indirect paths show that use of conservative media in March predicted less mask wearing as mediated by increases in conspiracy belief in July apart from any change attributable to previous actions taken.

Figure 2. Standardized relations between media use in March and subsequent conspiracy beliefs and action taken to prevent spread of COVID infection in July. All paths within 99% CIs, except between social media and conspiracy beliefs in March, which excluded zero with a 95% CI.



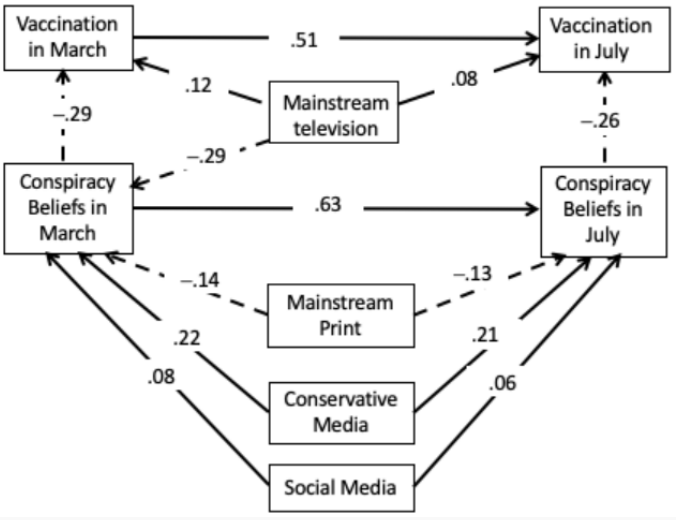
With regard to use of mainstream print in March, the overall relation with mask wearing in July was positive (.057, 99% CI .030-.086), indicating that it was associated with greater use of this behavior. In addition, in contrast with conservative media use, its relationship with mask wearing as mediated by change in belief in pandemic conspiracy theories was positive (.027, 99% CI .008-.042). This amounted to about 47% (.027/.057) of the relation between use of mainstream print in March and mask wearing in July. The rest of the relation was mediated by carryover in conspiracy belief from March to July (.026, 99% CI .008-.043) and through action taken in March (.005, 99% CI .001-.010).

Social media use in March also predicted less mask wearing in July with a total indirect relation of –.039, 99% CI –.067 to –.018. Here more than half of this relation (–.022/–.039=56%) was attributable to change in conspiracy beliefs (–.022, 99% CI –.040 to –.006).

Relations With Vaccination in July

Figure 3 shows the model that explored the role of various media uses in March as predictors of change in vaccination intentions and associated change in conspiracy beliefs. This model also fit the data well, with an RMSEA of .058, 90% CI .046-.072; CFI=.97; TLI=.94, and SRMR=.021.

Figure 3. Standardized relations between media use in March and conspiracy beliefs and vaccine intentions in July. All paths within 99% CIs, except between social media and conspiracy beliefs in March and July, which excluded zero at the 95% CI.



As with mask wearing, vaccination intentions in July were inversely related to conspiracy beliefs in July apart from demographic variables and political ideology (–.26, 99% CI –.34 to –.17). Again, use of mainstream television was a positive predictor in July even though its relation was independent of belief in pandemic conspiracies (.075, 95% CI .016-.135). In the case of vaccination, mainstream television was also associated with greater intention to vaccinate in March (.124, 99% CI .031-.211).

Use of conservative media in March predicted reduced intentions to vaccinate in July as mediated in total by conspiracy beliefs and carryover in vaccination intentions (–.120, 99% CI –.165 to –.083). Use of conservative media in March was specifically

related to change in vaccination intentions stemming from changes in conspiracy beliefs in July (–.054, 99% CI –.083 to –.031). This indicates that about 45% (–.054/–.120) of the overall relation was attributable to conservative media's association with increased belief in pandemic conspiracies from March to July. The rest of the relation was attributable to carryover in conspiracy beliefs from March to July (–.035, 99% CI –.059 to –.019) and carryover from vaccination intention in March to July (–.031, 99% CI –.050 to –.018).

As with mask wearing, use of mainstream print had an overall positive relation with vaccination intention in July (.078, 99% CI .037-116). About 44% (.034/.078) of this relation was due to mainstream print's association with change in conspiracy belief (.034, 99% CI .017-.055). The rest of the mediation was attributable to carryover in conspiracy beliefs from March to July (.023, 99% CI .007-.040) and carryover in vaccination intentions from March to July (.021, 99% CI .007-.036).

Use of social media in March was also predictive of vaccination in July, with an overall negative indirect relation of –.041, 99% CI –.071 to –.014. However, this relation as mediated by change in conspiracy beliefs was weaker than for the other media, with only about 39% (–.016/–.041) attributable to this source (–.016, 99% CI –.035 to –.001). The rest of the relation was attributable to carryover in conspiracy beliefs and vaccination intentions from March to July.

Discussion

Principal Findings

Despite the characterization of conspiracy beliefs as paranoid [5] and stigmatized by some in public discourse [6], they remain robust sources of skepticism regarding important public health recommendations able to prevent the spread of COVID-19. In our analysis of the prevalence of three conspiracy beliefs regarding the COVID-19 pandemic, we found that acceptance of those beliefs ranged from 17% for belief that pharmaceutical companies created the virus to 38% for belief that the Chinese government did so. These levels of acceptance grew over the period from March to July. As with many conspiracies, these beliefs can be accepted by the same person despite their logical incompatibility [55,56]. Scholars have argued that it is the underlying distrust of governments rather than the consistency of their content that appears to motivate their acceptance [57,58].

We found that reliance on different types of media for information during the early months of the COVID-19 pandemic in the United States predicted changes in the strength of belief in the three prevalent conspiracy theories. In particular, use of conservative media, such as Fox News and the talk radio program hosted by the late Rush Limbaugh, was associated with increased acceptance of the three conspiracy beliefs, while use of mainstream print—such as the New York Times and the Wall Street Journal—was associated with increased rejection of them. Although mainstream broadcast television news was negatively associated with conspiracy beliefs in March, there was no direct relation in July. Despite the efforts of the platforms to interdict such content, there was also a small but reliable increase in the perceived truth of the three theories among users of social media in July.

We found partial support for the hypothesis (H1) that use of mainstream news outlets would reduce belief in conspiracy theories because those media tend to follow traditional journalistic norms of reporting based on substantiated information backed by credible sources [27,28]. Conspiracy theories are by their nature spread by questionable sources whose assertions are difficult to verify [29]. As a result, those theories are less likely to be covered in the news, except to undercut them (eg, NBC news coverage of the QAnon conspiracy [59]). This hypothesis was supported in regard to mainstream print, but not broadcast television. We also found support for the hypothesis (H2) that use of conservative media would sustain and strengthen conspiracy beliefs because the commentators on these venues discussed those theories as plausible causes of the pandemic and explanations for failures of the Trump administration to cope with the crisis [60]. As one would expect, we found that, regardless of media use patterns, politically conservative more so than liberal or independent respondents endorsed the conspiracy beliefs at both times. Early in the pandemic, a Republican member of Congress wore a gas mask to a vote on emergency relief for the pandemic, in an apparent attempt to poke fun at the need for such action [61]. Moreover, President Trump opined that the virus was no more serious than the seasonal flu [62], suggested that it was created at the Wuhan Institute of Virology [63], characterized a finding by the Department of Veteran Affairs that hydroxychloroquine was not effective as a "Trump enemy statement" [64], and alleged that the "deep state" was delaying progress on a vaccine to thwart his reelection [65]. These moves to downplay the seriousness of the pandemic, blame its origins on a Chinese lab, and dismiss inconvenient science as the work of his enemies are consistent with two of the conspiracy beliefs on which we focused. Indeed, both the belief that some in the CDC were exaggerating the seriousness of the pandemic in order to undermine his presidency and that the Chinese created the virus as a weapon were endorsed by 63% and 55% of conservative respondents at the second wave, respectively. Conservatives were previously found to be more disposed to accept the COVID-19 conspiracy belief that "powerful people intentionally planned the COVID-19 outbreak" [66].

We did not make a directional prediction about the influence of social media on continued belief in conspiracy theories because, as we noted earlier, during the period between our two surveys, considerable effort was expended by major platforms to remove or downgrade misleading information about the pandemic [40,41]. In addition, a recent survey of content on Twitter concluded that despite the large amount of misinformation on social media, there is also a great amount of science-based information that circulates on those sites [67]. However, our findings suggest that these efforts did not remove the influence of the three theories on which we focused. This finding is consistent with evidence from a study conducted in April 2020 that found considerable evidence of tweeted comments supporting a conspiracy theory about the pandemic [68] and other work suggesting that conspiracy claims that are interdicted quickly reappear elsewhere [17,45]. Although

experience in China suggests that rapid refutation of pandemic rumors can be successful in reducing transmission of misinformation on social media [69], the expanse, variety, and decentralized nature of US media make such efforts more challenging than in countries whose media are largely government controlled. We also found that acceptance of conspiracy theories was associated with less action to prevent the spread of the disease, such as mask wearing. Similar patterns were observed in willingness to accept a vaccine. In both cases, use of conservative and social media were related to reductions in preventive behavior associated with changes in conspiracy beliefs.

Although it is difficult to debunk conspiracy theories, we did find that use of mainstream print predicted a subsequent decrease in the three beliefs. In addition, persons with greater education were more likely to move away from belief in those theories by July. Nevertheless, despite the potential ability of mainstream print to reduce belief in pandemic conspiracies, its influence appeared to be weaker than that of conservative and social media, which together predicted more than twice the effect size of mainstream print (.23 versus –.09). If this pattern of influence were to persist, it could lead to further increases in conspiracy belief among users of social and conservative media and, because of the smaller influence of print, a net increase in those beliefs in the US population at large.

On a positive note, the influence of mainstream television news, which also reaches a more sizeable audience than social and conservative media, appeared to outweigh their potentially negative effects on mask wearing and future vaccination. Indeed, mainstream television was the most used source of information in our sample. Despite the greater *change* in conspiracy beliefs associated with conservative and social media use, when the direct relations with vaccination intentions (.08) and mask wearing (.17) are included, the difference between social and conservative versus mainstream media use was negligible for vaccination (–.161 versus .158, respectively) and weaker for mask wearing (–.132 versus .227, respectively). Considering the larger reach of mainstream news, the overall potential effects of those news sources could well have outweighed the effects of conservative and social media use on the public's acceptance of vaccination and mask wearing.

Some have suggested that use of mainstream news is so dominant in the United States that disinformation transmitted through the internet and social media is unlikely to exert much influence [70]. Our findings are consistent with the conclusion that although mainstream news use is extensive and correlated with positive protective behaviors, the influence of social and conservative media is nonetheless significant. The three conspiracy beliefs that we studied are associated with media use outside of mainstream news, suggesting that these sources have a worrisome influence on the US public, despite their smaller share of the media market.

It is notable that use of mainstream television news was associated with greater mask wearing and intentions to vaccinate whether one accepted the three conspiracies or not. Use of broadcast television was also associated with taking recommended action in the United Kingdom [26] and Canada [25]. This pattern suggests that this popular source of information enhances compliance with recommended behavior by means other than dispelling conspiracy theories. Through its wide reach and its audiovisual capacities, it may provide direct exposure to persons wearing masks and taking vaccines, actions that would increase the normative acceptance of these behaviors. It is also likely that mainstream news sources such as television transmit more public health recommendations than are routinely accessed on social media [25] or conservative cable and opinion sites [60]. Trust in mainstream news is also associated with greater rejection of conspiracy theories [71].

It is noteworthy as well that some individual characteristics were no longer directly related to increased belief in COVID-19 conspiracies in July despite being associated with them in March. Although both Black and Hispanic respondents reported more belief in those theories than white respondents in March, Black and Hispanic respondents did not increase their belief by the second wave in July. This suggests that they were not exposed to greater sources of COVID-19 conspiracy information in the media during the intervening period. Importantly, neither Black nor Hispanic respondents were more likely to consume information on conservative media (*r* values ranging from –.015 to .057) and were only somewhat more likely to use information on social media (*r* values ranging from .017 to .099). At the same time, Black respondents were more likely to use mainstream television than others at both time points (*r* values=.167 and .144), while Hispanic respondents exhibited no differential use of information on any of the two remaining media groupings over the two time points. Nevertheless, despite the absence of an increase in conspiracy belief among Black and Hispanic respondents, their greater overall endorsement of such beliefs is still related to their unwillingness to accept vaccines.

Limitations

Although our two-wave panel study provides more sensitive evidence of the role of media use in the persistence of conspiracy beliefs regarding the COVID-19 pandemic, there are limitations that need to be recognized. We use a national probability sample to generalize to the US population, but the survey was conducted online, which restricted the sample to those with online access and experience. As a result, older citizens were underrepresented. Additionally, because we rely on self-reports of behavior, such as mask wearing, we cannot confirm that those reports reflect actual behavior. Our results also are only applicable to changes that occurred between March and July of 2020 in the United States. The effects of media use on pandemic conspiracy beliefs beyond that period remain to be studied. Finally, despite the ability to observe changes in conspiracy beliefs associated with media use, we cannot make strong causal claims because it still remains possible that characteristics other than those for which we controlled drove the changes in those beliefs. Nevertheless, our controls for a wide range of demographic differences as well as for political ideology increase our confidence that media use predicts changes in conspiracy beliefs regarding the pandemic.

Conclusions

Exposure to conservative and social media during the period from March to July 2020 predicted greater belief in three COVID-19 conspiracy beliefs in the United States, and these beliefs were related to less intention to vaccinate in the future and lower reported use of masks in the present. Public health agencies tasked with communicating the need for effective action to prevent the spread of the virus should seek opportunities to present accurate information about the pandemic to users of those media. At the same time, reaching users of mainstream media is also important in that they were either less likely to subscribe to conspiracy beliefs (in the case of print) or more likely to adopt protective behavior (in the case of mainstream broadcast television news). Although social media platforms have attempted to remove misinformation and conspiracy theories related to the COVID-19 pandemic, users of these platforms were more likely to exhibit increases in the strength of such beliefs in July. This finding suggests that those venues need to exert even greater efforts to counteract exposure to problematic COVID-19–related content.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Questionnaire items for national survey.
[DOCX File , 22 KB - jmir_v23i4e25215_app1.docx]

References

1. Hall Jamieson K, Albarracín D. The Relation between Media Consumption and Misinformation at the Outset of the SARS-CoV-2 Pandemic in the US. HKS Misinfo Review 2020 Apr 17;1(2):1-22. [doi: 10.37016/mr-2020-012]
2. Miller JM. Do COVID-19 Conspiracy Theory Beliefs Form a Monological Belief System? Can J Pol Sci 2020 May 21;53(2):319-326. [doi: 10.1017/S0008423920000517]
3. Stephens M. A geospatial infodemic: Mapping Twitter conspiracy theories of COVID-19. Dialogues in Human Geography 2020 Jun 23;10(2):276-281. [doi: 10.1177/2043820620935683]
4. Uscinski JE, Enders AM, Klofstad C, Seelig M, Funchion J, Everett C, et al. Why do people believe COVID-19 conspiracy theories? HKS Misinfo Review 2020 Apr 28:1-12. [doi: 10.37016/mr-2020-015]
5. Hofstadter R. The Paranoid Style in American Politics. New York: Vintage Books; 2008.
6. Butter M. The Nature of Conspiracy Theories. Medford, MA: Polity Press; 2020.
7. Goreis A, Voracek M. A Systematic Review and Meta-Analysis of Psychological Research on Conspiracy Beliefs: Field Characteristics, Measurement Instruments, and Associations With Personality Traits. Front Psychol 2019 Feb 11;10:1-13 [FREE Full text] [doi: 10.3389/fpsyg.2019.00205] [Medline: 30853921]
8. Swami V, Voracek M, Stieger S, Tran US, Furnham A. Analytic thinking reduces belief in conspiracy theories. Cognition 2014 Dec;133(3):572-585. [doi: 10.1016/j.cognition.2014.08.006] [Medline: 25217762]
9. Lantian A, Muller D, Nurra C, Douglas KM. 'I Know Things They Don't Know!'. Social Psychology 2017 Jul;48(3):160-173. [doi: 10.1027/1864-9335/a000306]
10. Bogart LM, Thorburn S. Are HIV/AIDS conspiracy beliefs a barrier to HIV prevention among African Americans? J Acquir Immune Defic Syndr 2005 Feb 01;38(2):213-218. [doi: 10.1097/00126334-200502010-00014] [Medline: 15671808]
11. Oliver JE, Wood T. Medical conspiracy theories and health behaviors in the United States. JAMA Intern Med 2014 May;174(5):817-818. [doi: 10.1001/jamainternmed.2014.190] [Medline: 24638266]
12. Douglas KM, Uscinski JE, Sutton RM, Cichocka A, Nefes T, Ang CS, et al. Understanding Conspiracy Theories. Political Psychology 2019 Mar 20;40(S1):3-35. [doi: 10.1111/pops.12568]
13. Bierwiaczonek K, Kunst JR, Pich O. Belief in COVID-19 Conspiracy Theories Reduces Social Distancing over Time. Appl Psychol Health Well Being 2020 Dec 30;12(4):1270-1285. [doi: 10.1111/aphw.12223] [Medline: 32864837]
14. van Mulukom V, Pummerer L, Alper S. Antecedents and consequences of COVID-19 conspiracy theories: A rapid review of the evidence. PsyArXiv. Preprint posted online on November 13, 2020 [FREE Full text] [doi: 10.31234/osf.io/u8yah]
15. Imhoff R, Bruder M. Speaking (Un–)Truth to Power: Conspiracy Mentality as A Generalised Political Attitude. Eur J Pers 2020 Dec 02;28(1):25-43. [doi: 10.1002/per.1930]
16. Georgiou N, Delfabbro P, Balzan R. COVID-19-related conspiracy beliefs and their relationship with perceived stress and pre-existing conspiracy beliefs. Pers Individ Dif 2020 Nov 01;166:1-7 [FREE Full text] [doi: 10.1016/j.paid.2020.110201] [Medline: 32565592]
17. Kearney MD, Chiang SC, Massey PM. The Twitter origins and evolution of the COVID-19 'plandemic' conspiracy theory. HKS Misinfo Review 2020 Oct 9;1:1-18. [doi: 10.37016/mr-2020-42]
18. Romer D, Jamieson KH. Conspiracy theories as barriers to controlling the spread of COVID-19 in the US. Soc Sci Med 2020 Oct;263:1-8 [FREE Full text] [doi: 10.1016/j.socscimed.2020.113356] [Medline: 32967786]

19. Chan MS, Jones CR, Hall Jamieson K, Albarracín D. Debunking: A Meta-Analysis of the Psychological Efficacy of Messages Countering Misinformation. Psychol Sci 2017 Nov 12;28(11):1531-1546 [FREE Full text] [doi: 10.1177/0956797617714579] [Medline: 28895452]
20. Jolley D, Douglas KM. Prevention is better than cure: Addressing anti-vaccine conspiracy theories. J Appl Soc Psychol 2017 Jun 28;47(8):459-469. [doi: 10.1111/jasp.12453]
21. Uscinski JE, Klofstad C, Atkinson MD. What Drives Conspiratorial Beliefs? The Role of Informational Cues and Predispositions. Political Research Quarterly 2016 Jan 13;69(1):57-71. [doi: 10.1177/1065912915621621]
22. Imhoff R, Lamberty P. A Bioweapon or a Hoax? The Link Between Distinct Conspiracy Beliefs About the Coronavirus Disease (COVID-19) Outbreak and Pandemic Behavior. Social Psychological and Personality Science 2020 Jul 06;11(8):1110-1118. [doi: 10.1177/1948550620934692]
23. Islam M, Sarkar T, Khan S, Mostofa Kamal AH, Hasan S, Kabir A, et al. COVID-19-Related Infodemic and Its Impact on Public Health: A Global Social Media Analysis. Am J Trop Med Hyg 2020 Oct;103(4):1621-1629 [FREE Full text] [doi: 10.4269/ajtmh.20-0812] [Medline: 32783794]
24. Gollust SE, Vogel RI, Rothman A, Yzer M, Fowler EF, Nagler RH. Americans' perceptions of disparities in COVID-19 mortality: Results from a nationally-representative survey. Prev Med 2020 Dec;141:1-9 [FREE Full text] [doi: 10.1016/j.ypmed.2020.106278] [Medline: 33027615]
25. Bridgman A, Merkley E, Loewen PJ, Owen T, Ruths D, Teichmann L, et al. The causes and consequences of COVID-19 misperceptions: Understanding the role of news and social media. HKS Misinfo Review 2020 Jun 18;1:1-18. [doi: 10.37016/mr-2020-028]
26. Allington D, Duffy B, Wessely S, Dhavan N, Rubin J. Health-protective behaviour, social media usage and conspiracy belief during the COVID-19 public health emergency. Psychol Med 2020 Jun 09:1-7 [FREE Full text] [doi: 10.1017/S003329172000224X] [Medline: 32513320]
27. Benkler Y, Roberts H, Faris R. Network Propaganda: Manipulation, Disinformation, and Radicalization in American Politics. New York: Oxford University Press; 2018.
28. Ericson RV. How journalists visualize fact. Ann Am Acad Pol Soc Sci 1998;560(1):83-95. [doi: 10.1177/0002716298560001007]
29. Lewandowsky S, Cook J. The Conspiracy Theory Handbook. Bristol, UK: University of Bristol, UK; 2020 Mar. URL: https://www.climatechangecommunication.org/wp-content/uploads/2020/03/ConspiracyTheoryHandbook.pdf [accessed 2021-04-22]
30. Fox News. Tucker Carlson Tonight. 2020 Feb 10. URL: https://www.foxnews.com/transcript/adam-carolla-california-doesnt-care-about-the-homeless-because-they-dont-have-money [accessed 2020-10-15]
31. Bowden J. Limbaugh claims coronavirus is being 'weaponized' to 'bring down' Trump. The Hill. 2020 Feb 25. URL: https://thehill.com/homenews/media/484506-limbaugh-claims-coronavirus-is-being-weaponized-to-bring-down-trump [accessed 2021-04-15]
32. 106 sites in the US and Europe now promoting coronavirus misinformation. NewsGuard. 2020 Mar 03. URL: https://www.newsguardtech.com/press/106-sites-in-the-u-s-and-europe-now-promoting-coronavirus-misinformation/ [accessed 2021-04-22]
33. Thompson C. YouTube's plot to silence conspiracy theories. Wired Mag. 2020 Sep 18. URL: https://www.wired.com/story/youtube-algorithm-silence-conspiracy-theories/ [accessed 2020-10-15]
34. Gadde V, Derella M. An update on our continuing strategy during COVID-19. Twitter. 2020 Mar 16. URL: https://blog.twitter.com/en_us/topics/company/2020/An-update-on-our-continuity-strategy-during-COVID-19.html [accessed 2020-10-15]
35. O'Sullivan D. Twitter says it will label misleading coronavirus tweets--even if they're from Trump. CNN. 2020 May 11. URL: https://www.cnn.com/2020/05/11/tech/twitter-coronavirus-misinformation/index.html [accessed 2021-04-22]
36. Brodwin E. How Pinterest beat back vaccine misinformation--and what Facebook could learn from its approach. STAT. 2020 Sep 21. URL: https://www.statnews.com/2020/09/21/pinterest-facebook-vaccine-misinformation/ [accessed 2020-10-15]
37. Shieber J. Zuckerberg details the ways Facebook and Chan Zuckerberg initiative are responding to COVID-19. TechCrunch. 2020 Mar 03. URL: https://techcrunch.com/2020/03/03/zuckerberg-details-the-ways-facebook-and-chan-zuckerberg-initiative-are-responding-to-covid-19/ [accessed 2020-10-14]
38. Rosen G. An update on our work to keep people informed and limit misinformation about COVID-19. Facebook. 2020 Apr 16. URL: https://about.fb.com/news/2020/04/covid-19-misinfo-update/ [accessed 2020-10-15]
39. Lerman R. Facebook says it has taken down 7 million posts for spreading coronavirus misinformation. Washington Post. 2020 Aug 11. URL: https://www.washingtonpost.com/technology/2020/08/11/facebook-covid-misinformation-takedowns/ [accessed 2020-10-12]
40. DeSalvo K, Canegallo K. How you'll find accurate and timely information on COVID-19 vaccines. Google News. URL: https://blog.google/technology/health/accurate-timely-information-covid-19-vaccines/ [accessed 2020-12-22]
41. Ahmed W, Vidal-Alaball J, Downing J, López Seguí F. COVID-19 and the 5G Conspiracy Theory: Social Network Analysis of Twitter Data. J Med Internet Res 2020 May 06;22(5):e19458 [FREE Full text] [doi: 10.2196/19458] [Medline: 32352383]
42. Marchal N, Au H. 'Coronavirus EXPLAINED': YouTube, COVID-19, and the Socio-Technical Mediation of Expertise. Social Media + Society 2020 Aug 11;6(3):205630512094815. [doi: 10.1177/2056305120948158]

43. Herrera S. Coronavirus misinformation lives online, despite efforts to stamp it out. Wall Street Journal. 2020 Mar 03. URL: https://www.wsj.com/articles/coronavirus-misinformation-lives-online-despite-efforts-to-stamp-it-out-11583272556 [accessed 2020-10-14]
44. Fichera A, Spencer S, Gore D, Robertson L, Kiely E. The falsehoods of the 'Plandemic' video. FactCheck.org. Philadelphia, PA: Annenberg Public Policy Center, University of Pennsylvania; 2020 May 08. URL: https://www.factcheck.org/2020/05/the-falsehoods-of-the-plandemic-video/ [accessed 2020-10-15]
45. Frankel S, Decker B, Alba D. How the 'Plandemic' movie and its falsehoods spread widely online. New York Times. URL: https://www.nytimes.com/2020/05/20/technology/plandemic-movie-youtube-facebook-coronavirus.html [accessed 2020-10-14]
46. National Opinion Research Center. About AmeriSpeak from NORC Internet. NORC. URL: https://www.amerispeak.org/about-amerispeak [accessed 2020-05-30]
47. Lynas M. COVID: Top 10 current conspiracy theories. Ithaca, NY: Cornell Alliance for Science; 2020 Apr 20. URL: https://allianceforscience.cornell.edu/blog/2020/04/covid-top-10-current-conspiracy-theories [accessed 2020-10-13]
48. Enders AM, Uscinski JE, Klofstad C, Stoler J. The different forms of COVID-19 misinformation and their consequences. HKS Misinfo Review 2020 Nov 12;1(8):1-21. [doi: 10.37016/mr-2020-48]
49. Howard J, Huang A, Li Z, Tufekci Z, Zdimal V, van der Westhuizen HM, et al. An evidence review of face masks against COVID-19. Proc Natl Acad Sci U S A 2021 Jan 26;118(4):e2014564118. [doi: 10.1073/pnas.2014564118] [Medline: 33431650]
50. Edgerly S. Red Media, Blue Media, and Purple Media: News Repertoires in the Colorful Media Landscape. Journal of Broadcasting & Electronic Media 2015 Mar 11;59(1):1-21. [doi: 10.1080/08838151.2014.998220]
51. Pew Research Center. Republicans, Democrats move even further apart in Coronavirus concerns. Washington, DC: Pew Research Center; 2020 Jun 25. URL: https://www.pewresearch.org/politics/2020/06/25/republicans-democrats-move-even-further-apart-in-coronavirus-concerns/ [accessed 2020-08-16]
52. Weeks BE, Ksiazek TB, Holbert RL. Partisan Enclaves or Shared Media Experiences? A Network Approach to Understanding Citizens' Political News Environments. Journal of Broadcasting & Electronic Media 2016 May 16;60(2):248-268. [doi: 10.1080/08838151.2016.1164170]
53. Muthén LK, Muthén BO. Mplus user's guide. Los Angeles, CA: Muthén & Muthén; 2015. URL: https://www.statmodel.com/download/usersguide/MplusUserGuideVer_7.pdf [accessed 2021-04-22]
54. Bollen KA, Davis WR. Causal Indicator Models: Identification, Estimation, and Testing. Structural Equation Modeling: A Multidisciplinary Journal 2009 Jul 14;16(3):498-522. [doi: 10.1080/10705510903008253]
55. Irwin HJ, Dagnall N, Drinkwater K. Belief Inconsistency in Conspiracy Theorists. Comprehensive Psychology 2015 Jan;4:17.CP.4.19. [doi: 10.2466/17.cp.4.19]
56. Wood MJ, Douglas KM, Sutton RM. Dead or alive: Beliefs in contradictory conspiracy theories. Soc Psychol Personal Sci 2012 Jan 25;3(6):767-773. [doi: 10.1177/1948550611434786]
57. Crocker J, Luhtanen R, Broadnax S, Blaine BE. Belief in U.S. Government Conspiracies Against Blacks among Black and White College Students: Powerlessness or System Blame? Pers Soc Psychol Bull 1999;25(8):941-953. [doi: 10.1177/01461672992511003]
58. Uscinski JE, Parent JM. American Conspiracy Theories. New York: Oxford University Press; 2014.
59. What is QAnon?: Far-right conspiracy movement's growing influence. NBC News. 2020 Oct 01. URL: https://www.nbcnews.com/nightly-news/video/what-is-qanon-far-right-conspiracy-movement-s-growing-influence-92966469972 [accessed 2020-10-15]
60. Motta M, Stecula D, Farhart C. How Right-Leaning Media Coverage of COVID-19 Facilitated the Spread of Misinformation in the Early Stages of the Pandemic in the US. Can J Pol Sci 2020 May 01;53(2):335-342 [FREE Full text] [doi: 10.1017/S0008423920000396]
61. Cummings W. 'I was quite serious': Rep. Matt Gaetz denies he wore gas mask to make light of coronavirus after constituent dies. Washington, DC: USA Today; 2020 Mar 9. URL: https://www.usatoday.com/story/news/politics/2020/03/09/coronavirus-matt-gaetz-defends-use-gas-mask-after-constituent-dies/5000255002 [accessed 2020-05-27]
62. Montanaro D. Trump compares coronavirus to the flu, but it could be 10 times deadlier. National Public Radio. 2020 Mar 24. URL: https://www.npr.org/sections/coronavirus-live-updates/2020/03/24/820797301/fact-check-trump-compares-coronavirus-to-the-flu-but-they-are-not-the-same [accessed 2020-05-30]
63. Marcus J. Coronavirus: Trump stands by China lab origin theory for virus. BBC News. 2020 May 01. URL: https://www.bbc.com/news/world-us-canada-52496098 [accessed 2021-02-16]
64. Associated Press. AP fact check: Trump trashing virus science he doesn't like. WTTW Radio. 2020 May 23. URL: https://news.wttw.com/2020/05/23/ap-fact-check-trump-trashing-virus-science-he-doesn-t [accessed 2021-02-16]
65. Herper M, Florko N. Drug makers rebut Trump tweet that FDA 'deep state' is delaying Covid-19 vaccines and drugs. STAT. 2020 Aug 22. URL: https://apnews.com/article/d976ab2b204ac14b220d5bf6dc565b2d [accessed 2021-04-16]
66. Schaeffer K. A look at Americans who believe there is some truth to the conspiracy theory that COVID-19 was planned. Pew Research Center. 2020 Jul 24. URL: https://www.pewresearch.org/fact-tank/2020/07/24/a-look-at-the-americans-who-believe-there-is-some-truth-to-the-conspiracy-theory-that-covid-19-was-planned/

67. Pulido CM, Villarejo-Carballido B, Redondo-Sama G, Gómez A. COVID-19 infodemic: More retweets for science-based information on coronavirus than for false information. International Sociology 2020 Apr 15;35(4):377-392. [doi: 10.1177/0268580920914755]
68. Ahmed W, López Seguí F, Vidal-Alaball J, Katz MS. COVID-19 and the "Film Your Hospital" Conspiracy Theory: Social Network Analysis of Twitter Data. J Med Internet Res 2020 Oct 05;22(10):1-8 [FREE Full text] [doi: 10.2196/22374] [Medline: 32936771]
69. Chen B, Chen X, Pan J, Liu K, Xie B, Wang W, et al. Dissemination and Refutation of Rumors During the COVID-19 Outbreak in China: Infodemiology Study. J Med Internet Res 2021 Feb 15;23(2):1-16 [FREE Full text] [doi: 10.2196/22427] [Medline: 33493124]
70. Allen J, Howland B, Mobius M, Rothschild D, Watts DJ. Evaluating the fake news problem at the scale of the information ecosystem. Sci Adv 2020 Apr 03;6(14):1-6 [FREE Full text] [doi: 10.1126/sciadv.aay3539] [Medline: 32284969]
71. van Mulukom V. The role of trust and information in protective behaviors and conspiracy belief during the COVID-19 pandemic. PsyArXiv. Preprint posted online on August 1, 2020 [FREE Full text] [doi: 10.31234/osf.io/chy4b]

Abbreviations

CDC: Centers for Disease Control and Prevention
CFI: comparative fit index
RMSEA: root mean square error of approximation
SRMR: standardized root mean square residual
TLI: Tucker-Lewis Index

Edited by C Basch; submitted 22.10.20; peer-reviewed by D Allington, A Chang, A van Stekelenburg; comments to author 02.12.20; revised version received 23.12.20; accepted 08.04.21; published 27.04.21.

Please cite as:
Romer D, Jamieson KH
Patterns of Media Use, Strength of Belief in COVID-19 Conspiracy Theories, and the Prevention of COVID-19 From March to July 2020 in the United States: Survey Study
J Med Internet Res 2021;23(4):e25215
URL: https://www.jmir.org/2021/4/e25215
doi:10.2196/25215
PMID:33857008

Original Paper

Predictability of COVID-19 Hospitalizations, Intensive Care Unit Admissions, and Respiratory Assistance in Portugal: Longitudinal Cohort Study

André Patrício[1], BSc; Rafael S Costa[2,3*], PhD; Rui Henriques[1,4*], PhD

[1]Instituto Superior Técnico, Universidade de Lisboa, Lisboa, Portugal

[2]LAQV-REQUIMTE, NOVA School of Science and Technology, Universidade NOVA de Lisboa, Caparica, Portugal

[3]IDMEC, Instituto Superior Técnico, Universidade de Lisboa, Lisboa, Portugal

[4]Instituto de Engenharia de Sistemas e Computadores–Investigação e Desenvolvimento, Lisboa, Portugal

[*]these authors contributed equally

Corresponding Author:
Rafael S Costa, PhD
LAQV-REQUIMTE
NOVA School of Science and Technology
Universidade NOVA de Lisboa
Campus Caparica, 2829-516
Caparica, 2829-516
Portugal
Phone: 351 21 294 8351
Email: rs.costa@fct.unl.pt

Abstract

Background: In the face of the current COVID-19 pandemic, the timely prediction of upcoming medical needs for infected individuals enables better and quicker care provision when necessary and management decisions within health care systems.

Objective: This work aims to predict the medical needs (hospitalizations, intensive care unit admissions, and respiratory assistance) and survivability of individuals testing positive for SARS-CoV-2 infection in Portugal.

Methods: A retrospective cohort of 38,545 infected individuals during 2020 was used. Predictions of medical needs were performed using state-of-the-art machine learning approaches at various stages of a patient's cycle, namely, at testing (prehospitalization), at posthospitalization, and during postintensive care. A thorough optimization of state-of-the-art predictors was undertaken to assess the ability to anticipate medical needs and infection outcomes using demographic and comorbidity variables, as well as dates associated with symptom onset, testing, and hospitalization.

Results: For the target cohort, 75% of hospitalization needs could be identified at the time of testing for SARS-CoV-2 infection. Over 60% of respiratory needs could be identified at the time of hospitalization. Both predictions had >50% precision.

Conclusions: The conducted study pinpoints the relevance of the proposed predictive models as good candidates to support medical decisions in the Portuguese population, including both monitoring and in-hospital care decisions. A clinical decision support system is further provided to this end.

(J Med Internet Res 2021;23(4):e26075) doi:10.2196/26075

KEYWORDS

COVID-19; machine learning; intensive care admissions; respiratory assistance; predictive models; data modeling; clinical informatics

Introduction

Background

COVID-19 is a disease caused by the novel coronavirus SARS-CoV-2, transmissible from person to person and associated with acute respiratory complications in severe cases [1,2]. The main symptoms of patients infected are fever, cough, and fatigue; others are asymptomatic [3]. The COVID-19 pandemic presents a substantial threat to global health and has been directly responsible for many deaths. Since the first

outbreak in December 2019 in Wuhan, China, the number of confirmed infected patients worldwide has exceeded 55 million cases, and nearly 1.3 million people have died from COVID-19 [4]. Current literature has shown that infected patients with specific comorbidities or preconditions (eg, hypertension, respiratory problems, diabetes) and of old age are expected to develop a more severe response to the infection and may consequently need longer hospitalizations and intensive care [5-7]. Strict social confinement measures have been implemented to decrease the COVID-19 R_0 value (average number of individuals infected by each infected person) and guarantee the optimal use of equipment and beds at normal, continuous, and intensive care units (ICUs). However, although public health responses aim to delay the spread of the infection, several countries such as the United States, Brazil, Italy, and India have faced severe health care crises.

Without effective antiviral drugs and a vaccine, prognostic tools related to COVID-19 are required. Statistical and computational models could assist clinical staff in triaging patients at high risk for respiratory failure to better guide the allocation of medical resources. Recently, several predictive models ranging from statistical and score-based systems to more recent machine learning models have been proposed in response to COVID-19. Guan et al [8] proposed a Cox regression model to infer potential risk factors associated with serious adverse outcomes in patients with COVID-19. Univariate and multivariate logistic regression models have been used to determine risk factors associated with mortality [9]. Scoring systems have been proposed to predict COVID-19 patient mortality but are limited by small sample sizes, with a poor discriminatory ability [10-12]. Other statistical approaches have also been emerging to aid prognostics [13,14]. Complementarily, machine learning methods offer the possibility to model more complex data relationships, generally yielding powerful capabilities to predict outcomes of infectious and noninfectious diseases in medical practice [15-17]. To this end, classification and regression models have been proposed for risk stratification of patients and to screen the spread of COVID-19 [18-20]. Despite the inherent potentialities of ongoing efforts, studies in the context of COVID-19 are limited by either the size of available cohorts or the lack of a systematic comparison of different models [21-24], and generally neglect the predictability of medical needs (instead the focus is commonly placed on measurable disease factors, early detection of infection, and mortality risk prediction [25-28]). None of these studies have comprehensively targeted the Portuguese population at the present time.

This Study

This study provides a structured view on the predictability of hospitalizations, ICU admissions, respiratory assistance needs, and survivability outcomes using a retrospective cohort encompassing individuals with a SARS-CoV-2–positive result in Portugal as of June 30, 2020.

To this end, and considering demographic, comorbidity, and care provision variables collected for the infected individuals, an assessment methodology was conducted, whereby state-of-the-art predictive models were hyperparameterized and robustly evaluated in order to assess the upper bounds on the predictive performance for each one of the targeted variables. In addition, whenever applicable, this analysis was extended toward the various stages of a patient's cycle: prehospitalization (at the time of testing), after hospitalization, and after ICU admission.

This study offers a solid methodology for the robust assessment of the predictability guarantees of future care needs of infected individuals, contrasting with the dominant correlation-based guarantees in literature. As comparable studies demonstrated in other populations, it lays a solid ground to compare type-I and type-II predictive errors and assess population-wise differences.

Methods

Overview

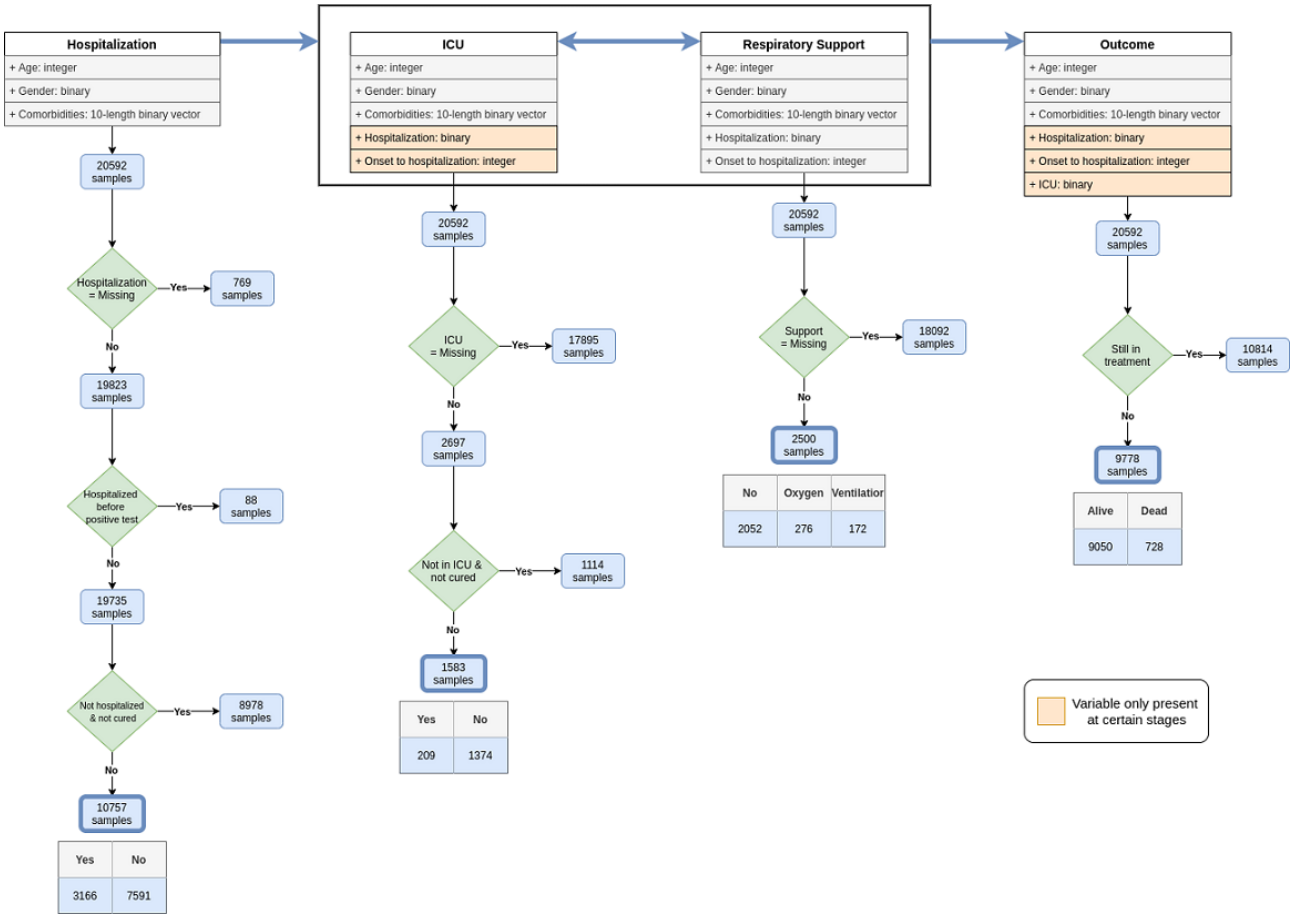
Complete subpopulations from the target cohort were identified for each output (Figure 1), guaranteeing the presence of all individuals undertaking the target forms of care (hospitalization, ICU admission, respiratory support) with a recovery-or-death outcome.

After the sampling and data curation steps (Figure 1), we proceeded to the optimization of data preprocessing options and classifiers' parameterization for each of the target variables separately. To this end, we applied a nested 10-fold cross-validation assessment methodology, whereby we first create train-test partitions (outer cross-validation) to assess the performance of an optimized classification method, and within each training fold we further create train-test partitions (inner cross-validation) for hyperparameterizing the predictive model under assessment. This methodology guarantees that all observations are used to assess the final performance and prevents biases as hyperparameterization takes place within each training folds.

Within each inner train-test fold, Bayesian optimization [29] was applied to find the hyperparameters that best fit the pipeline. The optimization measures are:

- F1 score and 0.7 × recall + 0.3 × precision for binary classes. These two views generate two sets of classifiers: one that equally weights recall-and-precision views, and other that, similar to the F2 score (F_β, where $\beta=2$), prioritizes the optimization of the true-positive rate (recall) at the cost of a lower positive predictive value (precision);
- Cohen kappa and average class recall for target variables with more than 2 classes (respiratory support).

Figure 1. Exclusion and inclusion criteria for composing the outcome-conditional cohorts: hospitalization, respiratory assistance support, intensive care unit (ICU) admission, and survivability. Blank/unknown cells in the Direcção-Geral da Saúde data set are classified as missing values.



Hospitalization, UCI admission, respiratory support, and recovery-or-death outcomes for SARS-CoV-2–infected individuals are considerably imbalanced, hence, the relevance of the placed recall-precision and multiclass recall views. In particular, considering both a balanced recall-precision optimization and recall-oriented optimization is relevant for clinical decisions. When the allocated teams have capacity to remotely monitor SARS-CoV-2–infected patients, the predictive models optimized with a schema that prioritizes recall should be pursued to guarantee that no vulnerable patient is left out. Nevertheless, when monitoring capacity is limited, greater attention to precision is necessary, and only more vulnerable patients (as suggested by the predictive models optimized with balanced recall-precision views) should be attemptively monitored.

The allowed preprocessing options are as follows: imputation of missing values using median-mode imputation, KNNImputer, or none; class balancing using subsampling, oversampling, SMOTE (Synthetic Minority Oversampling Technique), or none; and normalization of real-valued variables using standardization, scaling, or none. The selected classifiers are as follows: Bernoulli naive Bayes, Gaussian naive Bayes, k-nearest neighbors (KNN), decision tree (DT), random forest, XGBoost (XGB), logistic regression, Light Gradient Boosting Machine (LightGBM), Super Learner, and multilayer perceptron (MLP). Super Learner uses folding to hyperparameterize models and selects predictors for out-of-fold predictions from individual performance estimates per fold. In this context, Super Learner's performance is generally coincident with the best predictive model and thus not always disclosed in the *Results* section to allow the identification of the best underlying predictors. We considered the implementations provided in the *scikit-learn* [30] and *xgboost* [31] packages in Python (Python Software Foundation). For each classifier, all supported parameters in scikit-learn were subjected to hyperparameterization. Regarding the MLP, we placed upper limits on the number of hidden layers (3) and nodes per layer (20) given the low-dimensionality nature of the target data set. The hyperparameters were subjected to a total of 50 iterations. Multimedia Appendix 1 displays the optimized parameters for the best-performing predictive models per outcome.

Differences in performance from the paired-error estimates collected per fold were statistically tested using *t* tests when estimates passed the Shapiro–Wilk normality test. When this condition was not satisfied, Wilcoxon signed-rank tests were applied.

In addition to the conducted analysis, the best predictors trained on the whole data set were made available within a clinical decision support system built using flask technology and dash facilities in Python [32], which can run as an offline web application.

Data Source

A retrospective cohort (from March 1 to June 30, 2020) of patients with confirmed COVID-19 in Portugal was used in this

study. The anonymized data set was provided by the Directorate General of Health (Direcção-Geral da Saúde, DGS), the Portuguese health authority. The gathered data, called the *covid19-DGS* database, contains information pertaining to the demographic and clinical patient characteristics as well as preexisting conditions.

Data are available upon reasonable request.

Ethical Considerations

The COVID-19 data set is provided by the DGS under the collaborative *score4COVID* research project proposal. The tasks conducted in the *score4COVID* project were further validated by the Ethical Committee of the NOVA School of Science and Technology.

Results

Results on the predictability of hospitalization needs, ICU admissions, respiratory assistance, and outcome of infected individuals living in Portugal, as of June 30, 2020, are discussed below.

Cohort Characteristics

The target cohort comprised 38,545 individuals who were SARS-CoV-2 positive: 17,046 recoveries (SARS-CoV-2 negative after positive testing) and 1155 deaths. Four individuals were excluded from the data set due to inconsistent recordings related to age and pregnancy-gender variables. Table 1 provides essential statistics. Figure 2 further describes sex and age distributions in deaths, hospitalizations, ICU admissions, and average number of days from symptom onset (traced by the public health line for COVID-19) to a positive test result and hospitalization.

Within the target population, there were 4326 hospitalizations (11.2% of population base) and 253 admissions to the ICU (5.8% of hospitalizations). Among ICU internments, there were 82 recoveries and 61 deaths. In terms of respiratory support, a total of 180 individuals undertook assisted ventilation, 292 submitted to oxygen therapy, and 9 underwent alternative modes of respiratory support such as extracorporeal membrane oxygenation.

The major classes of comorbidities monitored were neoplasm, diabetes, asthma, pulmonary, hepatic, hematological, renal, neurological, neuromuscular, and immune deficiency conditions. The representativity of individuals with one or more comorbidities, as well as their impact on survivability, is depicted in Figure 3.

Table 1. Characteristics of SARS-CoV-2–infected patients in the target cohort.

Characteristic	Value
Numeric variables, mean (SD); range	
Age at notification (years)	48.3 (22.1); 0-105
Onset to hospitalization (days)	1.1 (5.1); 0-169
Categoric variables, n (%)	
Gender	
Female	11,252 (54.64)
Male	9340 (45.36)
Hospitalization	
Yes	16,651 (84.00)
No	3172 (16.00)
ICU[a] admission	
Yes	209 (7.75)
No	2488 (92.25)
Respiratory support	
Oxygen therapy	276 (11.04)
Assisted ventilation	172 (6.88)
No support	2052 (82.08)
Comorbidities	
Cancer	940 (4.56)
Cardiac disease	3025 (14.69)
Diabetes	2134 (10.36)
Immune deficiency	222 (1.08)
Renal disease	718 (3.49)
Liver disease	206 (1.00)
Lung disease	794 (3.86)
Chronic neurological disease	1087 (5.28)
Mortality	
Yes	728 (7.45)
No	9050 (92.55)

[a]ICU: intensive care unit.

Figure 2. Cohort statistics: (A-C) demographic distribution of infected individuals with known outcome (death and recovery) and stage in the care life cycle (hospitalization and intensive care unit [ICU] admission); (D) average number of days between care stages (the plotted negative bin [ie, negative occurrences] corresponds to hospitalizations before SARS-CoV-2 testing).

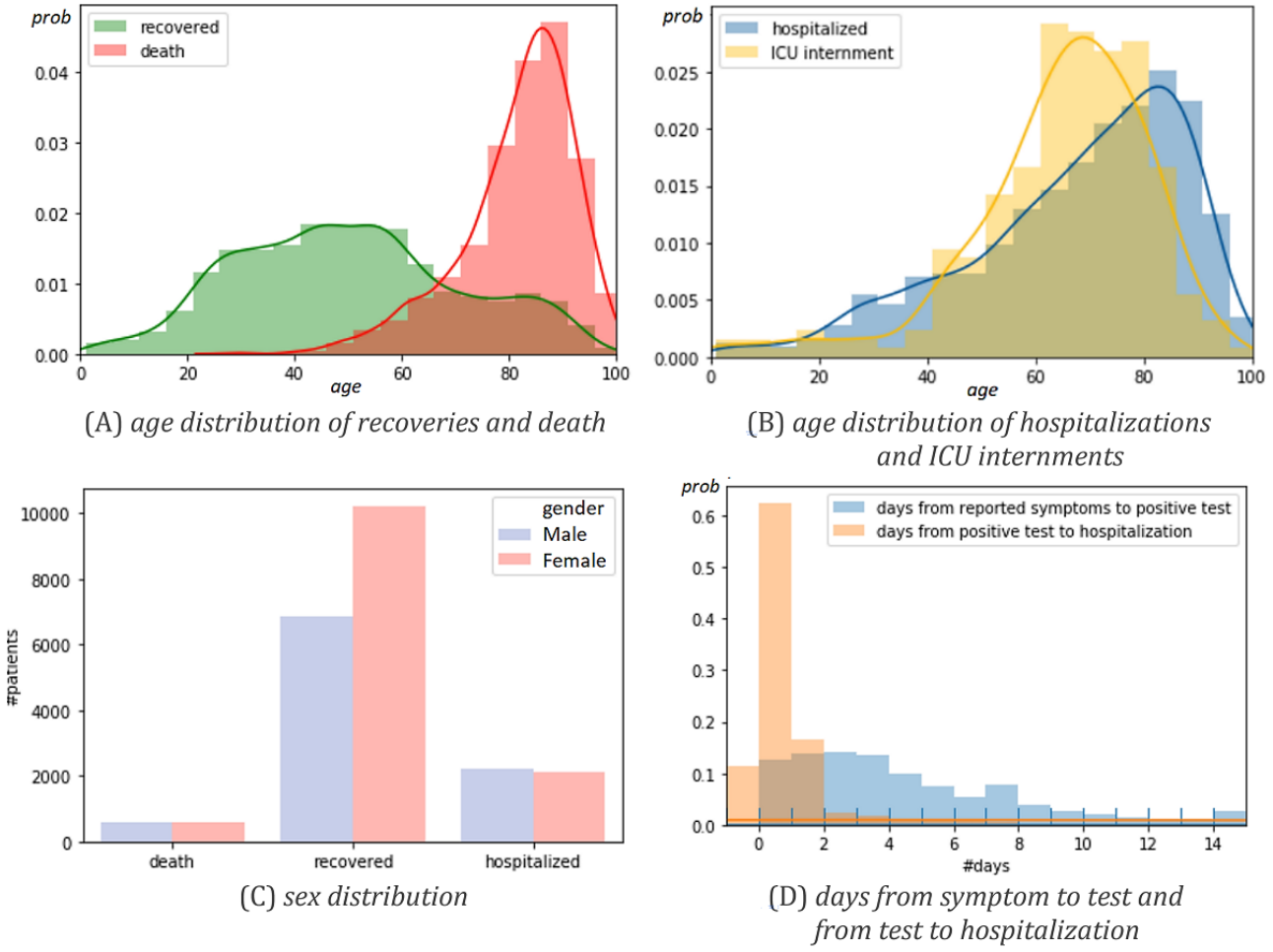
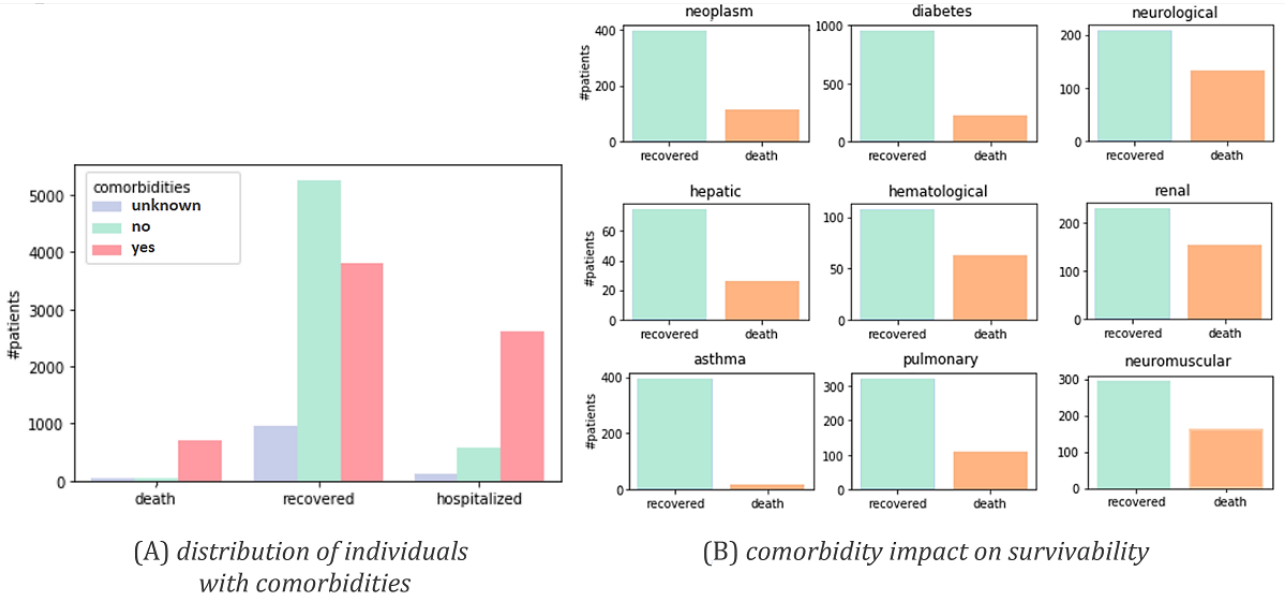


Figure 3. Cohort statistics: (A) distribution of individuals with one or more comorbidities among deaths, recovered cases, and hospitalizations; and (B) association between individual comorbidities and survivability outcomes.



Hospitalization

Figure 4 and Table 2 provide results pertaining to the models' ability to predict the need for individuals to be hospitalized once they are tested as SARS-CoV-2 positive given their (1) demographic group (age and gender) and (2) comorbidity factors. Comorbidity factors were categorized in accordance with the presence or absence of kidney, asthma, lung, cancer,

neuromuscular, diabetes, HIV, cardiac, and pregnancy conditions. Nonhospitalized individuals without a clear outcome (recovery or death) were excluded from this analysis. Figure 5 provides the receiver operating characteristic curve per predictor for each optimization setting.

Generally, we observed that nearly 90% of hospitalization needs could be identified at the time of SARS-CoV-2 testing. This level of recall/sensitivity was observed at the expense of an approximate 55% precision, meaning that more than half of the predicted hospitalization needs were in fact observed. Logistic regression and MLP were the best-performing classification models according to F1-score and recall, respectively. Statistical superiority was verified for logistic regression but not MLP against peer models (at =.05). These results provide empirical evidence toward the role of these predictors in supporting individual remote monitoring decisions.

Figure 4. Predictability of hospitalizations for individuals testing SARS-CoV-2 positive. Recall, precision, and F1 for the best predictors in F1 (left) and recall-oriented (right) scores on the validation set after nested cross-validation. MPL: multilayer perceptron.

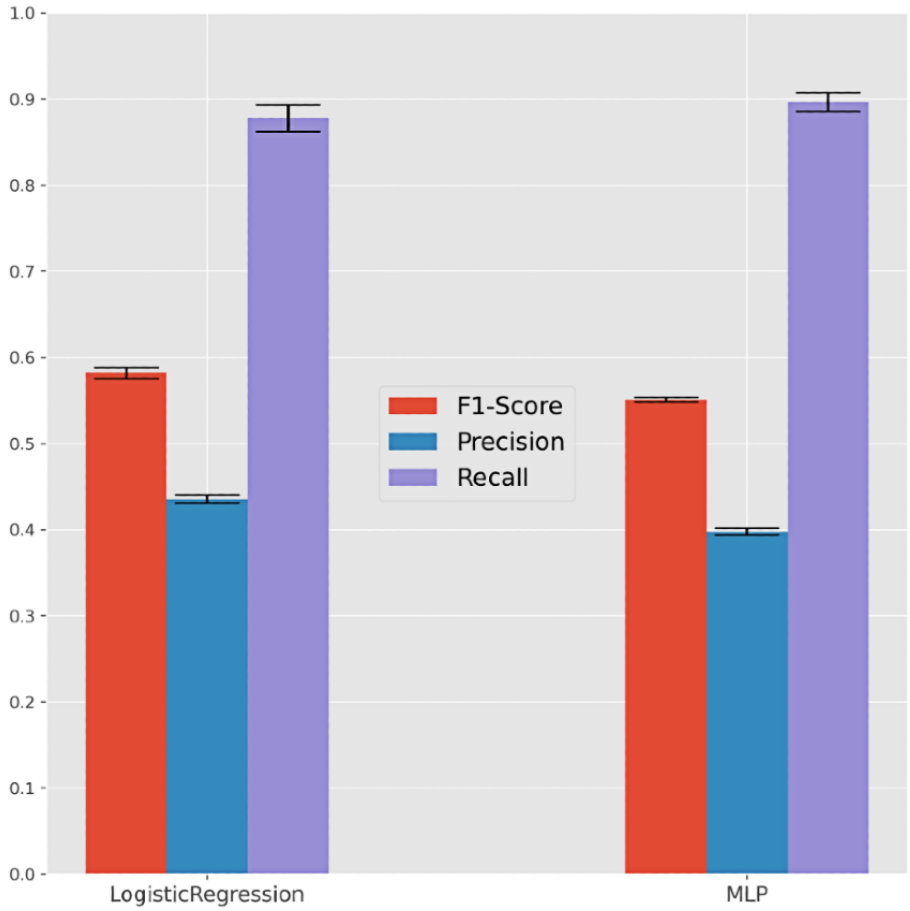
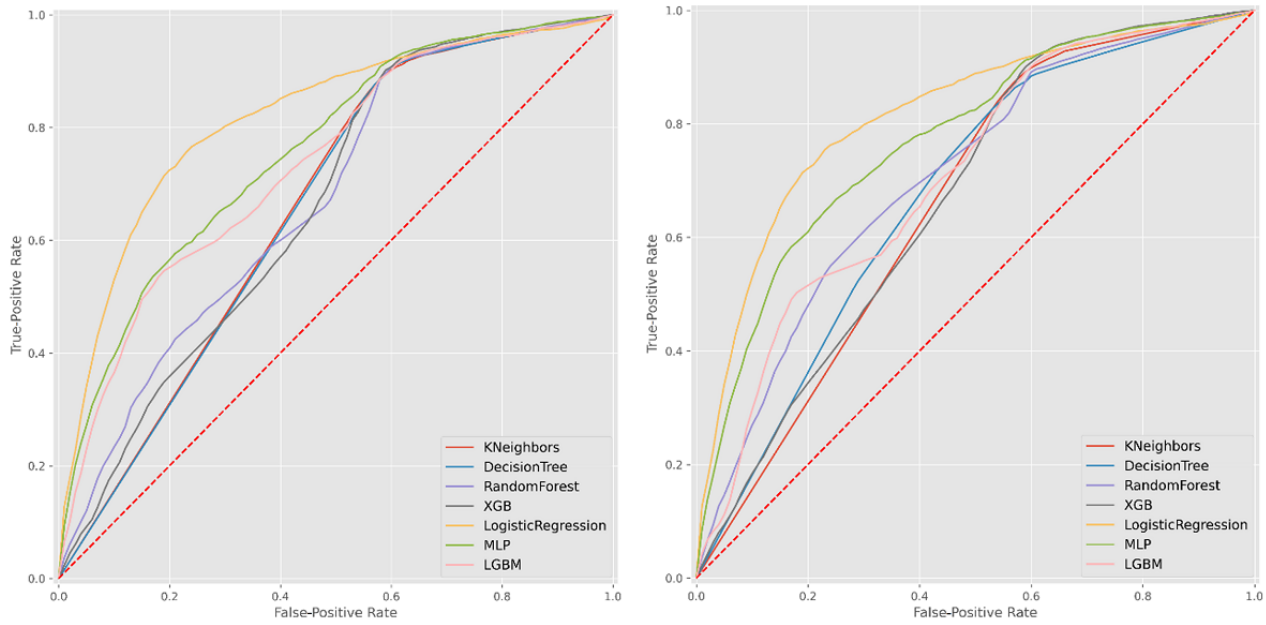


Table 2. Predictability of hospitalizations per predictive model.

Model	F1 optimization, mean (SD)		Recall-oriented ($F_{\beta=2}$) optimization, mean (SD)	
	F1-score	Recall	F1-score	Recall
KNN[a]	0.544 (0.007)	0.883 (0.020)	0.545 (0.005)	0.890 (0.017)
DT[b]	0.562 (0.030)	0.837 (0.069)	0.548 (0.004)	0.897 (0.007)
RF[c]	0.535 (0.010)	0.878 (0.016)	0.541 (0.005)	0.874 (0.029)
XGB[d]	0.546 (0.004)	0.897 (0.012)[e]	0.545 (0.004)	0.895 (0.011)
LR[f]	0.582 (0.006)[e]	0.878 (0.015)	0.583 (0.010)[e]	0.879 (0.015)
MLP[g]	0.549 (0.006)	0.892 (0.010)	0.551 (0.003)	0.897 (0.011)[e]
LGBM[h]	0.545 (0.005)	0.893 (0.013)	0.545 (0.005)	0.893 (0.016)

[a]KNN: k-nearest neighbors.
[b]DT: decision tree.
[c]RF: random forest.
[d]XGB: XGBoost.
[e]Best-performing models
[f]LR: logistic regression.
[g]MLP: multilayer perceptron.
[h]LGBM: LightGBM.

Figure 5. Receiver operating characteristic curves with the predictive behavior of the selected classifiers in asserting hospitalization needs at the time of SARS-CoV-2 testing. XGB: XGBoost, MLP: multilayer perceptron, LGBM: LightGBM.



(A) *F1 optimization* (B) *Fβ=2 (recall-oriented) optimization*

ICU Admissions

Figures 6 and 7 and Table 3 assess the ability to anticipate intensive care needs for infected individuals at two stages: before hospitalization and after hospitalization. To this end, the proposed methodology was pursued considering demographic factors, comorbidity factors, and the time to hospitalization for hospitalized individuals. Individuals without a SARS-CoV-2–negative test result after infection were excluded.

Figure 6. Predictability of intensive care unit admission. Results for the best F1 predictor (left) and recall-oriented predictor (right). XGB: XGBoost.

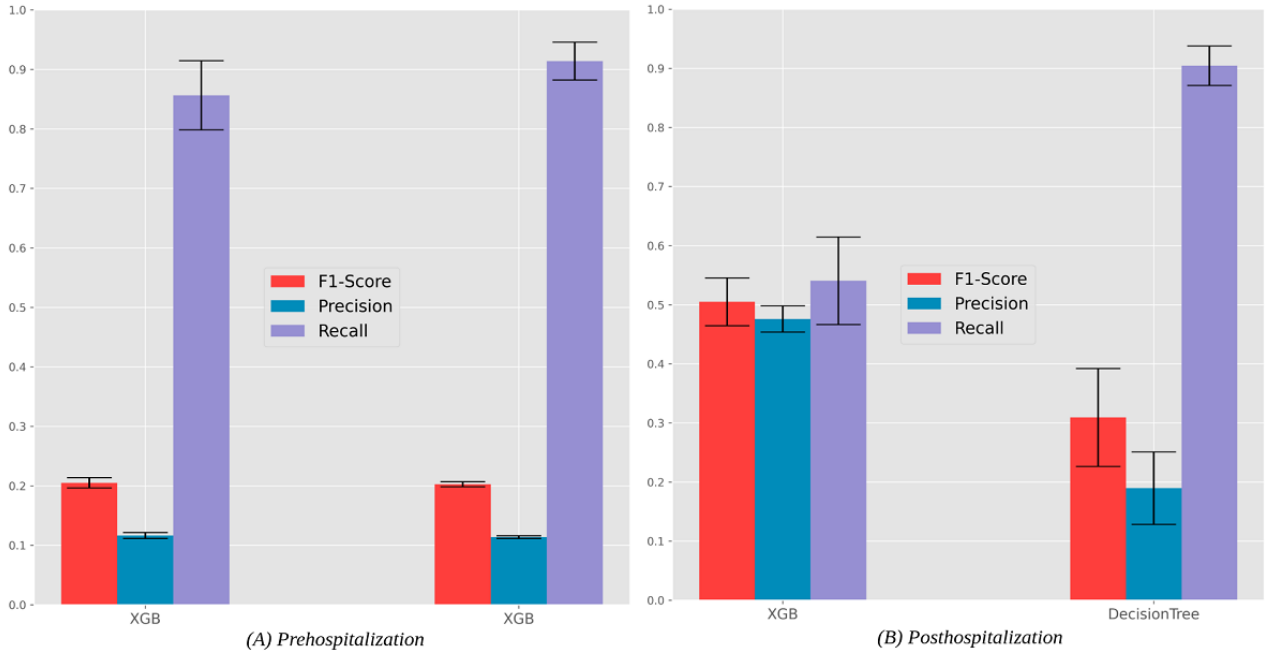


Figure 7. Receiver operating characteristic curves with the predictive behavior of the selected classifiers in predicting intensive care unit admission needs at the time of hospitalization. XGB: XGBoost, MLP: multilayer perceptron, LGBM: LightGBM.

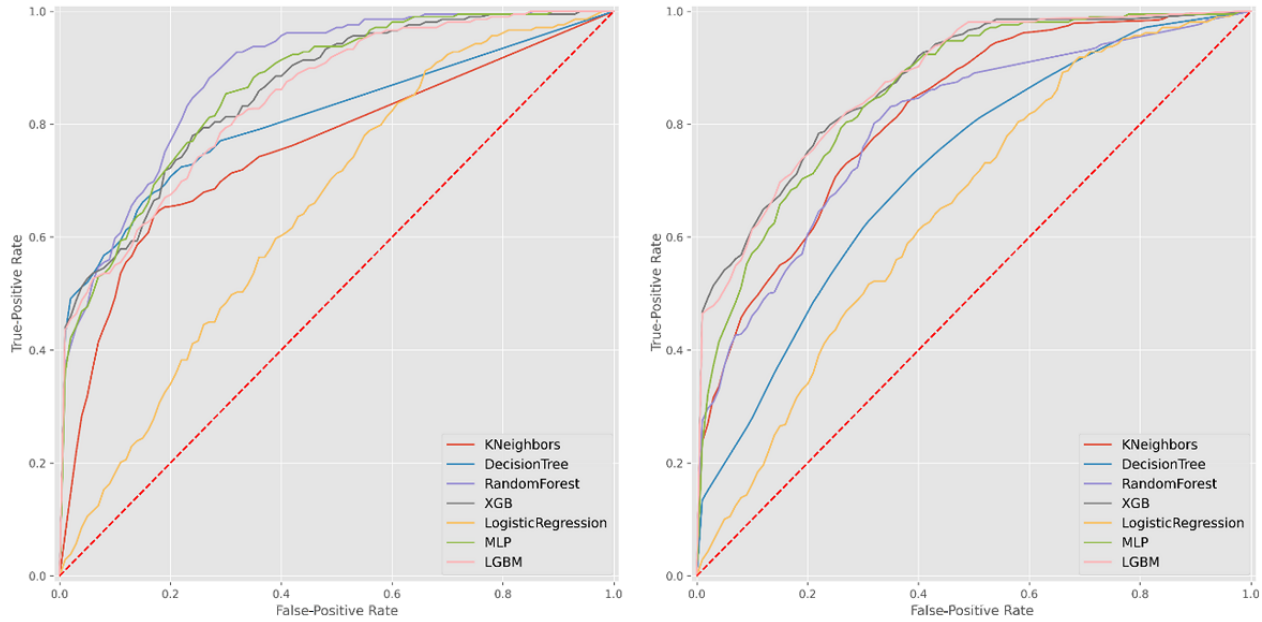


Table 3. Predictability of intensive care unit admissions per predictive model.

Model	F1 optimization, mean (SD)		$F_{\beta=2}$ (recall-oriented) optimization, mean (SD)	
	F1-score	Recall	F1-score	Recall
At the time of hospitalization				
KNN[a]	0.428 (0.039)	0.574 (0.156)	0.369 (0.037)	0.741 (0.088)
DT[b]	0.461 (0.016)	0.651 (0.056)	0.309 (0.083)	0.904 (0.033)[c]
RF[d]	0.454 (0.027)	0.713 (0.088)[c]	0.382 (0.103)	0.794 (0.128)
XGB[e]	0.505 (0.040)[c]	0.541 (0.074)	0.431 (0.040)	0.766 (0.084)
LR[f]	0.250 (0.015)	0.622 (0.041)	0.248 (0.013)	0.651 (0.074)
MLP[g]	0.449 (0.060)	0.703 (0.145)	0.410 (0.039)	0.818 (0.050)
LGBM[h]	0.480 (0.023)	0.536 (0.048)	0.435 (0.025)[c]	0.770 (0.051)
At the time of SARS-CoV-2 testing				
KNN	0.195 (0.016)	0.818 (0.090)	0.198 (0.007)	0.852 (0.039)
DT	0.209 (0.012)	0.752 (0.135)	0.201 (0.007)	0.890 (0.036)
RF	0.200 (0.008)	0.880 (0.038)	0.200 (0.008)	0.880 (0.044)
XGB	0.205 (0.009)	0.857 (0.058)	0.203 (0.004)	0.914 (0.032)
LR	0.200 (0.008)	0.847 (0.054)	0.201 (0.007)	0.880 (0.034)
MLP	0.202 (0.006)	0.871 (0.049)	0.200 (0.008)	0.880 (0.037)
LGBM	0.204 (0.012)	0.871 (0.074)	0.197 (0.012)	0.861 (0.066)

[a]KNN: k-nearest neighbors.

[b]DT: decision tree.

[c]Best-performing models.

[d]RF: random forest.

[e]XGB: XGBoost.

[f]LR: logistic regression.

[g]MLP: multilayer perceptron.

[h]LGBM: LightGBM.

The predictability of ICU needs is less satisfactory than hospitalization needs, particularly for the prehospitalization stage. We hypothesize that this difficulty was partially related to the smaller number of individuals with ICU internments, together with the presence of missing values associated with ICU internment needs for most individuals. Even though we can achieve recall levels over 90% with gradient boosting (XGBoost) in a posthospitalization setting, it comes at the cost of a considerably low precision (with one-third of predictions seen in practice). Still, the best-performing predictive models are suggested to support monitoring decisions at the hospital bedside, as their recall and specificity are considerably high.

Respiratory Support

Figure 8 and Table 4 assess respiratory assistance needs for hospitalized individuals with SARS-CoV-2, considering three assistance modes: (1) ventilation support, (2) oxygen therapy, and (3) combined ventilation and oxygen therapies. Demographic, comorbidity, and time-to-hospitalization factors were used as input variables.

Individuals without a SARS-CoV-2–negative test result after infection were excluded from this analysis. As respiratory support is a multiclass variable, we considered a different performance evaluation by focusing on (1) the recall for each major class (ventilation, oxygen, and nonrequired support), (2) the precision of individuals with oxygen or ventilation assistance, and (3) the Cohen kappa coefficient.

XGBoost, LightGBM, and random forests attained a satisfactory identification of hospitalized individuals who may require respiratory support in the future, generally providing recalls for each assistance mode around 60% at the cost of a 40% precision. According to the conducted methodology, they are thus pinpointed as good candidates to support in-hospital care decisions.

Figure 8. Predictability of respiratory support needs—assisted ventilation, oxygen therapy, and combined support—for hospitalized individuals with SARS-CoV-2. Performance of the best F1 predictor (right) and recall-oriented predictor (left) is shown. XGB: XGBoost, LGBM: LightGBM.

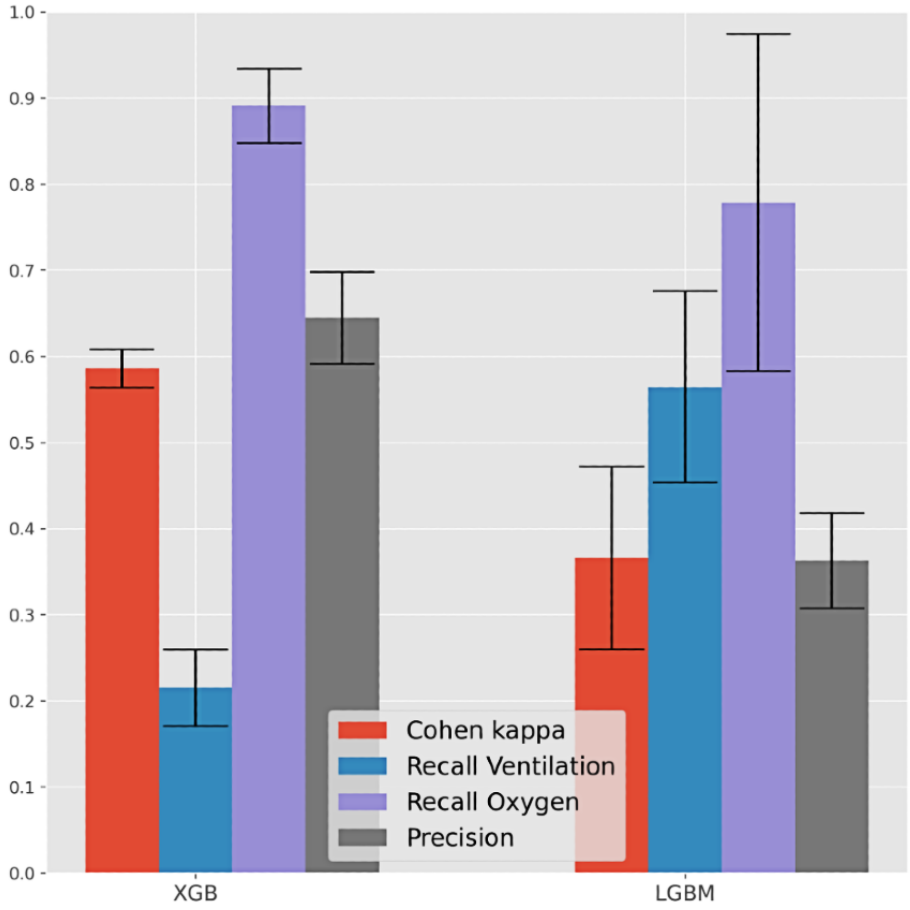


Table 4. Predictability of respiratory needs per predictive model.

Model	Kappa optimization, kappa, mean (SD)	Recalls optimization average, kappa, mean (SD)
KNN[a]	0.324 (0.050)	0.162 (0.022)
DT[b]	0.708 (0.017)[c]	0.110 (0.126)
RF[d]	0.518 (0.013)	0.017 (0.014)
XGB[e]	0.586 (0.022)	0.043 (0.013)
LR[f]	0.080 (0.009)	0.070 (0.009)
MLP[g]	0.464 (0.046)	0.204 (0.164)
LGBM[h]	0.567 (0.037)	0.366 (0.106)[c]

[a]KNN: k-nearest neighbors.
[b]DT: decision tree.
[c]Best-performing models.
[d]RF: random forest.
[e]XGB: XGBoost.
[f]LR: logistic regression.
[g]MLP: multilayer perceptron.
[h]LGBM: LightGBM.

Survivability (Outcome)

Finally, Figures 9 and 10 and Table 5 provide an analysis of the ability of the models to predict recovery-or-death outcomes for individuals with SARS-CoV-2 infection at three time points: (1) before hospitalization (at the time of testing), (2) after hospitalization, and (3) after ICU admission when applicable. To this end, we preserved the input variables and validation

methodology (see *Methods* section) considered in previous scenarios.

Our results showed a high ability to identify death outcomes. However, at the SARS-CoV-2 testing stage, this comes at a cost of incorrectly classifying two-thirds of individuals susceptible to death. In the posthospitalization scenario, we achieved more balanced results, with both precision and recall around 75% using gradient boosting (XGBoost and LightGBM). The introduction of the intensive care variable hampered the results since it restricted the analysis of deaths to individuals with acute needs and dependent on continuous care instruments.

Figure 9. Predictability of the survivability (outcome) of infected individuals at 3 stages. Results for the best F1 predictor (left) and recall-optimized predictor (right) per stage are shown. XGB: XGBoost, LGBM: LightGBM.

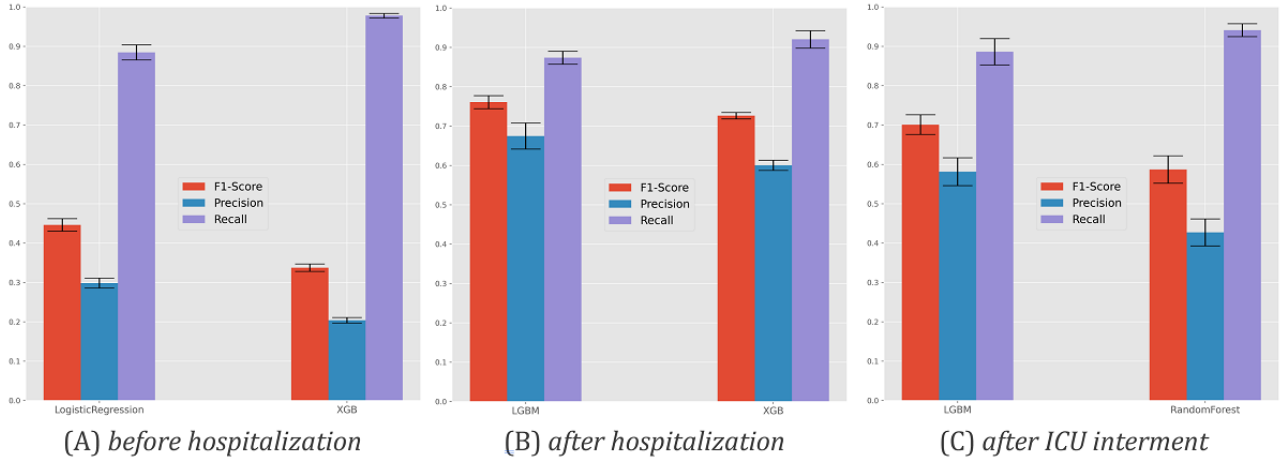


Figure 10. Receiver operating characteristic curves with the predictive behavior of the selected classifiers in asserting patient survivability at the time of hospitalization. XGB: XGBoost, MLP: multilayer perceptron, LGBM: LightGBM.

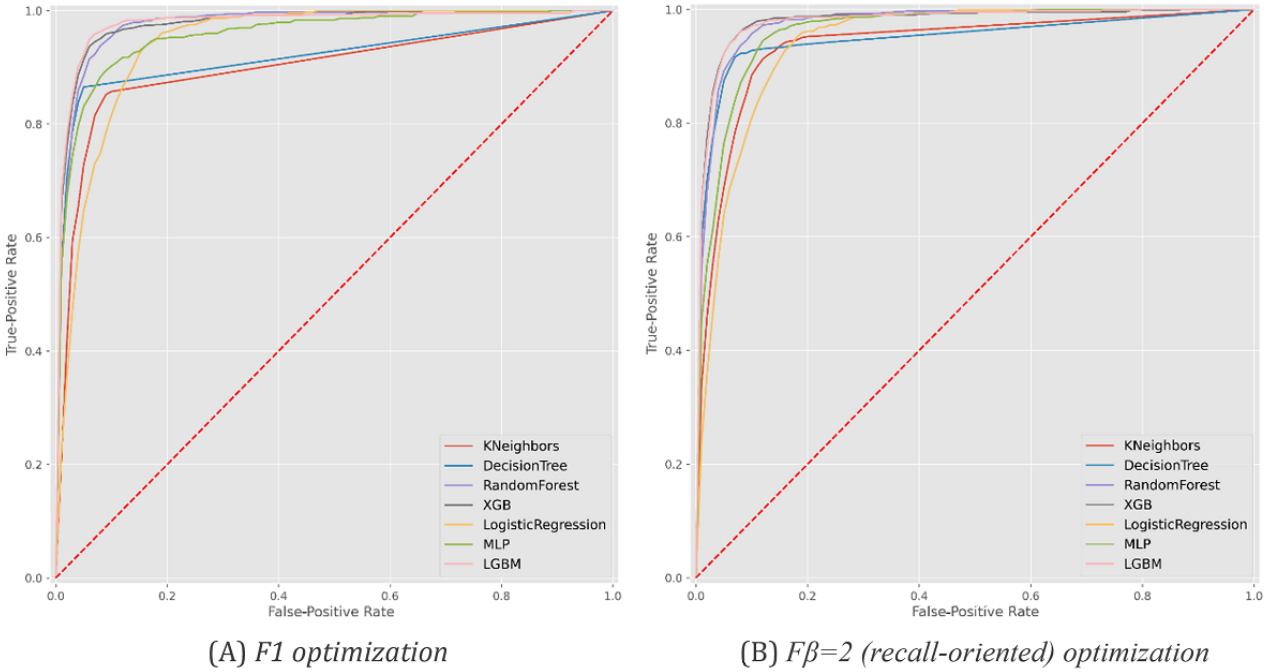


Table 5. Predictability of survivability per predictive model.

Model	F1 optimization, mean (SD)		$F_{\beta=2}$ (recall-oriented) optimization, mean (SD)	
	F1-score	Recall	F1-score	Recall
At the time of hospitalization				
KNN[a]	0.616 (0.035)	0.735 (0.048)	0.546 (0.021)	0.901 (0.030)
DT[b]	0.707 (0.013)	0.864 (0.011)	0.673 (0.022)	0.908 (0.017)
RF[c]	0.696 (0.030)	0.901 (0.030)	0.666 (0.021)	0.666 (0.021)
XGB[d]	0.765 (0.025)	0.834 (0.042)	0.726 (0.008)	0.920 (0.022)
LR[e]	0.492 (0.012)	0.909 (0.019)	0.476 (0.035)	0.916 (0.022)
MLP[f]	0.681 (0.024)	0.824 (0.027)	0.569 (0.020)	0.922 (0.023)
LGBM[g]	0.761 (0.017)	0.874 (0.016)	0.717 (0.036)	0.922 (0.021)
At the time of intensive care unit admission				
KNN	0.582 (0.040)	0.740 (0.053)	0.527 (0.030)	0.885 (0.049)
DT	0.652 (0.045)	0.879 (0.035)	0.638 (0.032)	0.922 (0.023)
RF	0.630 (0.018)	0.908 (0.039)	0.587 (0.035)	0.941 (0.016)
XGB	0.703 (0.035)	0.838 (0.068)	0.672 (0.021)	0.918 (0.051)
LR	0.497 (0.018)	0.908 (0.031)	0.470 (0.049)	0.920 (0.028)
MLP	0.633 (0.044)	0.790 (0.094)	0.529 (0.019)	0.935 (0.020)
LGBM	0.701 (0.025)	0.886 (0.034)	0.672 (0.024)	0.915 (0.027)

[a]KNN: k-nearest neighbors.

[b]DT: decision tree.

[c]RF: random forest.

[d]XGB: XGBoost.

[e]LR: logistic regression.

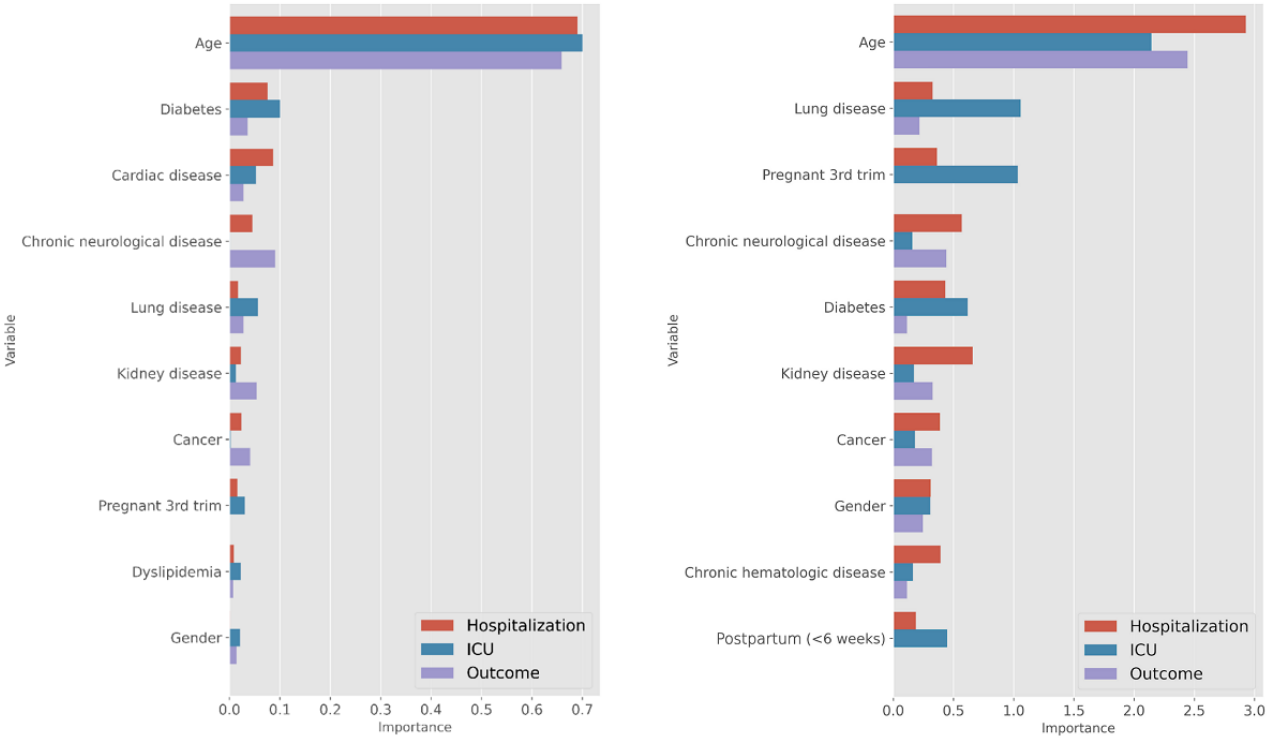
[f]MLP: multilayer perceptron.

[g]LGBM: LightGBM.

Determinants of Predictability

To assess the determinant factors underlying the achieved predictability levels, we first statistically tested the correlation between input and output variables using chi-square tests, ANOVA (analysis of variance), and their nonparametric counterparts, yielding results similar to those by Nogueira et al [33]. For a more in-depth understanding of the feature relevance for the assessed predictive models, Figures 11 and 12 illustrate the importance of the top features. To this end, we considered relevance outputs from gradient boosting (XGBoost) due to its competitively high performance across all outcomes, as well as the logistic regression for the hospitalization outcome by computing the Wald statistic to assess the significance of the coefficients for predictions.

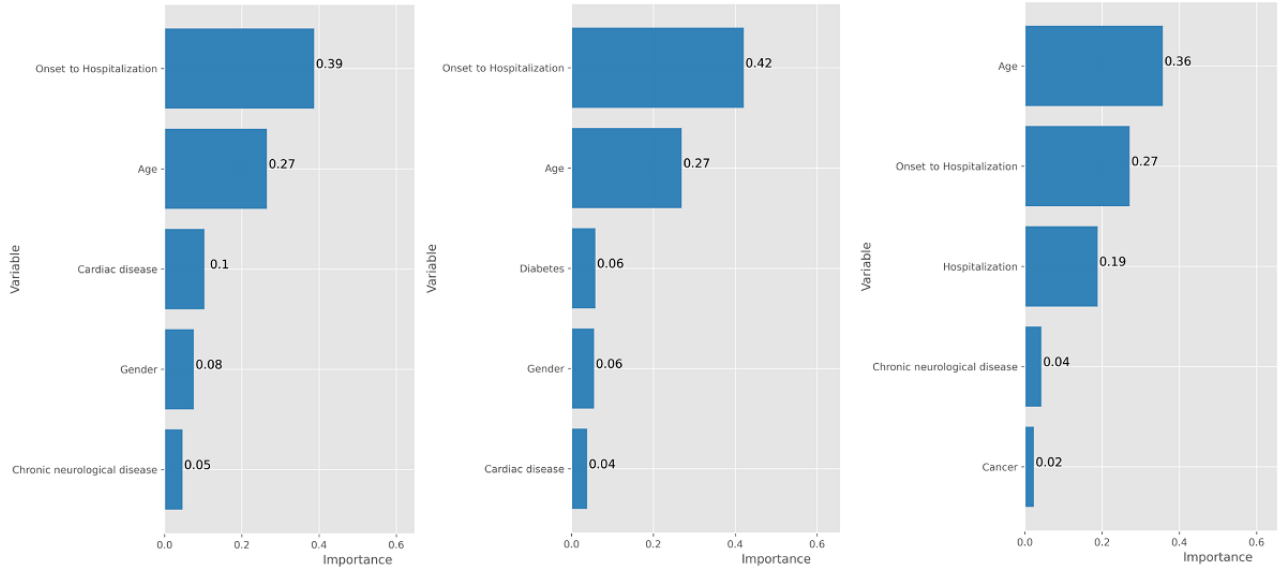
Figure 11. Top features and their importance for each target variable at the time of SARS-CoV-2 testing (prehospitalization). ICU: intensive care unit, LR: logistic regression.



(A) *XGBoost importance* (B) *LR importance (Wald statistic)*

We can observe that XGBoost distinguishes the relevance of different comorbidities for the target variables along each stage of the care process. In addition to the age variable, the onset period to hospitalization in days was also found to be a critical factor affecting the decisions (Figure 12). The high relevance of this variable consistently had top rank among associative models—XGBoost, random forests, and decision trees—pinpointing the importance of its collection for computer-aided predictions of ICU internment and respiratory needs.

Figure 12. XGBoost top features and their importance for the different outcomes at posthospitalization. ICU: intensive care unit.

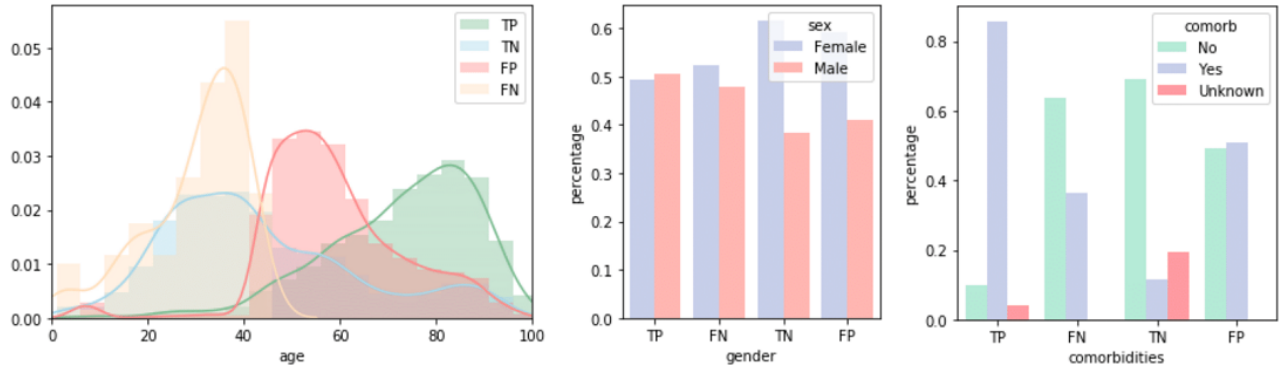


(A) ICU *admission* (B) *respiratory needs* (C) *survivability*

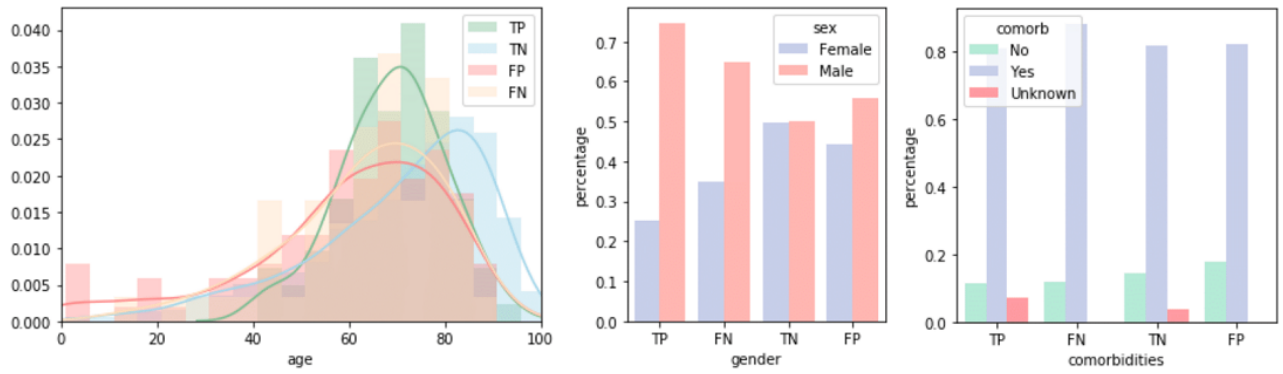
Complementarily, Figure 13 offers additional insights into the target predictive tasks by plotting some of the characteristics of the correctly classified individuals against incorrectly classified individuals with XGBoost. Particular attention should be paid to the differences between true positives and false negatives, that is, to the individuals requiring care, in order to guarantee their timely and proper assistance. The susceptibility

to false negatives is higher for individuals within the 40-60-year age category and without comorbidities.

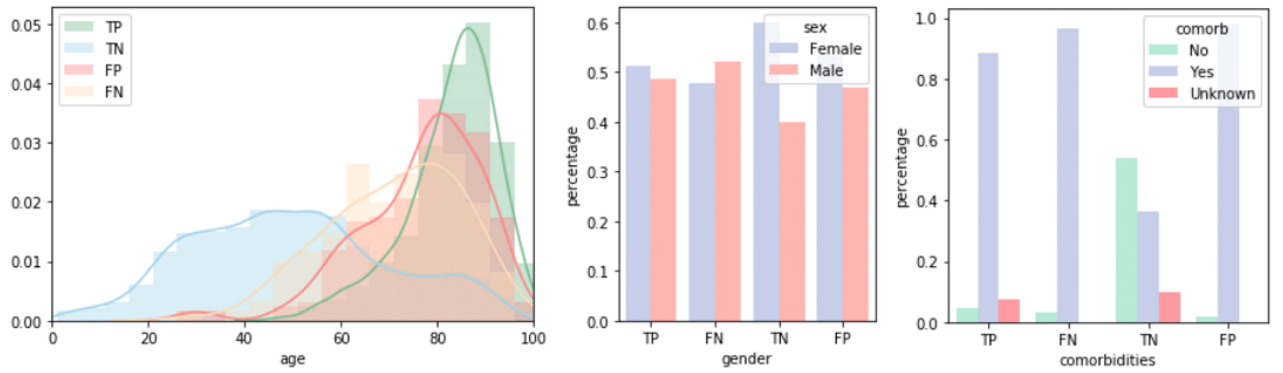
Figure 13. The characteristics of incorrectly predicted individuals with XGBoost. Particular attention should be paid to the differences between true-positive and false-negative individuals given their clinical relevance. ICU: intensive care unit.



(A) *Hospitalizations: profile of false-positive* (FP) *and false-negative* (FN) *individuals*



(B) ICU *admissions* (*posthospitalization*): *profile of false-positive* (FP) *and false-negative* (FN) *individuals*

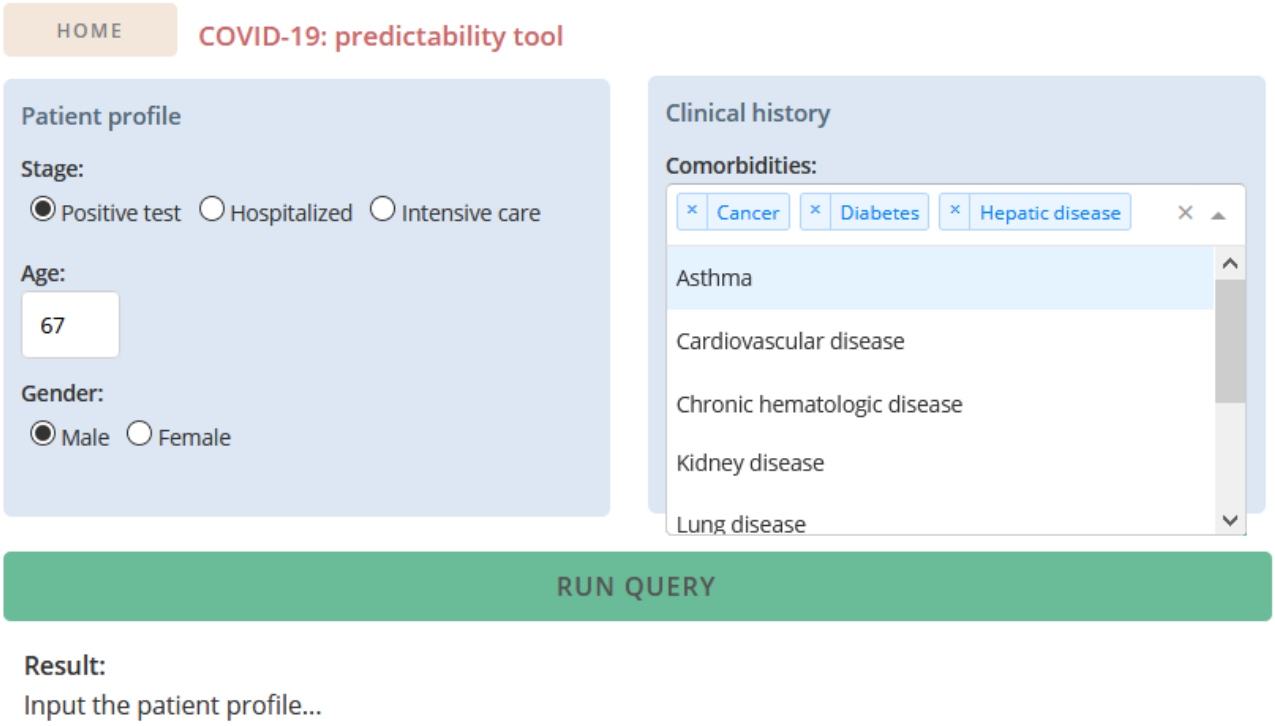


(C) *Survivability: profile of false-positive* (FP) *and false-negative* (FN) *individuals*

Clinical Decision Support System

The learned predictive models based on simple variables (stage, age, gender, and comorbidities) have been made available to health care providers within a recommendation system with graphical facilities. The serialized predictive models are used for the efficient testing of individuals at the different stages of the care cycle (testing, hospitalization, ICU admission) for the different outcome variables (care needs) after inserting essential demographic and comorbidity features. The output provides a bounded statistic based on the estimation returned by the predictive models achieving better recall and F1-measure for each outcome variable. Figure 14 provides a visualization of the graphical interface.

Figure 14. Snapshot of the provided clinical decision support system.



The variables required for each outcome score calculation are usually available at hospitals, and the tool is easy to use. Although recommendations are provided within a statistical frame, the tool does not categorize the risk into low- or high-risk patients as clinical experts are more informed to approximate this risk. In addition, we advise caution for clinicians who intend to use this tool as a predictive guide, especially for survivability analysis. Clinicians must balance the predictions from this tool against their practical experience.

In collaboration with DGS, our predictors are expected to be provided within public hospitals and care contact centers of the Portuguese Health Service (Serviço Nacional de Saúde), particularly to support remote care monitoring decisions.

The decision support system is available as a software tool on GitHub [34].

Discussion

Principal Findings

This work offers a discussion on the predictability of hospitalization needs, ICU admissions, respiratory assistance, and survivability outcome in individuals infected with SARS-CoV-2 in Portugal as of June 30, 2020. A retrospective cohort with all confirmed COVID-19 cases since March, encompassing demographic and comorbidity variables, was considered as the target population in this study.

The results for the given cohort reveal that (1) over 75% of hospitalization needs can be identified at the time of SARS-CoV-2 testing (with >50% precision); (2) ICU needs are generally less predictable at both the pre- and posthospitalization stages in the given cohort; (3) respiratory assistance needs (including ventilation support, oxygen therapy, and combined ventilation-oxygen support) achieved recall levels above 60% (with >50% precision); (4) death risk along different stages (testing time, after hospitalization, and after ICU admission) had the highest degree of predictability.

The predictive models yielding better accuracy performance were associative classifiers, particularly XGBoost and RandomForests, neural networks with hyperparameterized architectures, and logistic regressors, with the optimal choice varying in accordance with the target variable and evaluation measure.

Publications on COVID-19 using machine learning models for different outcomes have been rapidly increasing. Gao et al [35] developed a model that includes the mortality risk prediction and reported an F1 ranging from 0.65 to 0.69 (κ=0.61-0.65), in line with our findings. Alternative studies [28,36,37] offer additional results for generalizing results and identifying population-specific differences. Yet, most of these studies do not comprehensively assess models' performance or the cohort characteristics, impeding solid cross-population findings.

Limitations

This study has some inherent shortcomings that should be noted: (1) the number of clinical variables for the outcomes of interest were limited (eg, BMI and clinical symptoms were missing); (2) further external validation of the selected models is required; and (3) although some inconsistencies (listed in the *Cohort Description* section) and missing/unknown entries in the original DGS data set were excluded, data acquisition problems may still persist and influence the outcomes of this work. The fully autonomous and parameter-free nature of the proposed computational approach/models allows it to be dynamically retrained with updated data.

Concluding Remarks

In this work, we developed a web-based clinical decision support tool without biological variables as input that can be used by clinicians. The conducted work pinpoints the relevance of the proposed predictive models to aid medical decisions for the Portuguese population, including both remote monitoring and in-hospital care decisions. Predicting the most probable outcomes along the life cycle of a SARS-CoV-2–infected individual can identify patients who are expected to develop severe illness, thus optimizing the allocation of health care resources and supporting more vulnerable patients.

Acknowledgments

The authors thank Portuguese Directorate General of Health (DGS) for providing the data. Data are available upon reasonable request. This work was supported by Fundação para a Ciência e a Tecnologia (FCT), through IDMEC, under LAETA project (UIDB/50022/2020), IPOscore (DSAIPA/DS/0042/2018) and Data2Help (DSAIPA/AI/ 0044/2018) projects, the contract CEECIND/01399/2017 to RSC, FCT/MCTES funds for INESC-ID (UIDB/50021/2020) and the Associate Laboratory for Green Chemistry - LAQV (UIDB/50006/2020 and UIDP/50006/2020).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Optimized parameters for the best-performing classifiers (Figures 4, 6, 8, and 9).
[PNG File , 71 KB - jmir_v23i4e26075_app1.png]

References

1. Chan J, Yuan S, Kok K, To K, Chu H, Yang J, et al. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. The Lancet 2020 Feb;395(10223):514-523. [doi: 10.1016/s0140-6736(20)30154-9]
2. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, et al. A Novel Coronavirus from Patients with Pneumonia in China, 2019. N Engl J Med 2020 Feb 20;382(8):727-733. [doi: 10.1056/nejmoa2001017]
3. Fu L, Wang B, Yuan T, Chen X, Ao Y, Fitzpatrick T, et al. Clinical characteristics of coronavirus disease 2019 (COVID-19) in China: A systematic review and meta-analysis. J Infect 2020 Jun;80(6):656-665 [FREE Full text] [doi: 10.1016/j.jinf.2020.03.041] [Medline: 32283155]
4. World Health Organization. URL: https://www.who.int [accessed 2020-09-01]
5. Singh AK, Gupta R, Ghosh A, Misra A. Diabetes in COVID-19: Prevalence, pathophysiology, prognosis and practical considerations. Diabetes Metab Syndr 2020 Jul;14(4):303-310 [FREE Full text] [doi: 10.1016/j.dsx.2020.04.004] [Medline: 32298981]
6. Sanyaolu A, Okorie C, Marinkovic A, Patidar R, Younis K, Desai P, et al. Comorbidity and its Impact on Patients with COVID-19. SN Compr Clin Med 2020 Jun 25;2(8):1-8 [FREE Full text] [doi: 10.1007/s42399-020-00363-4] [Medline: 32838147]
7. Wang Y, Lu X, Li Y, Chen H, Chen T, Su N, et al. Clinical Course and Outcomes of 344 Intensive Care Patients with COVID-19. Am J Respir Crit Care Med 2020 Jun 01;201(11):1430-1434. [doi: 10.1164/rccm.202003-0736le]
8. Guan W, Liang W, Zhao Y, Liang H, Chen Z, Li Y, China Medical Treatment Expert Group for COVID-19. Comorbidity and its impact on 1590 patients with COVID-19 in China: a nationwide analysis. Eur Respir J 2020 May 26;55(5):2000547 [FREE Full text] [doi: 10.1183/13993003.00547-2020] [Medline: 32217650]
9. Du RH, Liang LR, Yang CQ, Wang W, Cao TZ, Li M, et al. Predictors of mortality for patients with COVID-19 pneumonia caused by SARS-CoV-2: a prospective cohort study. Eur Respir J 2020 Sep 24;56(3):2050524. [doi: 10.1183/13993003.50524-2020] [Medline: 32973076]
10. Shi Y, Yu X, Zhao H, Wang H, Zhao R, Sheng J. Host susceptibility to severe COVID-19 and establishment of a host risk score: findings of 487 cases outside Wuhan. Crit Care 2020 Mar 18;24(1):108 [FREE Full text] [doi: 10.1186/s13054-020-2833-7] [Medline: 32188484]
11. Hu H, Yao N, Qiu Y. Comparing Rapid Scoring Systems in Mortality Prediction of Critically Ill Patients With Novel Coronavirus Disease. Acad Emerg Med 2020 Jun 21;27(6):461-468 [FREE Full text] [doi: 10.1111/acem.13992] [Medline: 32311790]
12. Cho S, Park S, Song M, Bae YY, Lee D, Kim D. Prognosis Score System to Predict Survival for COVID-19 Cases: a Korean Nationwide Cohort Study. J Med Internet Res 2021 Feb 22;23(2):e26257 [FREE Full text] [doi: 10.2196/26257] [Medline: 33539312]
13. Agosto A, Giudici P. A Poisson Autoregressive Model to Understand COVID-19 Contagion Dynamics. SSRN Journal Preprint posted online Mar 10, 2020. [doi: 10.2139/ssrn.3551626]

14. Li X, Xu S, Yu M, Wang K, Tao Y, Zhou Y, et al. Risk factors for severity and mortality in adult COVID-19 inpatients in Wuhan. J Allergy Clin Immunol 2020 Jul;146(1):110-118 [FREE Full text] [doi: 10.1016/j.jaci.2020.04.006] [Medline: 32294485]
15. Barman RK, Mukhopadhyay A, Maulik U, Das S. Identification of infectious disease-associated host genes using machine learning techniques. BMC Bioinformatics 2019 Dec 27;20(1):736 [FREE Full text] [doi: 10.1186/s12859-019-3317-0] [Medline: 31881961]
16. Luz C, Vollmer M, Decruyenaere J, Nijsten M, Glasner C, Sinha B. Machine learning in infection management using routine electronic health records: tools, techniques, and reporting of future technologies. Clin Microbiol Infect 2020 Oct;26(10):1291-1299. [doi: 10.1016/j.cmi.2020.02.003] [Medline: 32061798]
17. Boutilier J, Chan T, Ranjan M, Deo S. Risk Stratification for Early Detection of Diabetes and Hypertension in Resource-Limited Settings: Machine Learning Analysis. J Med Internet Res 2021 Jan 21;23(1):e20123 [FREE Full text] [doi: 10.2196/20123] [Medline: 33475518]
18. Sujath R, Chatterjee JM, Hassanien AE. A machine learning forecasting model for COVID-19 pandemic in India. Stoch Environ Res Risk Assess 2020 May 30;34(7):959-972. [doi: 10.1007/s00477-020-01827-8]
19. Chassagnon G, Vakalopoulou M, Battistella E, Christodoulidis S, Hoang-Thi T, Dangeard S, et al. AI-driven quantification, staging and outcome prediction of COVID-19 pneumonia. Med Image Anal 2021 Jan;67:101860 [FREE Full text] [doi: 10.1016/j.media.2020.101860] [Medline: 33171345]
20. Assaf D, Gutman Y, Neuman Y, Segal G, Amit S, Gefen-Halevi S, et al. Utilization of machine-learning models to accurately predict the risk for critical COVID-19. Intern Emerg Med 2020 Nov 18;15(8):1435-1443 [FREE Full text] [doi: 10.1007/s11739-020-02475-0] [Medline: 32812204]
21. Wu G, Yang P, Xie Y, Woodruff HC, Rao X, Guiot J, et al. Development of a clinical decision support system for severity risk prediction and triage of COVID-19 patients at hospital admission: an international multicentre study. Eur Respir J 2020 Aug 02;56(2):2001104 [FREE Full text] [doi: 10.1183/13993003.01104-2020] [Medline: 32616597]
22. Wynants L, Van Calster B, Collins GS, Riley RD, Heinze G, Schuit E, et al. Prediction models for diagnosis and prognosis of covid-19 infection: systematic review and critical appraisal. BMJ 2020 Apr 07;369:m1328 [FREE Full text] [doi: 10.1136/bmj.m1328] [Medline: 32265220]
23. Burdick H, Lam C, Mataraso S, Siefkas A, Braden G, Dellinger RP, et al. Prediction of respiratory decompensation in Covid-19 patients using machine learning: The READY trial. Comput Biol Med 2020 Sep;124:103949 [FREE Full text] [doi: 10.1016/j.compbiomed.2020.103949] [Medline: 32798922]
24. Liang W, Liang H, Ou L, Chen B, Chen A, Li C, China Medical Treatment Expert Group for COVID-19. Development and Validation of a Clinical Risk Score to Predict the Occurrence of Critical Illness in Hospitalized Patients With COVID-19. JAMA Intern Med 2020 Aug 01;180(8):1081-1089 [FREE Full text] [doi: 10.1001/jamainternmed.2020.2033] [Medline: 32396163]
25. Gayam V, Chobufo MD, Merghani MA, Lamichhane S, Garlapati PR, Adler MK. Clinical characteristics and predictors of mortality in African-Americans with COVID-19 from an inner-city community teaching hospital in New York. J Med Virol 2021 Feb 05;93(2):812-819 [FREE Full text] [doi: 10.1002/jmv.26306] [Medline: 32672844]
26. Núñez-Gil IJ, Fernández-Pérez C, Estrada V, Becerra-Muñoz VM, El-Battrawy I, Uribarri A, HOPE COVID-19 Investigators. Mortality risk assessment in Spain and Italy, insights of the HOPE COVID-19 registry. Intern Emerg Med 2020 Nov 09;51(1):e13436 [FREE Full text] [doi: 10.1007/s11739-020-02543-5] [Medline: 33165755]
27. Iwendi C, Bashir AK, Peshkar A, Sujatha R, Chatterjee JM, Pasupuleti S, et al. COVID-19 Patient Health Prediction Using Boosted Random Forest Algorithm. Front Public Health 2020 Jul 3;8:357 [FREE Full text] [doi: 10.3389/fpubh.2020.00357] [Medline: 32719767]
28. Vaid A, Somani S, Russak AJ, De Freitas JK, Chaudhry FF, Paranjpe I, et al. Machine Learning to Predict Mortality and Critical Events in a Cohort of Patients With COVID-19 in New York City: Model Development and Validation. J Med Internet Res 2020 Nov 06;22(11):e24018 [FREE Full text] [doi: 10.2196/24018] [Medline: 33027032]
29. Snoek J, Larochelle H, Adams RP. Practical Bayesian optimization of machine learning algorithms. In: NIPS'12: Proceedings of the 25th International Conference on Neural Information Processing Systems. Red Hook, NY: Curran Associates Inc; 2012 Presented at: International Conference on Neural Information Processing Systems; Dec 3-6; Nevada p. 2951-2959 URL: https://papers.nips.cc/paper/2012/file/05311655a15b75fab86956663e1819cd-Paper.pdf
30. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine learning in Python. J Mach Learn Res 2011;12:2825-2830 [FREE Full text]
31. Chen T, Guestrin C. XGBoost: A scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2016 Presented at: KDD '16: The 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; Aug 13-16; San Francisco, CA p. 785-794. [doi: 10.1145/2939672.2939785]
32. Plotly. URL: https://plotly.com/dash/ [accessed 2021-04-13]
33. Nogueira PJ, de Araújo Nobre M, Costa A, Ribeiro RM, Furtado C, Bacelar Nicolau L, et al. The Role of Health Preconditions on COVID-19 Deaths in Portugal: Evidence from Surveillance Data of the First 20293 Infection Cases. J Clin Med 2020 Jul 24;9(8):3442 [FREE Full text] [doi: 10.3390/jcm9082368] [Medline: 32722159]

34. GitHub. URL: https://github.com/Andrempp/COVIDCalculator [accessed 2021-04-13]
35. Gao Y, Cai G, Fang W, Li H, Wang S, Chen L, et al. Machine learning based early warning system enables accurate mortality risk prediction for COVID-19. Nat Commun 2020 Oct 06;11(1):5033 [FREE Full text] [doi: 10.1038/s41467-020-18684-2] [Medline: 33024092]
36. Hu C, Liu Z, Jiang Y, Shi O, Zhang X, Xu K, et al. Early prediction of mortality risk among patients with severe COVID-19, using machine learning. Int J Epidemiol 2021 Jan 23;49(6):1918-1929 [FREE Full text] [doi: 10.1093/ije/dyaa171] [Medline: 32997743]
37. Yadaw AS, Li Y, Bose S, Iyengar R, Bunyavanich S, Pandey G. Clinical features of COVID-19 mortality: development and validation of a clinical prediction model. The Lancet Digital Health 2020 Oct;2(10):e516-e525. [doi: 10.1016/s2589-7500(20)30217-x]

Abbreviations

ANOVA: analysis of variance
DGS: Direcção-Geral da Saúde
DT: decision tree
ICU: intensive care unit
KNN: k-nearest neighbors
LGBM: Light Gradient Boosting Machine
MLP: multilayer perceptron
SMOTE: Synthetic Minority Oversampling Technique
XGB: XGBoost

Edited by C Basch; submitted 26.11.20; peer-reviewed by S Hyland, X Zhang, Z Zhang; comments to author 28.01.21; revised version received 14.02.21; accepted 18.03.21; published 28.04.21.

Please cite as:
Patrício A, Costa RS, Henriques R
Predictability of COVID-19 Hospitalizations, Intensive Care Unit Admissions, and Respiratory Assistance in Portugal: Longitudinal Cohort Study
J Med Internet Res 2021;23(4):e26075
URL: https://www.jmir.org/2021/4/e26075
doi:10.2196/26075
PMID:33835931

Original Paper

Telemanagement of Home-Isolated COVID-19 Patients Using Oxygen Therapy With Noninvasive Positive Pressure Ventilation and Physical Therapy Techniques: Randomized Clinical Trial

Aya Sedky Adly[1,2], MSc, PhD; Mahmoud Sedky Adly[3,4], MSc, DDS, HDD; Afnan Sedky Adly[5,6], MSc

[1]Faculty of Computers and Artificial Intelligence, Helwan University, Cairo, Egypt

[2]Faculty of Engineering and Technology, Badr University in Cairo (BUC), Cairo, Egypt

[3]Faculty of Oral and Dental Medicine, Cairo University, Cairo, Egypt

[4]Royal College of Surgeons of Edinburgh, Scotland, United Kingdom

[5]Faculty of Physical Therapy, Cardiovascular-Respiratory Disorders and Geriatrics, Laser Applications in Physical Medicine, Cairo University, Cairo, Egypt

[6]Faculty of Physical Therapy, Internal Medicine, Beni-Suef University, Beni-Suef, Egypt

Corresponding Author:
Aya Sedky Adly, MSc, PhD
Faculty of Engineering and Technology
Badr University in Cairo (BUC)
Cairo Suez Road
Badr City
Cairo
Egypt
Phone: 20 1145559778
Email: aya.sedky@gmail.com

Related Articles:

Comment in: https://www.jmir.org/2022/5/e34437

Comment in: https://www.jmir.org/2022/5/e37413

Abstract

Background: With the growing stress on hospitals caused by the COVID-19 pandemic, the need for home-based solutions has become a necessity to support these overwhelmed hospitals.

Objective: The goal of this study was to compare two nonpharmacological respiratory treatment methods for home-isolated COVID-19 patients using a newly developed telemanagement health care system.

Methods: In this single-blinded randomized clinical trial, 60 patients with stage 1 pneumonia caused by SARS-CoV-2 infection were treated. Group A (n=30) received oxygen therapy with bilevel positive airway pressure (BiPAP) ventilation, and Group B (n=30) received osteopathic manipulative respiratory and physical therapy techniques. Arterial blood gases of PaO_2 and $PaCO_2$, pH, vital signs (ie, temperature, respiratory rate, oxygen saturation, heart rate, and blood pressure), and chest computed tomography scans were used for follow-up and for assessment of the course and duration of recovery.

Results: Analysis of the results showed a significant difference between the two groups (P<.05), with Group A showing shorter recovery periods than Group B (mean 14.9, SD 1.7 days, and mean 23.9, SD 2.3 days, respectively). Significant differences were also observed between baseline and final readings in all of the outcome measures in both groups (P<.05). Regarding posttreatment satisfaction with our proposed telemanagement health care system, positive responses were given by most of the patients in both groups.

Conclusions: It was found that home-based oxygen therapy with BiPAP can be a more effective prophylactic treatment approach than osteopathic manipulative respiratory and physical therapy techniques, as it can impede exacerbation of early-stage COVID-19 pneumonia. Telemanagement health care systems are promising methods to help in the pandemic-related shortage of hospital

beds, as they showed reasonable effectiveness and reliability in the monitoring and management of patients with early-stage COVID-19 pneumonia.

Trial Registration: ClinicalTrials.gov NCT04368923; https://clinicaltrials.gov/ct2/show/NCT04368923

(J Med Internet Res 2021;23(4):e23446) doi:10.2196/23446

KEYWORDS

telemedicine; oxygen therapy; noninvasive positive airway pressure; BiPAP; osteopathic medicine; physical therapy; SARS-CoV-2; COVID-19; teletherapy; telemanagement

Introduction

COVID-19 has become a global pandemic that has had a gross dramatic impact on hospitals and health care systems worldwide [1-3]. Therefore, home isolation may become the only available option to stop the spread in most countries [4-6]. A worldwide shortage of medical devices, protective equipment, and pharmacological treatment with growing stress on hospital resources have led to an obvious drop in the performance of the majority of frontline health care workers [7-9]. Therefore, introducing telemanagement approaches has become a necessity for coping with the exponential increase in the number of people infected with SARS-CoV-2 [10-12].

Home care isolation can be considered for COVID-19 patients with mild illness when inpatient isolation is unavailable. However, advice and precautions regarding respiratory hygiene, environmental ventilation, hand hygiene, shared space confinement, and optimal nutritional intake should be followed as a part of the management process [13,14].

Current evidence suggests that the application of noninvasive bilevel positive airway pressure (BiPAP) can reduce pulmonary complications and raise pulmonary oxygen pressure by activation of collapsing alveoli and reduction of the degree of shunt [15]. However, noninvasive positive pressure ventilation is usually considered the last line of treatment in the initial management of hospitalized COVID-19 patients, as it is considered an aerosol-generating procedure that can increase the risk of infection in a hospital setting [16,17]. Yet, there is an apparent lack of research on its impact on COVID-19 patients, despite evidence from several previous studies that confirm its beneficial effects in different types of pneumonia.

COVID-19 disease progression and complications have been found to be unpredictable. Patients' deterioration has been commonly reported to be a result of pulmonary edema due to interstitial fluid accumulation from pulmonary capillary leakage, cytokine storms, and microvascular thrombosis. Pulmonary edema is characterized by acute onset and rapid progression [18]. Noninvasive BiPAP may prevent this deterioration if applied at an early stage by impeding the consequent pulmonary edema via positive pressure [19].

Recent studies indicate that osteopathic manipulative respiratory and physical therapy techniques can improve pulmonary function in both chronic and acute pulmonary conditions [20]. These techniques are directed toward the respiratory musculoskeletal components inducing thoracic pressure changes, which are essential for effective respiratory processes. Additionally, these techniques have enormous potential in the alleviation of pulmonary disease complications through increasing mobility of chest wall muscles and the diaphragm [21].

Osteopathic manipulative respiratory and physical therapy techniques have an important advantage of being low cost with the ability to be modified for home or self-application, which can reduce the risk of infection and overcome the shortage in pharmacological treatments and medical devices [22].

In the presence of COVID-19, telepractice has transformed physical therapy, as communication-based platforms beyond telerehabilitation, telemedicine, telemanipulation, and telehealth are utilized to advance remote access to therapy [23].

Recent evidence supported the use of self-directed web-based physical therapy over traditional outpatient physical therapy. Moreover, web-based physical therapy at home has gained interest owing to its interactive nature, which offers a more formal structure and direction for the patients [24]. Consequently, self-directed web-based physical therapy with remote supervision can be considered as an effective solution to offset the risks associated with infection among COVID-19 patients without compromising outcomes [24-26].

Noninvasive ventilation (NIV) has been utilized as an adjunct to physical therapy in patients with respiratory diseases. Several studies have demonstrated the positive effects of NIV and supplemental oxygen as adjuncts to physical therapy [27]. NIV was found to unload the ventilatory muscles and reverse the neural drive of fatigue through reducing the amount of necessary patient effort as well as reducing the muscle load during an assisted interactive breath [28]. Thus, investigation of these two treatment modalities among COVID-19 patients would be a great benefit.

Evidence has reported that real-time telemedicine in patients with acute respiratory infections showed similar rates of clinical management when compared to traditionally treated patients. Furthermore, the frequency of follow-up visits in real-time telemedicine was higher than in traditional follow-up visits [29].

Preliminary regulations have been issued regarding the control of infection, diagnosis, and monitoring of COVID-19 patients, while there is limited and unclear guidance on the effect of starting care management of these patients from an early stage [30-33]. The divergence of results in the majority of research studies on this topic and the need for high-quality randomized clinical trials that follow the recommended standards of reporting were the reasons for implementing this study.

Advances in information and communication technologies (ICTs) have turned the world into a connected village. Remarkable challenges have been addressed by ICTs in various health care sectors [34-36].

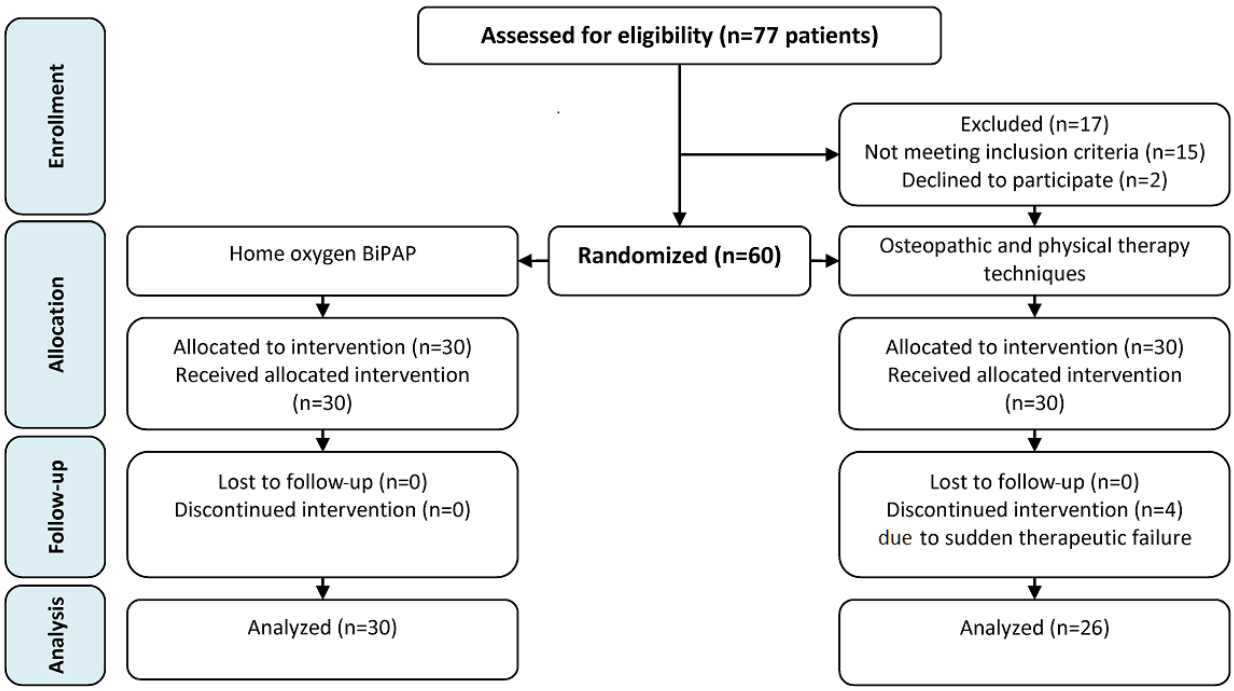
Thus, the aim of this study was to assess and compare oxygen therapy combined with noninvasive positive pressure ventilation with osteopathic manipulative respiratory and physical therapy techniques using a telemanagement health care system, which was applied to home-isolated COVID-19 patients. We also assessed patient satisfaction with the COVID-19 telemanagement system.

Methods

Overview

This single-blinded, parallel-group, randomized clinical trial was approved by the Research Ethics Committee of Cairo University and was registered at ClinicalTrials.gov (NCT04368923). The study was performed in accordance with the ethical standards of the Declaration of Helsinki and followed CONSORT guidelines for conducting randomized trials (see Figure 1).

Figure 1. CONSORT flow diagram. BiPAP: bilevel positive airway pressure.



Patient recruitment was realized through social media using the snowball subject recruitment technique. Initially, invitations were targeted to patients who met the study's eligibility criteria through the health professionals' social media networks. Then, those patients were requested to distribute invitations through their individual social media networks; no restrictions were made on what social media platform they should use. The participants were obligated to send the required data online in order to be assessed for eligibility.

A total of 60 patients were randomized into two groups in a 1:1 allocation ratio using the computer-generated randomization software StatsDirect, version 2.7.7 (StatsDirect Ltd); the allocation was done by a blinded and independent coworker to ensure that the randomization process was totally concealed. This study was conducted with a sample of home-isolated patients. Informed consent was obtained from all patients who were included in this study.

The inclusion criteria for patients were as follows:

- American Society of Anesthesiologists Class I patient before the onset of COVID-19
- History of close contact to a confirmed positive COVID-19 case as defined by the national guidelines for public health units [37]
- SARS-CoV-2 infection determined by chest computed tomography (CT) scan showing typical ground-glass abnormalities with the onset of two or more clinical symptoms and the patient condition classified as stage 1 pneumonia according to Pan et al [38]
- Patient indicated for home isolation.

Patients were excluded if they demonstrated the following characteristics:

- Not being consistent with home-isolation regimen, including nutritional supplementation, proper hygiene, and room aeration
- Inability to deal with the telemanagement system provided
- Therapeutic failure during the study, which is characterized by severe respiratory distress with respiratory rate that is more than or equal to 30 breaths per minute, oxygen saturation less than or equal to 93% in room air, and a requirement of intubation or mechanical ventilation [39].

Patients were divided equally into two groups, with 30 patients in each group. The first group received oxygen therapy with BiPAP ventilation (Group A), and the second group received osteopathic manipulative respiratory and physical therapy techniques (Group B). All patients received the same nutritional supplementation, including multivitamins and adequate supportive diets.

Regarding Group A, an oxygen concentrator with BiPAP was given using an AirFit F30 face mask (ResMed) with inspiratory positive airway pressure/expiratory positive airway pressure of 15/3 cm H_2O and 5 L/min oxygen flow by an oxygen flow meter, which was done for 4 hours per day [40]. Procedures for oxygen administration were consistent across all patients, clarified in detail, and accomplished via a teleconference for each patient that was supervised by an expert respiratory physiotherapist.

Group B received osteopathic manipulative respiratory and physical therapy techniques in the form of the following:

- Prone reverse Trendelenburg positioning for 4 hours/day [41,42]
- Cephalic traction with approximate duration of 1 minute/day [43]
- Muscle energy technique for scalene muscles 5 times/day [43]
- Rib raising technique (5 cycles for each rib group with a total of 15 cycles and approximately 5 minutes' duration) [44]
- Suboccipital area intermittent rhythmic pressure (according to each patient sensitivity) [45]
- Osteopathic lymphatic thoracic pump techniques with respiratory assist for 4 minutes [46]
- Pedal lymphatic pump for 1 minute [47]
- Thoracic inlet myofascial release for 1 minute [48]
- Diaphragmatic doming for 3 to 5 sequential respiratory cycles [49-52].

A real-time videoconference was established between the patient and the physiotherapist for training, directing, and supervising the patient during self-application. All techniques were given on a daily basis. The total duration of each therapeutic session was approximately 4 hours and 30 minutes.

In both groups, therapy ended after the assessment of all the outcome values, making sure that all of them fell within the normal range and continued to be stable for three consecutive outcomes without any clinical symptoms, which indicated recovery.

Evaluation procedures included the following:

- Recording of arterial blood gases for both oxygen (PaO_2) and carbon dioxide ($PaCO_2$) in addition to pH for each patient in both groups every 48 hours [53]
- Telemonitoring of vital signs (ie, temperature, respiratory rate, oxygen saturation, heart rate, and blood pressure) for each patient in both groups every 24 hours [54]
- Pretreatment and 14 days' posttreatment chest CT scans.

All evaluation procedures were recorded, monitored, and analyzed with an application developed by the authors, which was used for management of the patients' data and tracking of their therapeutic progress. Continuous online support with comprehensive supervision was done via videoconferencing by expert respiratory physiotherapists whenever requested by the patient.

The primary outcome measures for this study were the times to reach normal levels of both PaO_2 and $PaCO_2$ in addition to pH, which were assessed every 48 hours. Secondary outcome measures were temperature, respiratory rate, oxygen saturation, heart rate, and blood pressure, which were evaluated every 24 hours.

The participants were provided with wearable devices that have been used for telemonitoring of the required vital signs. The devices were capable of transmitting vital signs data via Bluetooth connection to the authorized gateway (ie, mobile, tablet, or any other gateway).

Collection and reporting of the primary outcome measures were done by three laboratory technicians who were blinded to the groups of study; they were then analyzed via the system. The technicians were assigned to take the samples from patients' homes for analysis. Interrater reliability was assessed using the intraclass correlation coefficient, which was 0.97; intrarater reliability was assessed using the Pearson correlation coefficient and was found to be 0.99.

Secondary outcome measures were collected and reported by the patient himself or herself, and the evaluation of these readings was done by a single clinician who was also blinded to the groups under study. All patients were instructed to be observed for another 14 days after ending therapy; chest CT scans were then performed.

Method for Telemanagement of COVID-19 Patients

A telemanagement system was developed by the authors, which followed the Health Level Seven Version 3 Standard in order to simplify its integration and facilitate receiving and/or retrieving information from other sources and applications, as well as to enable the usage of Internet of Things. This system was able to support live transmission of the vital signs data and allowed for incorporation of different medical sensors through wireless connections. The patient was able to access the system via mobile phone, tablet, or any web platform.

The major components of the platform included (1) a thorough monitoring plan, (2) the patient-specific interventions, (3) definitions of alarms and indicators, (4) the user-environment configurations of the therapist's device, and (5) the user-environment configurations of the patient's device.

Additionally, the system was able to provide an efficient, flexible, scalable integrated solution that incorporates artificial intelligence to provide support in planning, predicting, and decision making. The platform was specifically designed for managing and providing teletherapeutic services to home-isolated COVID-19 patients. It also covered the services of therapists and managing staff while coordinating the work of all the involved professionals. The system was also able to allow the patients to receive an individualized therapeutic program. It also allowed the therapists to set up a plan and

thresholds that could be personalized in accordance with each patient's profile.

In addition, it provided the option to create combined alerts via several variables. The system provided the patient with step-by-step written instructions as well as precautions about each procedure, along with a video that explained how to perform the procedures. Another advantage of this system was the auto-reminder feature to help patients with sending their data on time.

After outcome data were collected for each patient, the results were sent immediately to a server so they would be available for analysis. The system included a decision support option, which provided tailored feedback for each patient. Depending on the progress of the outcome variables, the system provided an alerting option when certain outcomes were reached or when patient performance required in-person counseling. At the end of the therapy, the system enabled each patient to answer posttreatment questions to assess patient satisfaction with the quality of this novel telemanagement service.

The decision support engine entailed workflow management routines that were used for coordinating reception of the inputs and managing their interactions, in addition to handling decision support outputs by means of tasks or actions.

The alert messages were sent to the intended devices with features that could be adjusted, such as color-coded displays, preferred choices for creating alerts, and interface personalization options. The system presented the information in meaningful medical-related ways, where it first presented the alerts followed by their related information, which was gradually presented in a more detailed manner.

With the goal of adapting the designed platform in line with the specific needs of COVID-19 patients, we conducted consecutive meetings with the patients and the therapists. The platform took into consideration the normal values of the outcome measures according to their follow-up schedule. Furthermore, we incorporated a group of educational elements for improving patients' knowledge of COVID-19, which was strengthened via interactive materials.

Analysis began after the reception of data, or was periodically scheduled, in accordance with decision support requirements. The input-module functions were dependent on the routines of feature extraction for the characterization of the data patterns.

Data sets were generated through the collection of questionnaires and clinical measurements; these measurements were assessed as being a function of time in order to detect sudden deviations, which can indicate health deterioration. The feature extraction routines for the distribution analysis and thresholding were implemented as separated modules; processed data were then retrieved and the notable clinical data patterns were fed back into a decision support engine.

For proper identification of the significant downward or upward values over the analysis window period, an identification analysis routine was used. In addition, a threshold analysis routine was used for comparing the values to the adaptive thresholds. The data in the outer ranges were defined through the confidence intervals, and any detected deviation was flagged as being an abnormal measurement. Furthermore, we designed the system to have the ability to predict the probability that a future patient would experience therapeutic failure using deep learning; however, a sufficient training data set is yet to be obtained in order for the system to provide accurate results.

Deep learning was based on feed-forward multilayer artificial neural networks that used back-propagation processes for updating the weights among the hidden layers and the output; a back-propagation of the resulting error was then applied. For the predictions, the mean square error, as well as the R^2 error, were calculated.

Statistical Analysis

The sample size was calculated with a power of 80% and level of significance of 5% (α=.05). The analysis was based on post hoc power analyses utilizing G*Power software, version 3 (Heinrich-Heine-Universität Düsseldorf) [55], accounting for 15% missing data, and based on the minimal meaningful effect size. This analysis showed that the sample size was adequate. Differences between the two groups were assessed by means and standard deviations. Comparison between the two groups was done using Student *t* test. Nominal data were summarized as frequencies and percentages. All data were statistically analyzed by an independent statistician, who was blinded to the interventions, using SPSS statistical software, version 20 (IBM Corp).

Results

A total of 60 patients (22 males [37%] and 38 females [63%]) were included in this study. Their ages ranged from 21 to 40 years with a mean age of 31.6 (SD 5.8) years. The characteristics that presented most frequently were fever (56/60, 93%) and dyspnea (55/60, 92%), as shown in Table 1.

Table 1. Demographics and baseline characteristics of patients.

Characteristic	Total (N=60)	Group A (n=30)	Group B (n=30)
Age (years), mean (SD)	31.6 (5.8)	32.2 (5.4)	30.9 (6.2)
Gender, n (%)			
Male	22 (37)	10 (33)	12 (40)
Female	38 (63)	20 (67)	18 (60)
Clinical symptom, n (%)[a]			
Fever	56 (93)	27 (90)	29 (97)
Tachypnea	51 (85)	29 (97)	22 (73)
Tachycardia	29 (48)	17 (57)	12 (40)
Hypertension	19 (32)	11 (37)	8 (27)
Dyspnea	55 (92)	27 (90)	28 (93)
Cough	43 (72)	17 (57)	26 (87)
Chest tightness	12 (20)	5 (17)	7 (23)

[a]Percentages in this category add up to greater than 100% because patients could report multiple symptoms.

A significant difference was observed between the two groups regarding the mean number of days needed for recovery (P<.05), with Group A showing a lower recovery period than Group B (mean 14.9, SD 1.7 days, and mean 23.9, SD 2.3 days, respectively), as shown in Figures 2 and 3. The unpaired t test (df=58) value was 16.55 with a mean difference of –9.056 between Group A and Group B.

Figure 2. Number of days needed for Group A and Group B to recover. Values and whiskers on bars are means and SDs, respectively. BiPAP: bilevel positive airway pressure.

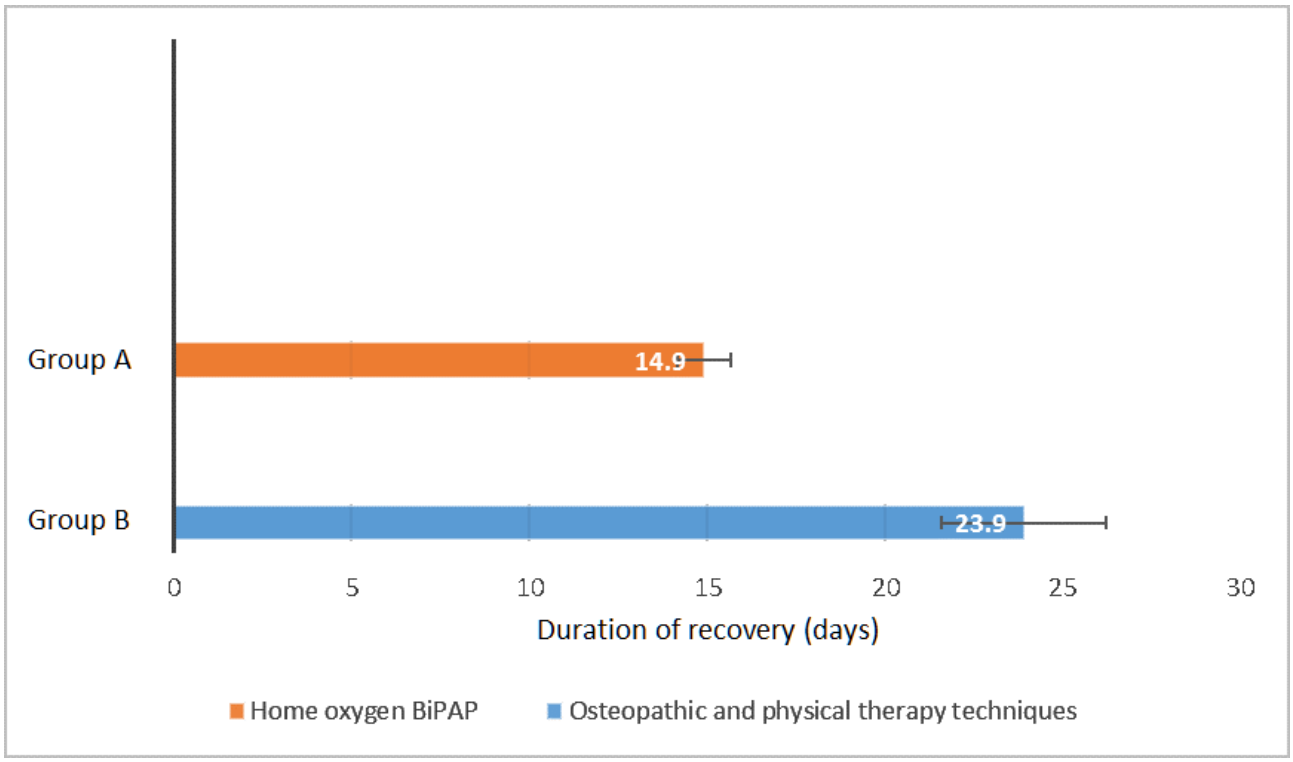
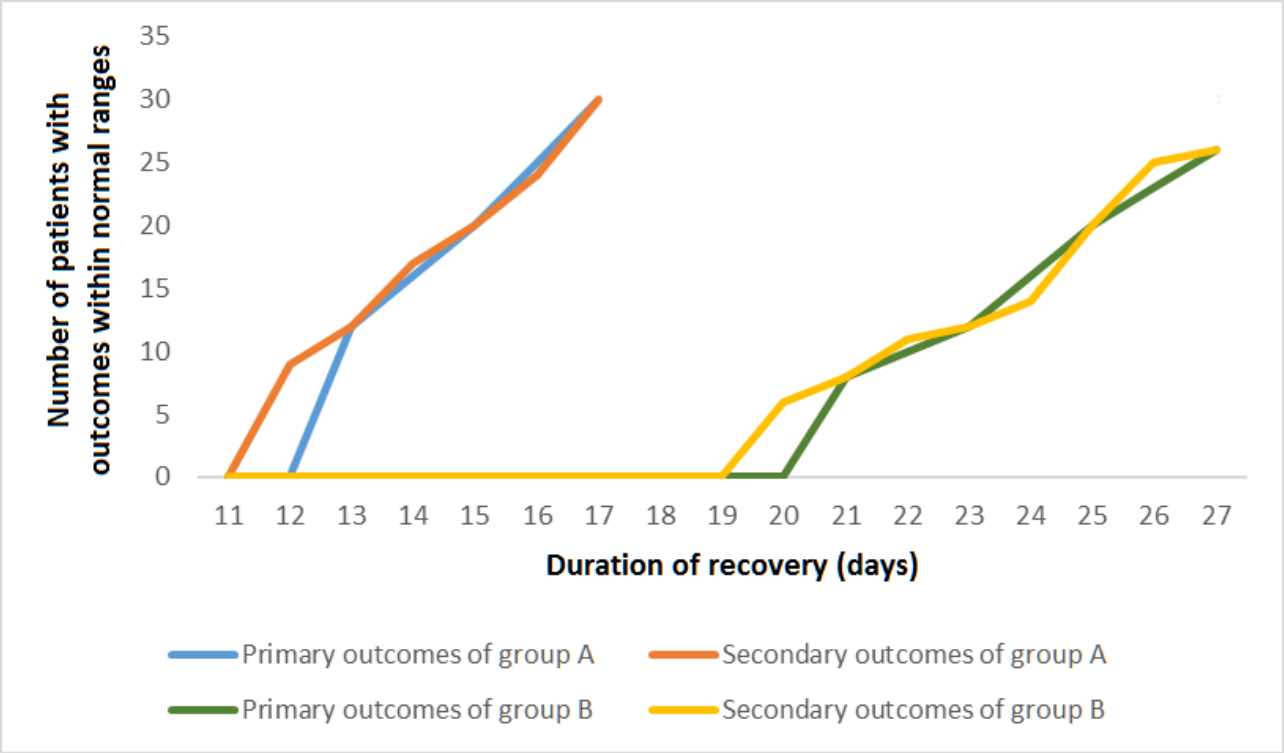


Figure 3. Patient outcome recovery rates.



A total of 4 patients out of 30 (13%) from Group B were excluded because of sudden therapeutic failure during the course of osteopathic manipulative respiratory and physical therapy techniques, as these patients required hospitalization and/or intubation.

All patients who were included in the analysis underwent chest CT scans two times, one before starting therapy and the other 14 days after ending therapy. Early-stage COVID-19 pneumonia mainly appeared as minor subpleural, bilateral, or, less commonly, unilateral ground-glass opacities in the lower lobes.

A total of 51 of the 60 patients (85%) had bilateral lung pneumonia, while 9 of the 60 patients (15%) had unilateral lung involvement. The unilateral lung involvement cases were comprised of 6 patients who had right lung involvement and 3 patients who had left lung involvement.

A total of 2 weeks after ending therapy, the CT scans showed complete resolution in Group A patients; however, in Group B patients, the lesions were mostly absorbed compared with images from before therapy in 23 out of 30 patients (77%), while 3 patients (10%) showed no worsening compared to the previous CT scan results. The 4 patients who were excluded from the study due to therapeutic failure were instructed to be admitted to the hospital.

The data sets from the study were based on the daily reports. These reports included analysis of arterial blood gases and vital signs. The data set consisted of 1735 daily records. After filtering the data set by removing the records of the excluded patients, the final analyzed data set included 1686 records. The parameters of the submitted records for each attribute are shown in Table 2.

Table 2. Parameters of the submitted records.

Attribute	Duration	Total submitted records for Group A, n	Total submitted records for Group B, n
PaO_2	Every 48 hours	238	324
$PaCO_2$	Every 48 hours	238	324
pH	Every 48 hours	238	324
Temperature	Every 24 hours	476	648
Respiratory rate	Every 24 hours	476	648
Oxygen saturation	Every 24 hours	476	648
Heart rate	Every 24 hours	476	648
Blood pressure	Every 24 hours	476	648

Our system predictions showed an R^2 of 0.965 with a mean square error of 0.27, which means that more than 96% of the variations were predicted.

The posttreatment patient satisfaction questions focused on a group of technical aspects—simplicity, effectiveness, acceptability, usability, reliability, and level of confidence—as well as the system's ability to assess a patient's status remotely. The results of the satisfaction questions were very promising and most of the patients responded positively to the questions, as shown in Table 3.

Table 3. Posttreatment patient satisfaction questions for the telemanagement system.

Question	Response (n=56), n (%)[a]		
	Yes	No	No answer
Were the telemanagement procedures simple?	41 (73)	12 (21)	3 (5)
Were the telemanagement treatment procedures useful?	54 (96)	0 (0)	2 (4)
Were the telemanagement procedures well tolerated?	56 (100)	0 (0)	0 (0)
Were educational elements and interactive materials useful?	56 (100)	0 (0)	0 (0)
Did you regret using this telemanagement system and, instead, prefer admission to the hospital?	1 (2)	52 (93)	3 (5)
Did the auto-reminder option in this system help you in sending data on time?	55 (98)	0 (0)	1 (2)
Do you think that the time spent by therapists with you was adequate?	38 (68)	13 (23)	5 (9)
Do you think that this telemanagement system was consistent?	49 (88)	2 (4)	5 (9)
Do you think that this telemanagement system is an acceptable way to receive treatment services?	56 (100)	0 (0)	0 (0)
Overall, are you satisfied with the quality of the services provided by this telemanagement system?	55 (98)	1 (2)	0 (0)
Would you recommend this telemanagement system to anyone?	51 (91)	1 (2)	4 (7)
Any comments? (Comments added to a free-text field are shown below) • I regret not being admitted to the hospital as I think that I would not have transmitted the infection to my wife. • This telemanagement system saved efforts, costs, and time. • I think that the time spent with the therapist needs to be more frequent. • The system could be more helpful if it would send periodic information often about my status by informing me if I am getting better or worse.	4 (7)	48 (86)	4 (7)

[a]Percentages may not add up to 100% due to rounding.

The telemanagement system showed positive satisfaction responses from the patients for most categories, including simplicity (41/56, 73%), effectiveness (101/112, 90.2%), acceptability (56/56, 100%), usability (110/112, 98.2%), reliability (162/168, 96.4%), level of confidence (55/56, 98%), and ability for remote assessment (38/56, 68%).

Discussion

Principal Findings

Our study explores the feasibility of oxygen therapy with noninvasive positive pressure ventilation therapy versus osteopathic manipulative respiratory and physical therapy techniques in COVID-19 patients.

Only 56 of the 60 patients (93%) completed the treatment until recovery and achieved all of the required outcomes. The therapeutic procedures were well tolerated, and the clinical symptoms significantly improved over a relatively short time.

The telemanagement application addressed variance in the clinical outcomes and maximized the benefits from the specialists' expertise. Telemanagement and telemedicine applications have a wide range of implementations, and there has been robust evidence about their clinical benefits, cost-effectiveness, simplicity, and positive impact on critical care safety and quality [56].

The telemanagement decision support option provided many potential benefits by reducing the specialist's workload caused by regular revision of all the results; this option allowed patients' measurement patterns to indicate health status deteriorations and the identification of patients with higher priorities for revision.

Regardless of the broad usage of smart devices in health telemanagement and the collection of data, medical applications of deep learning approaches for making predictions are still considered to be challenging [57]. In this study, depending on the prevailing health conditions, deep learning was used to assist in defining the essential features, predictions, and contextual detections of patterns.

Deep learning was also used as a recognized candidate to predict the probability that a future patient would experience therapeutic failure; this was due to its ability to exploit the intramodality correlations efficiently that would allow for extraction of the hierarchical representations of the data, as well as its ability to perform feature extractions.

The results obtained from this study clearly demonstrated that noninvasive BiPAP was able to significantly improve the clinical

status of patients infected with SARS-CoV-2. Patients' PaO_2 and oxygen saturation were elevated over a relatively short time. Patients' respiratory frequency and heart rate decreased. Furthermore, 100% (30/30) of Group A patients did not require any hospitalization or intubation and did not experience any complications.

The results of this study suggest that the application of home-based oxygen BiPAP therapy could improve the respiratory status of patients with COVID-19 pneumonia. It could also significantly improve the arterial blood gas status of the patients without any negative influence on hemodynamics.

It was also found that the application of oxygen with BiPAP at an early stage could reduce the need for intubation along with its related complications. The respiratory complications of COVID-19 can be attributed to a rise in the capillary alveolar membrane permeability, which can lead to pulmonary edema. Thus, noninvasive BiPAP may have contributed to prevention of deterioration by impeding the consequent pulmonary edema via positive pressure [19].

Nevertheless, some side effects of BiPAP were noticed in this study, such as facial skin and eye irritation, mild oropharyngeal dryness, mild abdominal gaseous distention, and stomach pain. Using appropriate face masks with good compatibility, along with avoiding mouth respiration and guiding nasal respiration, have been found to be effective in decreasing these side effects.

Regarding the group treated with osteopathic manipulative respiratory and physical therapy techniques, the improvement in chest CT scans was not significantly different from baseline but seemed to be clinically relevant, while there was a significant improvement in the rest of the outcomes. The clinical symptoms had also improved over a relatively short time. The osteopathic manipulative respiratory techniques that were used with this group have been reported in several studies to have beneficial effects in treating pneumonia [58]. Our main motives for the use of these techniques were focused on immunity improvement, blood clotting prevention, as well as the absence of any noticeable side effect.

Although NIV was found to be more effective, combining the reverse Trendelenburg with prone positioning can be beneficial as well. In our study, it was reported to be tolerable and comfortable by the patients. Some studies investigated the effects of each position [41,42]; one of those studies demonstrated that prone positioning contributed to improvement of the ventilation perfusion mismatch in COVID-19 patients by inducing dorsal lung region recruitment, alveolar shunt reduction, tidal volume, and end-expiratory lung volume improvement [41]. However, we found that prone positioning can induce abdominal push on the diaphragm, especially with patients that have a protruded abdomen, which can limit diaphragmatic excursion. In another study, reverse Trendelenburg positioning by tilting the patient's head up by 25 degrees showed a decrease in abdominal push on the diaphragm; therefore, an increase in functional residual capacity and lung compliance were observed [42]. Thus, we combined the two positions to avoid any limitations in diaphragmatic movement.

While the group treated with physical therapy showed less-significant results, most of the patients in our study reported relaxation effects immediately after application of cephalic traction and muscle energy techniques for scalene muscles. Similarly, another study demonstrated that these techniques had an effective role in improving vital capacity, increasing respiratory muscle efficiency, increasing cervical flexibility, and decreasing fatigability levels [43]. This study also revealed that the more the lengths of the scalene muscles changed per unit volume, the lower the alveolar pressure became. Thus, the ventilation volume through thoracic expansion increases when the scalene muscles maintain reasonable lengths. Likewise, the rib-raising technique was reported to be relaxing to the patients, which is consistent with another study that found this technique to have an immediate reduction in the activity of the sympathetic nervous system without causing any alteration in parasympathetic activity or in the hypothalamic-pituitary-adrenal axis [44].

The patients' feelings of arousal that were reported after having intermittent pressure to the suboccipital area can be attributed to the ability of this technique to cause arterial vasomotion at rates usually associated with the cranial rhythmic impulse, as demonstrated in another study [45].

In our study, osteopathic lymphatic thoracic pump techniques with respiratory assist, myofascial release to the thoracic inlet, and the pedal lymphatic pump technique also indicated good tolerability; they were also reported in another study to be advantageous for treating pneumonia by targeting the lymphatic flow, activating autonomic-mediated intrinsic lymphatic contractility, improving respiratory function, and improving circulation. In addition, the thoracic lymphatic pump techniques and thoracic inlet release have been shown to increase chemokines and cytokines in the thoracic vessels as well as in the intestinal lymph vessels, while the pedal lymphatic pump technique improves flow into the lymphatic systems [59]. The outcomes of the lymphatic pump techniques, including improvement in serum interferon levels [46], were carefully directed toward our study sample as we have taken into consideration that COVID-19 induces hyperactivation of the immune system in the severe stages, while it usually induces impairment of the immune system in the early stages. Therefore, immune response suppression may be targeted in the severe stages, while in the early stages, which was the case in our study, reduction of the viral load by stimulating type I interferon should be targeted [60].

It was also observed that in the first group, NIV promoted a significant increase in chest wall volumes directly after application, which could be due to passive expansion. On the other hand, manual diaphragmatic releasing techniques in the second group contributed to positive outcomes of our study by improving the mobility of the chest wall immediately after the intervention, which was in accordance with several studies [49-52].

Compared to physical therapy techniques, BiPAP therapy outcomes were confirmed by this study to be much more effective and promising. Even though physical therapy

techniques do not require equipment, those techniques need to be investigated further in order to be considered promising.

Overall, regarding posttreatment satisfaction with our proposed telemanagement health care systems, positive responses were given by most of the patients, even in the group treated with physical therapy, despite longer recovery periods. This was attributed to the advantage of the second group's costs being lower than those of the first group.

Limitations

Our study had some limitations. The study included a relatively small number of patients. Thus, further large randomized controlled trials with larger sample sizes are recommended. In addition, osteopathic manipulative respiratory and physical therapy techniques may have a role in elevating recovery rates and improving outcomes of patients with COVID-19. Therefore, further randomized controlled trials with larger sample sizes would be required to determine the therapeutic extent of these techniques.

Conclusions

From this study, it was found that in the early stages of SARS-CoV-2 pneumonia, home-based oxygen BiPAP ventilation can reduce the need for endotracheal intubation. It can also be an effective prophylactic treatment approach to avoid exacerbation of this disease and the need for hospitalization. Home-based oxygen BiPAP ventilation was more effective than osteopathic manipulative respiratory and physical therapy techniques, as it was associated with shorter recovery periods. A home-based COVID-19 telemanagement system with decision support services showed satisfying outcomes and may be recommended in certain cases as an effective solution for the extreme shortage of hospital beds caused by this pandemic. Further investigations are still required to determine the effectiveness of the osteopathic manipulative respiratory and physical therapy techniques in the management of COVID-19 patients in the early stages.

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT-EHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1428 KB - jmir_v23i4e23446_app1.pdf]

References

1. Eysenbach G. How to fight an infodemic: The four pillars of infodemic management. J Med Internet Res 2020 Jun 29;22(6):e21820 [FREE Full text] [doi: 10.2196/21820] [Medline: 32589589]
2. Pahayahay A, Khalili-Mahani N. What media helps, what media hurts: A mixed methods survey study of coping with COVID-19 using the media repertoire framework and the appraisal theory of stress. J Med Internet Res 2020 Aug 06;22(8):e20186 [FREE Full text] [doi: 10.2196/20186] [Medline: 32701459]
3. Hou Z, Du F, Zhou X, Jiang H, Martin S, Larson H, et al. Cross-country comparison of public awareness, rumors, and behavioral responses to the COVID-19 epidemic: Infodemiology study. J Med Internet Res 2020 Aug 03;22(8):e21143 [FREE Full text] [doi: 10.2196/21143] [Medline: 32701460]
4. Ding L, She Q, Chen F, Chen Z, Jiang M, Huang H, et al. The internet hospital plus drug delivery platform for health management during the COVID-19 pandemic: Observational study. J Med Internet Res 2020 Aug 06;22(8):e19678 [FREE Full text] [doi: 10.2196/19678] [Medline: 32716892]
5. Pépin JL, Bruno RM, Yang R, Vercamer V, Jouhaud P, Escourrou P, et al. Wearable activity trackers for monitoring adherence to home confinement during the COVID-19 pandemic worldwide: Data aggregation and analysis. J Med Internet Res 2020 Jun 19;22(6):e19787 [FREE Full text] [doi: 10.2196/19787] [Medline: 32501803]
6. Farooq A, Laato S, Islam AKMN. Impact of online information on self-isolation intention during the COVID-19 pandemic: Cross-sectional study. J Med Internet Res 2020 May 06;22(5):e19128 [FREE Full text] [doi: 10.2196/19128] [Medline: 32330115]
7. Kukafka R. Digital health consumers on the road to the future. J Med Internet Res 2019 Nov 21;21(11):e16359 [FREE Full text] [doi: 10.2196/16359] [Medline: 31750835]
8. Xu C, Zhang X, Wang Y. Mapping of health literacy and social panic via web search data during the COVID-19 public health emergency: Infodemiological study. J Med Internet Res 2020 Jul 02;22(7):e18831 [FREE Full text] [doi: 10.2196/18831] [Medline: 32540844]
9. Crawford A, Serhal E. Digital health equity and COVID-19: The innovation curve cannot reinforce the social gradient of health. J Med Internet Res 2020 Jun 02;22(6):e19361 [FREE Full text] [doi: 10.2196/19361] [Medline: 32452816]
10. Adly AS, Adly AS, Adly MS. Approaches based on artificial intelligence and the Internet of Intelligent Things to prevent the spread of COVID-19: Scoping review. J Med Internet Res 2020 Aug 10;22(8):e19104 [FREE Full text] [doi: 10.2196/19104] [Medline: 32584780]

11. Geldsetzer P. Use of rapid online surveys to assess people's perceptions during infectious disease outbreaks: A cross-sectional survey on COVID-19. J Med Internet Res 2020 Apr 02;22(4):e18790 [FREE Full text] [doi: 10.2196/18790] [Medline: 32240094]
12. Kaxiras E, Neofotistos G. Multiple epidemic wave model of the COVID-19 pandemic: Modeling study. J Med Internet Res 2020 Jul 30;22(7):e20912 [FREE Full text] [doi: 10.2196/20912] [Medline: 32692690]
13. Home Care for Patients With Suspected or Confirmed COVID-19 and Management of Their Contacts: Interim Guidance. Geneva, Switzerland: World Health Organization; 2020 Aug 13. URL: https://www.who.int/publications/i/item/home-care-for-patients-with-suspected-novel-coronavirus-(ncov)-infection-presenting-with-mild-symptoms-and-management-of-contacts [accessed 2021-04-20]
14. Fernández-Quintela A, Milton-Laskibar I, Trepiana J, Gómez-Zorita S, Kajarabille N, Léniz A, et al. Key aspects in nutritional management of COVID-19 patients. J Clin Med 2020 Aug 10;9(8):2589 [FREE Full text] [doi: 10.3390/jcm9082589] [Medline: 32785121]
15. Ballenberger M, Ishikawa O, Mina B, Esquinas AM. Noninvasive ventilation in hypoxemic respiratory failure. In: Esquinas AM, Fiorentino G, Insalaco G, Mina B, Duan J, Mondardini MC, et al, editors. Noninvasive Ventilation in Sleep Medicine and Pulmonary Critical Care. Cham, Switzerland: Springer; May 29, 2020:325-336.
16. Schünemann HJ, Khabsa J, Solo K, Khamis AM, Brignardello-Petersen R, El-Harakeh A, et al. Ventilation techniques and risk for transmission of coronavirus disease, including COVID-19: A living systematic review of multiple streams of evidence. Ann Intern Med 2020 Aug 04;173(3):204-216 [FREE Full text] [doi: 10.7326/M20-2306] [Medline: 32442035]
17. Arulkumaran N, Brealey D, Howell D, Singer M. Use of non-invasive ventilation for patients with COVID-19: A cause for concern? Lancet Respir Med 2020 Jun;8(6):e45. [doi: 10.1016/s2213-2600(20)30181-8]
18. Verma K. Pulmonary edema in COVID19-A neural hypothesis. ACS Chem Neurosci 2020 Jul 15;11(14):2048-2050 [FREE Full text] [doi: 10.1021/acschemneuro.0c00370] [Medline: 32614178]
19. Mastroianni F, Narasimhan M. Measures of improvement for noninvasive positive pressure ventilation in the ICU. In: Esquinas AM, Fiorentino G, Insalaco G, Mina B, Duan J, Mondardini MC, et al, editors. Noninvasive Ventilation in Sleep Medicine and Pulmonary Critical Care. Cham, Switzerland: Springer; May 29, 2020:301-305.
20. Thomas P, Baldwin C, Bissett B, Boden I, Gosselink R, Granger CL, et al. Physiotherapy management for COVID-19 in the acute hospital setting: Clinical practice recommendations. J Physiother 2020 Apr;66(2):73-82 [FREE Full text] [doi: 10.1016/j.jphys.2020.03.011] [Medline: 32312646]
21. Lorenzo S, Nicotra CM, Mentreddy AR, Padia HJ, Stewart DO, Hussein MO, et al. Assessment of pulmonary function after osteopathic manipulative treatment vs standard pulmonary rehabilitation in a healthy population. J Am Osteopath Assoc 2019 Feb 11;119(3):155-163. [doi: 10.7556/jaoa.2019.026] [Medline: 30741314]
22. Seffinger MA, editor. Foundations of Osteopathic Medicine: Philosophy, Science, Clinical Applications, and Research. 4th edition. Philadelphia, PA: Wolters Kluwer; 2018.
23. Lee AC. COVID-19 and the advancement of digital physical therapist practice and telehealth. Phys Ther 2020 Jul 19;100(7):1054-1057 [FREE Full text] [doi: 10.1093/ptj/pzaa079] [Medline: 32343836]
24. Zachwieja E, Theosmy EG, Yacovelli SJ, Beatty EW, McGrath ME, Lonner JH. Web-based self-directed exercise program is cost-effective compared to formal physical therapy after primary total knee arthroplasty. J Arthroplasty 2020 Sep;35(9):2335-2341. [doi: 10.1016/j.arth.2020.04.061] [Medline: 32423757]
25. Turolla A, Rossettini G, Viceconti A, Palese A, Geri T. Musculoskeletal physical therapy during the COVID-19 pandemic: Is telerehabilitation the answer? Phys Ther 2020 Aug 12;100(8):1260-1264 [FREE Full text] [doi: 10.1093/ptj/pzaa093] [Medline: 32386218]
26. Lewis J, Mc Auliffe S, O'Sullivan K, O'Sullivan P, Whiteley R. Musculoskeletal physical therapy after COVID-19: Time for a new "normal". J Orthop Sports Phys Ther 2021 Jan;51(1):5-7. [doi: 10.2519/jospt.2021.0102] [Medline: 33383997]
27. Borghi-Silva A, Goncalves Mendes R, Toledo AC, Malosa Sampaio LM, Patrocínio da Silva T, Kunikushita LN, et al. Adjuncts to physical training of patients with severe COPD: Oxygen or noninvasive ventilation? Respir Care 2010 Jul;55(7):885-894 [FREE Full text]
28. MacIntyre NR. Physiologic effects of noninvasive ventilation. Respir Care 2019 Jun;64(6):617-628 [FREE Full text] [doi: 10.4187/respcare.06635] [Medline: 31110031]
29. Shi Z, Mehrotra A, Gidengil CA, Poon SJ, Uscher-Pines L, Ray KN. Quality of care for acute respiratory infections during direct-to-consumer telemedicine visits for adults. Health Aff (Millwood) 2018 Dec;37(12):2014-2023 [FREE Full text] [doi: 10.1377/hlthaff.2018.05091] [Medline: 30633682]
30. Fagherazzi G, Goetzinger C, Rashid MA, Aguayo GA, Huiart L. Digital health strategies to fight COVID-19 worldwide: Challenges, recommendations, and a call for papers. J Med Internet Res 2020 Jun 16;22(6):e19284 [FREE Full text] [doi: 10.2196/19284] [Medline: 32501804]
31. Portz JD, Brungardt A, Shanbhag P, Staton EW, Bose-Brill S, Lin C, et al. Advance care planning among users of a patient portal during the COVID-19 pandemic: Retrospective observational study. J Med Internet Res 2020 Aug 11;22(8):e21385 [FREE Full text] [doi: 10.2196/21385] [Medline: 32716900]
32. Katapally TR. A global digital citizen science policy to tackle pandemics like COVID-19. J Med Internet Res 2020 May 26;22(5):e19357 [FREE Full text] [doi: 10.2196/19357] [Medline: 32408267]

33. Parush A, Wacht O, Gomes R, Frenkel A. Human factor considerations in using personal protective equipment in the COVID-19 pandemic context: Binational survey study. J Med Internet Res 2020 Jun 17;22(6):e19947 [FREE Full text] [doi: 10.2196/19947] [Medline: 32511099]
34. Waqas A, Teoh SH, Lapão LV, Messina LA, Correia JC. Harnessing telemedicine for the provision of health care: Bibliometric and scientometric analysis. J Med Internet Res 2020 Oct 02;22(10):e18835 [FREE Full text] [doi: 10.2196/18835] [Medline: 33006571]
35. Almathami HKY, Win KT, Vlahu-Gjorgievska E. Barriers and facilitators that influence telemedicine-based, real-time, online consultation at patients' homes: Systematic literature review. J Med Internet Res 2020 Feb 20;22(2):e16407 [FREE Full text] [doi: 10.2196/16407] [Medline: 32130131]
36. Adly MS, Adly AS, Adly AS. Assessment of early orthodontic treatment on functional shifts by telemonitoring mandibular movements using a smart phone. J Telemed Telecare 2020 Apr;26(3):150-160. [doi: 10.1177/1357633X18800876] [Medline: 30373447]
37. Communicable Diseases Network Australia (CDNA). Coronavirus Disease 2019 (COVID-19): CDNA National Guidelines for Public Health Units. Canberra, Australia: Australian Government, Department of Health; 2020 Mar 24. URL: https://www1.health.gov.au/internet/main/publishing.nsf/Content/cdna-song-novel-coronavirus.htm [accessed 2021-04-20]
38. Pan F, Ye T, Sun P, Gui S, Liang B, Li L, et al. Time course of lung changes at chest CT during recovery from coronavirus disease 2019 (COVID-19). Radiology 2020 Jun;295(3):715-721. [doi: 10.1148/radiol.2020200370]
39. Grifoni E, Valoriani A, Cei F, Vannucchi V, Moroni F, Pelagatti L, et al. The CALL score for predicting outcomes in patients with COVID-19. Clin Infect Dis 2021 Jan 23;72(1):182-183 [FREE Full text] [doi: 10.1093/cid/ciaa686] [Medline: 32474605]
40. Gibson V, Waters D, editors. Respiratory Care. Boca Raton, FL: CRC Press; 2017.
41. Ghelichkhani P, Esmaeili M. Prone position in management of COVID-19 patients: A commentary. Arch Acad Emerg Med 2020;8(1):e48 [FREE Full text] [Medline: 32309812]
42. Stewart LE, Long MJ. Positioning the patient for procedures. In: Gross KA, editor. Advanced Practice and Leadership in Radiology Nursing. Cham, Switzerland: Springer; 2020:91-103.
43. Lee J, Hwang S, Han S, Han D. Effects of stretching the scalene muscles on slow vital capacity. J Phys Ther Sci 2016 Jun;28(6):1825-1828 [FREE Full text] [doi: 10.1589/jpts.28.1825] [Medline: 27390425]
44. Rechberger V, Biberschick M, Porthun J. Effectiveness of an osteopathic treatment on the autonomic nervous system: A systematic review of the literature. Eur J Med Res 2019 Oct 25;24(1):36 [FREE Full text] [doi: 10.1186/s40001-019-0394-5] [Medline: 31653268]
45. Guillaud A, Darbois N, Monvoisin R, Pinsault N. Reliability of diagnosis and clinical efficacy of cranial osteopathy: A systematic review. PLoS One 2016;11(12):e0167823 [FREE Full text] [doi: 10.1371/journal.pone.0167823] [Medline: 27936211]
46. Hodge LM. Osteopathic lymphatic pump techniques to enhance immunity and treat pneumonia. Int J Osteopath Med 2012 Mar;15(1):13-21 [FREE Full text] [doi: 10.1016/j.ijosm.2011.11.004] [Medline: 22977459]
47. Franzini D, Cuny LA, Pierce-Talsma S. Osteopathic lymphatic pump techniques. J Am Osteopath Assoc 2018 Jul 01;118(7):e43-e44. [doi: 10.7556/jaoa.2018.112] [Medline: 29946678]
48. Bordoni B. The five diaphragms in osteopathic manipulative medicine: Myofascial relationships, Part 1. Cureus 2020 Apr 23;12(4):e7794 [FREE Full text] [doi: 10.7759/cureus.7794] [Medline: 32461863]
49. Marizeiro DF, Florêncio ACL, Nunes ACL, Campos NG, Lima PODP. Immediate effects of diaphragmatic myofascial release on the physical and functional outcomes in sedentary women: A randomized placebo-controlled trial. J Bodyw Mov Ther 2018 Oct;22(4):924-929. [doi: 10.1016/j.jbmt.2017.10.008] [Medline: 30368336]
50. Nair A, Alaparthi GK, Krishnan S, Rai S, Anand R, Acharya V, et al. Comparison of diaphragmatic stretch technique and manual diaphragm release technique on diaphragmatic excursion in chronic obstructive pulmonary disease: A randomized crossover trial. Pulm Med 2019;2019:6364376. [doi: 10.1155/2019/6364376] [Medline: 30719351]
51. Elnaggar RK, Shendy MA, Mahmoud MZ. Prospective effects of manual diaphragmatic release and thoracic lymphatic pumping in childhood asthma. Respir Care 2019 Nov;64(11):1422-1432 [FREE Full text] [doi: 10.4187/respcare.06716] [Medline: 31337743]
52. Leonés-Macías E, Torres-Sánchez I, Cabrera-Martos I, Ortiz-Rubio A, López-López L, Valenza MC. Effects of manual therapy on the diaphragm in asthmatic patients: A randomized pilot study. Int J Osteopath Med 2018 Sep;29:26-31. [doi: 10.1016/j.ijosm.2018.07.006]
53. Grange ES, Neil EJ, Stoffel M, Singh AP, Tseng E, Resco-Summers K, et al. Responding to COVID-19: The UW Medicine Information Technology Services experience. Appl Clin Inform 2020 Mar;11(2):265-275 [FREE Full text] [doi: 10.1055/s-0040-1709715] [Medline: 32268390]
54. Assessing COVID-19-Related Symptoms in Outpatient Adult and Adolescent Subjects in Clinical Trials of Drugs and Biological Products for COVID-19 Prevention or Treatment: Guidance for Industry. Silver Spring, MD: US Department of Health and Human Services, Food and Drug Administration; 2020 Sep. URL: https://www.fda.gov/media/142143/download [accessed 2021-04-21]

55. Erdfelder E, Faul F, Buchner A. GPOWER: A general power analysis program. Behav Res Methods Instrum Comput 1996 Mar;28(1):1-11. [doi: 10.3758/bf03203630]
56. Atmojo JT, Sudaryanto WT, Widiyanto A, Ernawati E, Arradini D. Telemedicine, cost effectiveness, and patients satisfaction: A systematic review. J Health Policy Manage 2020;5(2):103-107 [FREE Full text] [doi: 10.26911/thejhpm.2020.05.02.02]
57. Majumder S, Mondal T, Deen M. Wearable sensors for remote health monitoring. Sensors (Basel) 2017 Jan 12;17(1):130 [FREE Full text] [doi: 10.3390/s17010130] [Medline: 28085085]
58. House S, Hodge L. The effect of lymph isolated during osteopathic lymphatic pump treatment on the immune response against acute pneumonia. In: Proceedings of Research Appreciation Day (RAD). 2019 Mar 05 Presented at: Research Appreciation Day (RAD); March 29, 2019; Fort Worth, TX.
59. Walkowski S, Singh M, Puertas J, Pate M, Goodrum K, Benencia F. Osteopathic manipulative therapy induces early plasma cytokine release and mobilization of a population of blood dendritic cells. PLoS One 2014;9(3):e90132 [FREE Full text] [doi: 10.1371/journal.pone.0090132] [Medline: 24614605]
60. Jamilloux Y, Henry T, Belot A, Viel S, Fauter M, El Jammal T, et al. Should we stimulate or suppress immune responses in COVID-19? Cytokine and anti-cytokine interventions. Autoimmun Rev 2020 Jul;19(7):102567 [FREE Full text] [doi: 10.1016/j.autrev.2020.102567] [Medline: 32376392]

Abbreviations

BiPAP: bilevel positive airway pressure
CT: computed tomography
ICT: information and communication technology
NIV: noninvasive ventilation

Edited by R Kukafka, C Basch; submitted 12.08.20; peer-reviewed by D Lipardo, N Freene, S Lalmuanawma; comments to author 14.12.20; revised version received 25.12.20; accepted 01.04.21; published 28.04.21.

Please cite as:
Adly AS, Adly MS, Adly AS
Telemanagement of Home-Isolated COVID-19 Patients Using Oxygen Therapy With Noninvasive Positive Pressure Ventilation and Physical Therapy Techniques: Randomized Clinical Trial
J Med Internet Res 2021;23(4):e23446
URL: https://www.jmir.org/2021/4/e23446
doi:10.2196/23446
PMID:33819166

Original Paper

Impact of the COVID-19 Pandemic on Health Care Utilization in a Large Integrated Health Care System: Retrospective Cohort Study

Stanley Xu[1], PhD; Sungching Glenn[1], MSc; Lina Sy[1], MPH; Lei Qian[1], PhD; Vennis Hong[1], MPH; Denison S Ryan[1], MPH; Steven Jacobsen[1], MD, PhD

Department of Research & Evaluation, Kaiser Permanente Southern California, Pasadena, CA, United States

Corresponding Author:
Stanley Xu, PhD
Department of Research & Evaluation
Kaiser Permanente Southern California
100 S Los Robles Ave, 5th Floor
Pasadena, CA, 91101
United States
Phone: 1 6263726807
Email: stan.xu@kp.org

Related Article:

This is a corrected version. See correction statement: https://www.jmir.org/2021/5/e30101/

Abstract

Background: The COVID-19 pandemic has caused an abrupt reduction in the use of in-person health care, accompanied by a corresponding surge in the use of telehealth services. However, the extent and nature of changes in health care utilization during the pandemic may differ by care setting. Knowledge of the impact of the pandemic on health care utilization is important to health care organizations and policy makers.

Objective: The aims of this study are (1) to evaluate changes in in-person health care utilization and telehealth visits during the COVID-19 pandemic and (2) to assess the difference in changes in health care utilization between the pandemic year 2020 and the prepandemic year 2019.

Methods: We retrospectively assembled a cohort consisting of members of a large integrated health care organization, who were enrolled between January 6 and November 2, 2019 (prepandemic year), and between January 5 and October 31, 2020 (pandemic year). The rates of visits were calculated weekly for four settings: inpatient, emergency department (ED), outpatient, and telehealth. Using Poisson models, we assessed the impact of the pandemic on health care utilization during the early days of the pandemic and conducted difference-in-deference (DID) analyses to measure the changes in health care utilization, adjusting for the trend of health care utilization in the prepandemic year.

Results: In the early days of the pandemic, we observed significant reductions in inpatient, ED, and outpatient utilization (by 30.2%, 37.0%, and 80.9%, respectively). By contrast, there was a 4-fold increase in telehealth visits between weeks 8 (February 23) and 12 (March 22) in 2020. DID analyses revealed that after adjusting for prepandemic secular trends, the reductions in inpatient, ED, and outpatient visit rates in the early days of the pandemic were 1.6, 8.9, and 367.2 visits per 100 person-years (P<.001), respectively, while the increase in telehealth visits was 272.9 visits per 100 person-years (P<.001). Further analyses suggested that the increase in telehealth visits offset the reduction in outpatient visits by week 26 (June 28, 2020).

Conclusions: In-person health care utilization decreased drastically during the early period of the pandemic, but there was a corresponding increase in telehealth visits during the same period. By end-June 2020, the combined outpatient and telehealth visits had recovered to prepandemic levels.

(J Med Internet Res 2021;23(4):e26558) doi:10.2196/26558

KEYWORDS

cohort; COVID-19; difference-in-difference analysis; health care utilization; health care worker; impact; knowledge; pandemic; policy; retrospective; telehealth; telemedicine; usage; utilization

Introduction

The COVID-19 pandemic has caused an abrupt reduction in the use of in-person health care, which has been accompanied by a corresponding surge in the use of telehealth services [1,2]. Health care visits such as inpatient visits, emergency department (ED) visits, and outpatient visits have significantly decreased since the start of the pandemic [3-6]. Two major factors have contributed to these changes. First, patients have chosen not to seek in-person health care owing to the fear of exposure to SARS-CoV-2 [3,7-9]. Second, in the early days of the pandemic, the Centers for Disease Control and Prevention (CDC) and Centers for Medicare and Medicaid Services (CMS) recommended delaying elective care to reduce the risk of SARS-CoV-2 transmission in health care facilities and to reduce the burden on health care systems [10]. Specifically, on March 4, 2020, the governor of California declared a state of emergency after the first official COVID-19 death in the state. On March 19, 2020, a stay-at-home order was enacted in California to slow the spread of SARS-CoV-2.

The CDC also encouraged the use of telehealth services to deliver care [11]. Telehealth is a health care provider's technology of choice to communicate information regarding the delivery of clinical and nonclinical care services. In addition to providing care for some medical conditions, telehealth has helped protect both providers and patients from the risk of exposure to SARS-CoV-2. It has also helped preserve critical personal protective equipment that was in short supply in the early days of the pandemic.

In response to this, Kaiser Permanente Southern California (KPSC) reported a drastic decline in in-person health care visits, coupled with an immediate increase in telehealth visits. The objectives of this study are to (1) evaluate changes in in-person health care utilization and telehealth visits at one of the largest integrated health care systems in the United States during the COVID-19 pandemic year 2020 and (2) assess the difference in changes in health care utilization between the pandemic year 2020 and the prepandemic year 2019.

Methods

Study Population and Study Period

We retrospectively assembled a cohort consisting of members from a large integrated health care system, KPSC. The KPSC serves 4.7 million members at 15 medical centers with at least 50% of its members belonging to racial or ethnic minorities, and 55% living in neighborhoods with a median annual household income of ≤US $75,000 [12]. The study period included the first 43 weeks in the pandemic year (January 5 to October 31, 2020) and the first 43 weeks in the prepandemic year 2019 (January 6 to November 2, 2019). In all analyses, health care utilization was considered for all members of the KPSC enrolled in this study during a given week. Because of data lags in inpatient and ED visits revealed from claims, we only included inpatient and ED visits in the first 35 weeks in the following analyses.

Data Source and Identification of Visits

We used electronic health record (EHR) data and claims data to identify visits in four settings: inpatient, ED, outpatient, and telehealth. Most of the encounters (approximately 90%) were from EHR data. While EHR data clearly indicated the encounter setting, for claims data, we used place-of-service and hospital revenue codes to determine the encounter setting. Multiple claims were consolidated to resemble a similar visit in the EHR. For example, a consolidated inpatient visit from claims data could include both institutional and professional claims. When a patient was admitted to the ED and then transferred to the hospital, both the ED visit and the hospital visit were considered. For encounters in the outpatient setting, we required a direct interaction between the provider and the patient and a documented diagnosis or procedure code. Encounters for a laboratory test or a procedure only were not included.

For telehealth encounters, telephone appointment visits and video visits were conducted synchronously using real-time telephone or live video-audio interaction, and they were billable and had a diagnosis or procedure code. Thus, telephone appointment visits and video visits were considered telehealth visits in this study. On the other hand, e-visits and message-only encounters were for patient self-triage and for communications without a real-time provider evaluation component. They were not considered telehealth visits in this study. Claims with a telehealth place-of-service code or with the 95 modifier, indicating that the services were delivered through telehealth, were considered telehealth visits in accordance with the CMS billing rules [13].

Rates of Health Care Utilization During the Pandemic and Prepandemic Years

The rates of visits from these 4 care settings were calculated weekly (Sunday to Saturday) for the prepandemic year and the pandemic year. The numerator was the visit counts of each type, and the denominator was 100 person-years of membership during a given week.

Statistical Analyses

We first plotted monthly KPSC member enrollment in 2019 and 2020. We examined the demographic characteristics of the cohort, including age, gender, race and ethnicity, and mean Charlson comorbidity index (CCI) of KPSC members in June 2019 and June 2020. CCI scores were calculated only for individuals aged ≥18 years. The visit rates by week during the prepandemic and pandemic years were plotted separately for inpatient, ED, outpatient, and telehealth visits.

In addition to plotting the trends, we used Poisson models to assess the significance of changes in health care utilization after versus before the onset of the pandemic in 2020 relative to

changes across the same time periods in 2019, using a difference-in-difference (DID) analysis. To achieve this goal, we selected week 8 (February 23, 2020) as the timepoint before the pandemic because the governor of California declared a state of emergency on March 4, 2020. We also chose week 12 (March 22, 2020) as the timepoint after the start of the pandemic because a stay-at-home order was enacted in California on March 19, 2020. We then selected the 2 corresponding time points during the prepandemic year. In Poisson models, the number of visits was the dependent variable, and an indicator variable for the 2 time points (ie, t=0 for week 8 and t=1 for week 12), an indicator variable for the year (2019 and 2020), and an interaction between the 2 variables were the independent variables. The interaction term was included in the DID analysis to directly assess the significance of the difference in the changes in the visit rates across the 2 years. In these Poisson models, we also included the natural log of person-years as an offset and adjusted for overdispersion of the count data. Because weekly visit data of the entire population were analyzed, individual-level covariates were not included in the analyses.

Results

Results Overview

Although the member enrollment number in the KPSC slightly decreased from July to October 2020 (4.57 million to 4.55 million), it remained steady during the pandemic year with a range of 4.55-4.57 million, slightly higher than 4.47-4.48 million in 2019 (Figure 1).

Similarly, the characteristics of KPSC members, such as age, gender, race and ethnicity, and mean CCI did not differ between June 2019 and June 2020 (Table 1). The impact of the pandemic on health care utilization in the KPSC was observed after week 8 (February 23) in 2020 (Figures 2-6).

Figure 1. Monthly member enrollment in the Kaiser Permanente Southern California in 2019 and 2020.

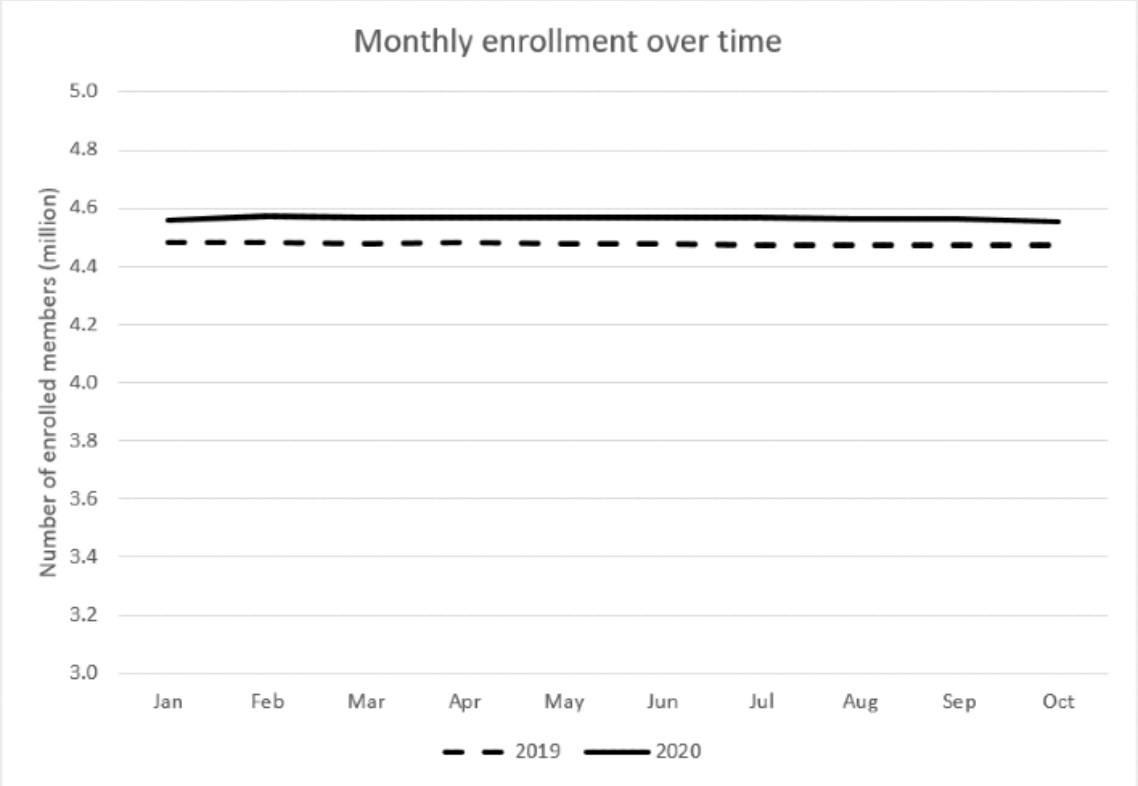


Table 1. Demographic characteristics and the Charlson comorbidity index of Kaiser Permanente Southern California members in June 2019 and June 2020.

Demographic characteristics and CCI[a]	June 2019 (n=4,475,819)	June 2020 (n=4,566,641)
Age (years), (%)		
0-17	20.8	20.9
18-44	38.1	39.1
45-64	26.2	26.6
≥65	14.9	15.4
Females, (%)	51.5	50.6
Race and Ethnicity, (%)		
Hispanic	40.9	41.3
Non-Hispanic White	31.4	31.0
Non-Hispanic Black	7.8	7.8
Non-Hispanic Asian or Pacific Islander	11.2	11.3
Non-Hispanic Native American or Alaskan	0.2	0.2
Non-Hispanic Multiple Races, others, or unknown	8.4	10.4
Mean CCI (SD)	0.48 (0.96)	0.45 (0.93)

[a]CCI: Charlson comorbidity index calculated for individuals aged ≥18 years with minimum 1 year of enrollment; n=3,055,756 in 2019 and n=3,115,974 in 2020.

Figure 2. Inpatient visit rate over time. DID: difference in difference.

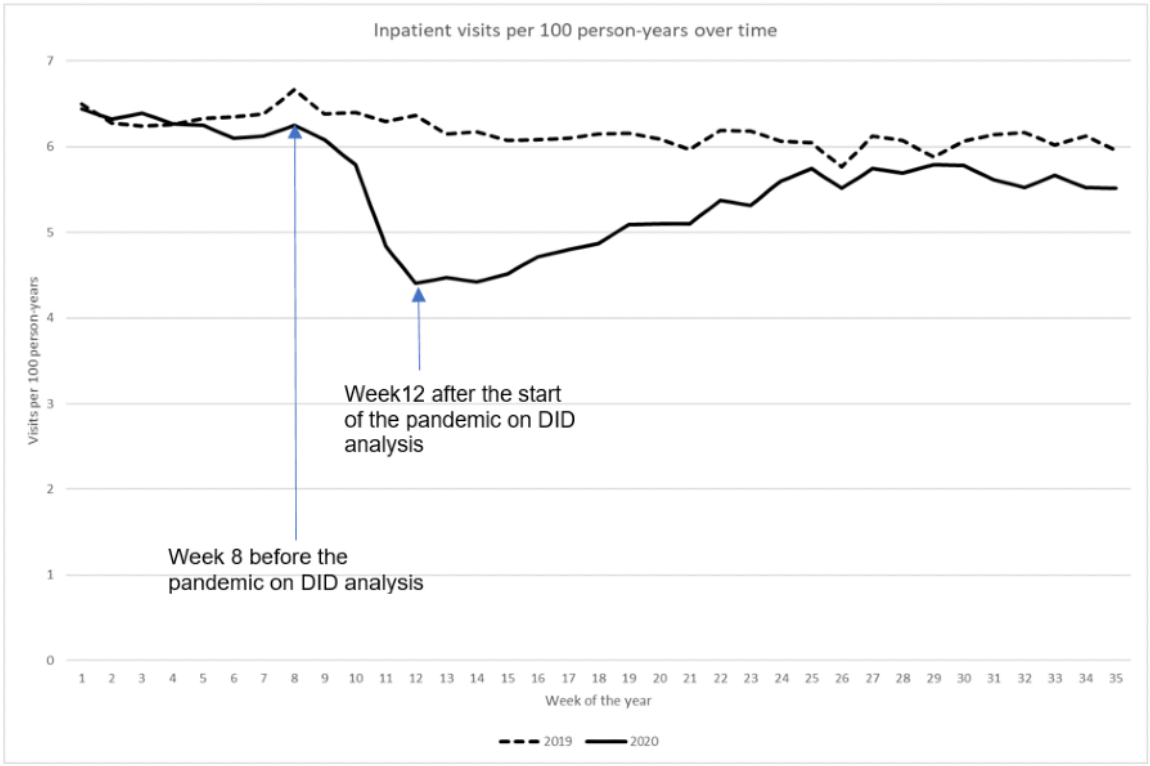


Figure 3. Emergency department visit rate over time. DID: difference in difference.

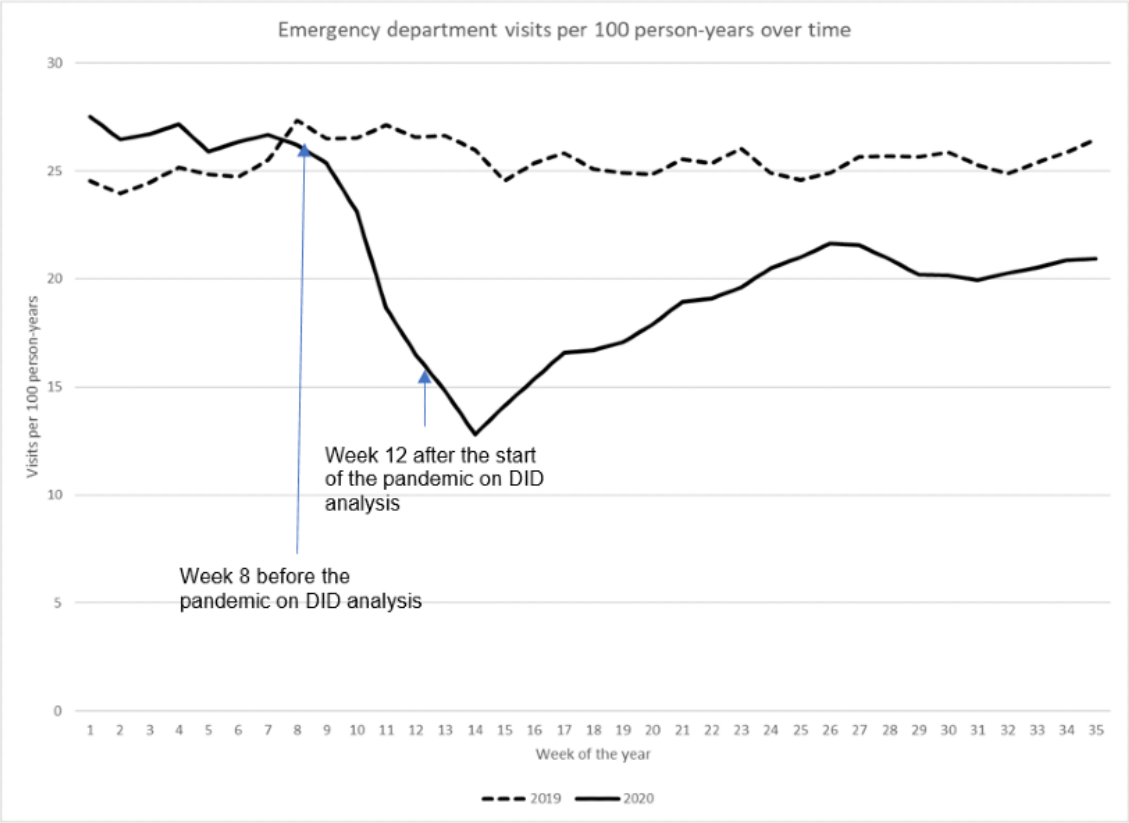


Figure 4. Outpatient visit rate over time. DID: difference in difference.

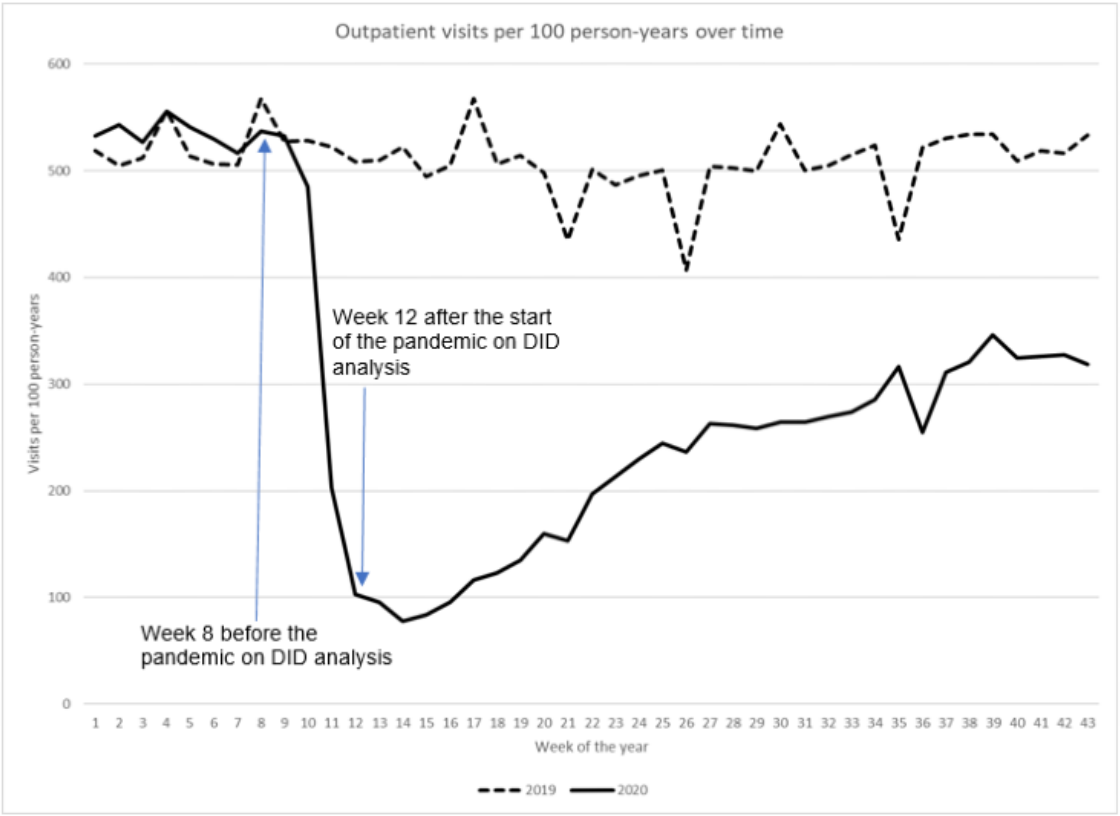


Figure 5. Telehealth visit rate over time. DID: difference in difference.

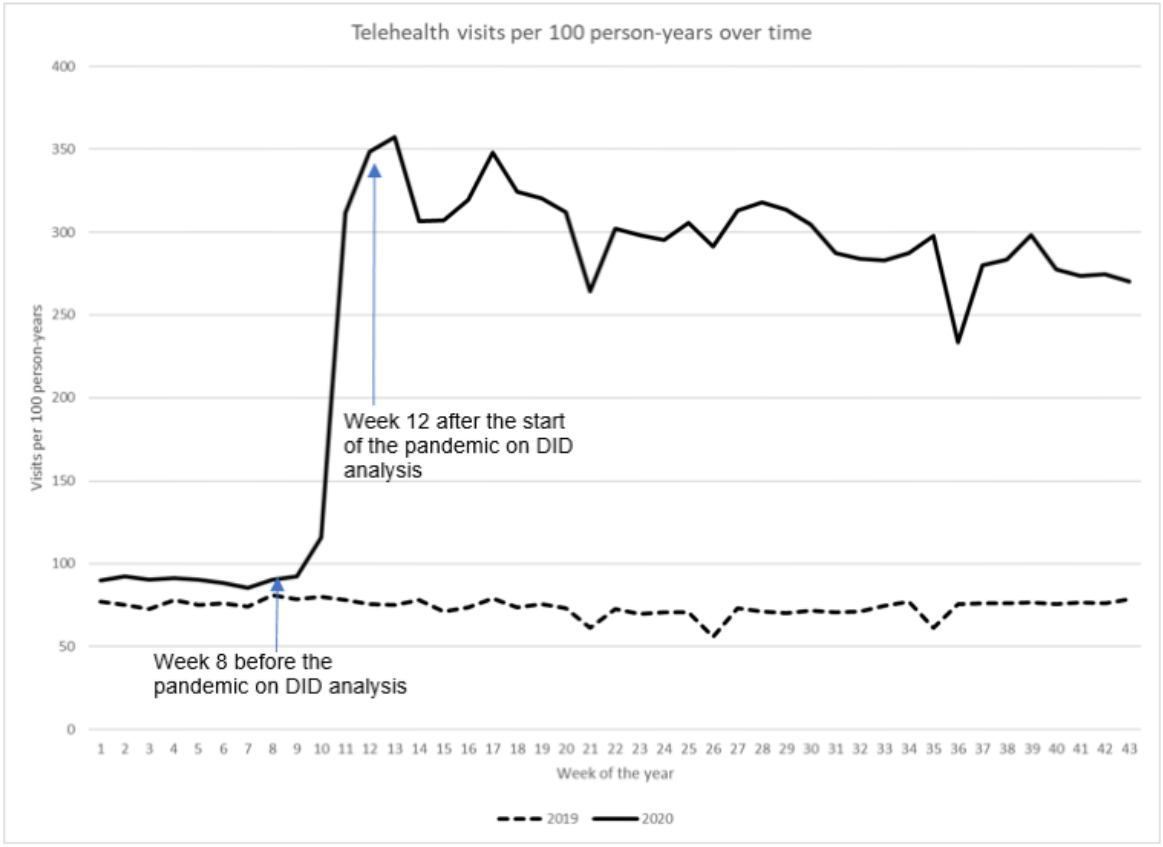
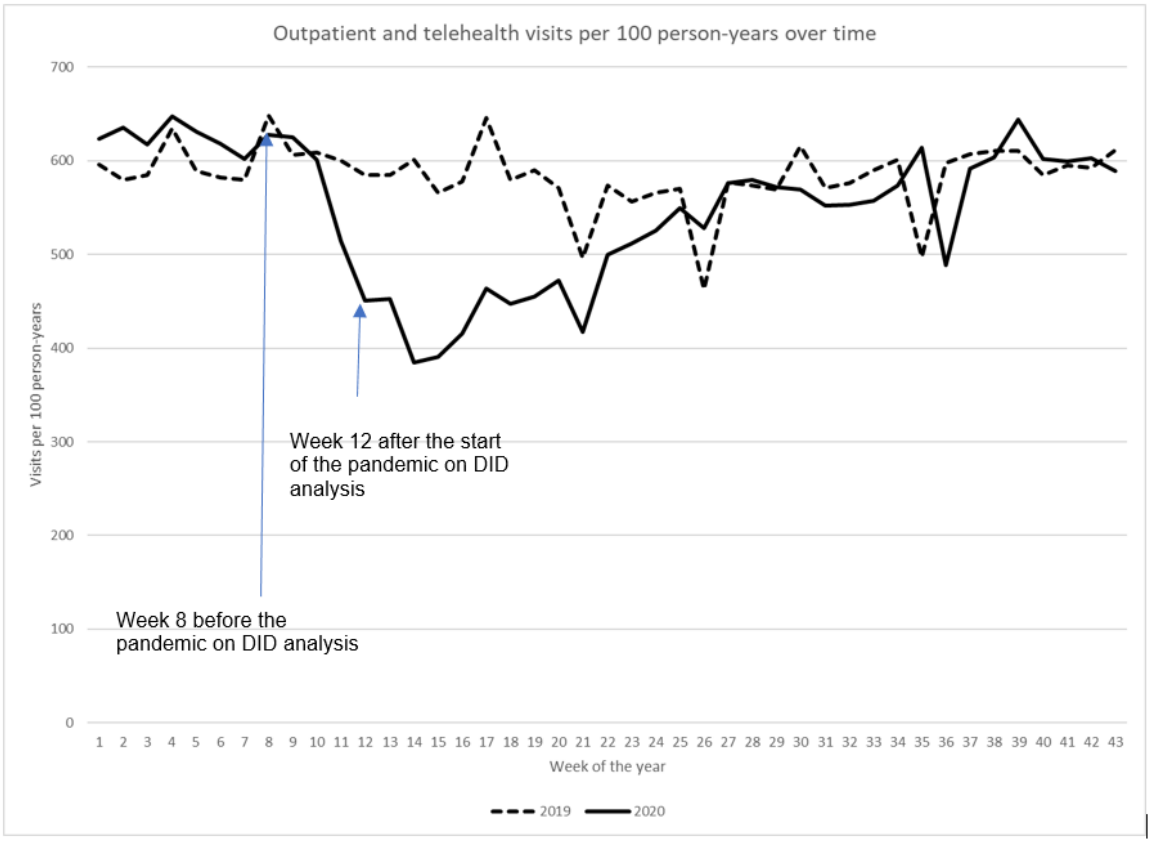


Figure 6. Combined outpatient and telehealth visit rate over time. DID: difference in difference.



Inpatient Visits

The inpatient visits per 100 person-years significantly decreased from 6.3 in week 8 (February 23) to 4.4 in week 12 (March 22) during the pandemic year (P<.001), thus displaying a 30.2% reduction, while the inpatient visit rate only slightly decreased from 6.7 to 6.4 between the same weeks in 2019 (Figure 2). DID analysis revealed that after adjusting for prepandemic secular trends, the reduction in inpatient visit rates from weeks 8-12 during the pandemic was 1.6 visits per 100 person-years (P<.001). After week 12 in the pandemic year, the inpatient visit rate increased until week 30 (July 26) but did not approach prepandemic levels (week 8 in 2020); the inpatient visit rate decreased again after week 30. In contrast, during the prepandemic year, the inpatient visit rate remained at approximately 6 per 100 person-years after week 12.

ED Visits

ED visits per 100 person-years significantly decreased from 26.2 in week 8 (February 23) to 16.5 in week 12 (March 22) during the pandemic year (P<.001), thus displaying a 37.0% reduction, while the ED visit rate slightly decreased from 27.3 in week 8 to 26.6 in week 12 during the prepandemic year (Figure 3). DID analysis revealed that after adjusting for prepandemic secular trends, the reduction in ED visit rates from week 8 to week 12 during the pandemic year was 8.9 visits per 100 person-years (P<.001). After week 12 during the pandemic year, the ED visit rate plummeted in week 14 (12.8) and increased to 21.6 in week 27 (July 5), but did not approach prepandemic levels. ED visit rates remained largely unchanged afterwards. In contrast, the average ED visit rate throughout the prepandemic year was 25.5 (range 24.0-27.3) per 100 person-years. The ED visit rates during February 24 to April 6, 2019, were slightly higher than those in the rest of 2019.

Outpatient Visits

The outpatient visits per 100 person-years drastically decreased from 537.3 in week 8 (February 23) to 102.8 in week 12 (March 22) during the pandemic year (P<.001), thus displaying a 80.9% reduction, while the outpatient visit rate slightly decreased from 567.2 to 508.9 for the same period during the prepandemic year (Figure 4). DID analysis revealed that after adjusting for prepandemic secular trends, the reduction in outpatient visit rates from week 8 to week 12 during the pandemic was 367.2 visits per 100 person-years (P<.001). After week 12 during the pandemic year, the outpatient visit rate decreased to 77.8 in week 14 and increased to 346.0 in week 39 (September 27), amounting to only 64.4% of the outpatient visit rate in week 8. In contrast, outpatient visit rates fluctuated during the prepandemic year with an average of 502.3 (SD 45.9) per 100 person-years and did not decrease abruptly as it did in 2020.

Telehealth Visits

In contrast with in-person visits, telehealth visits increased drastically after the onset of the pandemic (Figure 5). While the trend of the telehealth visit rate remained relatively steady during 2019, we observed an approximately 4-fold increase in the telehealth visit rate during the early days of the pandemic year: 90.4 visits per 100 person-years in week 8 (February 23) to 348.3 in week 12 (March 22). Although these rates decreased after week 13, the weekly rate at the end of the study period (October 25) was still almost 3-fold that in week 8. DID analysis revealed that after adjusting for prepandemic secular trends, the increase in telehealth visit rates from week 8 to week 12 was 272.9 visits per 100 person-years (P<.001) during the pandemic year.

Outpatient and Telehealth Visits

To determine whether the increase in telehealth visits offsets the reduction in outpatient visits, we calculated the rate of combined telehealth and outpatient visits (Figure 6). Although not as drastic as the rate of outpatient visits alone, the rate of combined telehealth and outpatient visits decreased from 627.7 visits per 100 person-years to 451.1, thus displaying a 28.1% reduction from week 8 (February 23) to week 12 (March 22) during the pandemic year. After week 12, the rate of combined telehealth and outpatient visits increased, having approached that in week 26 (June 28) in the prepandemic year.

Discussion

Principal Findings

In this study, we observed significant reductions in in-person medical visits as the pandemic progressed. The greatest reduction was observed in outpatient visits in the early days of the pandemic (80.9%). Although of lesser magnitude, inpatient and ED visits also decreased by 30.2% and 37.0%, respectively, during the early days of the pandemic. By contrast, we observed an approximately 4-fold increase in telehealth visits in weeks 8-12 in the pandemic year. Further analyses suggest that the increase in telehealth visits did not offset the reduction in outpatient visits during the early days of the pandemic; however, it did compensate for the reduction in outpatient visits by week 26 (June 28). In addition to the CDC recommendation for the use of telehealth services [11], federal and state governments have issued changes in reimbursement policies for these services [14,15]. Even though the pandemic continues to progress with periodic surges in COVID-19 cases and hospitalizations, these policy changes have helped providers deliver health care in telehealth settings.

Our study sheds light on the impact of the pandemic on health care utilization. With approximately 10 months' data during 2020, this study provides insights into patterns of health care utilization during the pandemic. By using visit rates as our outcomes, we could account for the changes in the underlying population denominator during the pandemic. We observed that KPSC membership generally remained stable during the pandemic, largely owing to the KPSC's decision to not cancel health coverage for groups or individuals who could not pay for most of the study period. By comparing health care utilization during the pandemic year to that in the prepandemic year through DID analyses, we show that these findings did not result from simply an exacerbation of seasonal effects. Robinson et al [16] recently described the transition to virtual care at the KPSC, but in contrast to our study, they used counts of visits instead of rates as outcomes; hence, they did not adjust for population size. In addition, they included all types of virtual care (including those intended for communication), did not use data from the prepandemic year, and did not conduct a DID

analyses. Furthermore, our study included data for 3 additional months.

Limitations

Some potential limitations in this study must be recognized. First, in addition to the COVID-19 pandemic, other factors such as civil unrest due to racial injustice and the wildfires on the West Coast may have influenced how patients sought health care. We could not differentiate the impact of these factors on health care utilization. Second, these results were derived from a large integrated health care organization that might have been able to change practices quickly, thus potentially not reflecting patterns in other health care systems. Third, while we studied the impact of the pandemic on health care utilization, we did not address the quality of care and population health.

Conclusions

In conclusion, in-person health care utilization decreased drastically during the early period of the pandemic, but there was a corresponding increase in telehealth visits during the same period. By the end of June 2020, the rate of combined outpatient and telehealth visits reverted to prepandemic levels.

Acknowledgments

The study was supported by internal funds of the KPSC.

Conflicts of Interest

SJ reports research grant funding from Dynavax Technologies. All other authors declare no conflicts.

References

1. Baum A, Kaboli P, Schwartz M. Reduced In-Person and Increased Telehealth Outpatient Visits During the COVID-19 Pandemic. Ann Intern Med 2021 Jan;174(1):129-131 [FREE Full text] [doi: 10.7326/M20-3026] [Medline: 32776780]
2. Patel SY, Mehrotra A, Huskamp HA, Uscher-Pines L, Ganguli I, Barnett ML. Trends in Outpatient Care Delivery and Telemedicine During the COVID-19 Pandemic in the US. JAMA Intern Med 2021 Mar 01;181(3):388-391. [doi: 10.1001/jamainternmed.2020.5928] [Medline: 33196765]
3. Kazi D, Wadhera R, Shen C, Ho K, Patell R, Selim M, et al. Decline in Emergent and Urgent Care during the COVID-19 Pandemic. medRxiv. Preprint posted online May 18, 2020. [doi: 10.1101/2020.05.14.20096602]
4. Jeffery MM, D'Onofrio G, Paek H, Platts-Mills TF, Soares WE, Hoppe JA, et al. Trends in Emergency Department Visits and Hospital Admissions in Health Care Systems in 5 States in the First Months of the COVID-19 Pandemic in the US. JAMA Intern Med 2020 Oct 01;180(10):1328-1333 [FREE Full text] [doi: 10.1001/jamainternmed.2020.3288] [Medline: 32744612]
5. Blecker S, Jones SA, Petrilli CM, Admon AJ, Weerahandi H, Francois F, et al. Hospitalizations for Chronic Disease and Acute Conditions in the Time of COVID-19. JAMA Intern Med 2021 Feb 01;181(2):269-271. [doi: 10.1001/jamainternmed.2020.3978] [Medline: 33104158]
6. Bhambhvani H, Rodrigues A, Yu J, Carr JB, Hayden Gephart M. Hospital Volumes of 5 Medical Emergencies in the COVID-19 Pandemic in 2 US Medical Centers. JAMA Intern Med 2021 Feb 01;181(2):272-274. [doi: 10.1001/jamainternmed.2020.3982] [Medline: 33104161]
7. Hartnett KP, Kite-Powell A, DeVies J, Coletta MA, Boehmer TK, Adjemian J, National Syndromic Surveillance Program Community of Practice. Impact of the COVID-19 Pandemic on Emergency Department Visits - United States, January 1, 2019-May 30, 2020. MMWR Morb Mortal Wkly Rep 2020 Jun 12;69(23):699-704 [FREE Full text] [doi: 10.15585/mmwr.mm6923e1] [Medline: 32525856]
8. Gavin C. As COVID-19 spreads, Massachusetts Hospitals See Decline in Visits for Other Illnesses. Boston.com. 2020 Mar 30. URL: https://www.boston.com/news/health/2020/03/30/massachusetts-coronavirus-emergency-rooms [accessed 2020-09-15]
9. Lange SJ, Ritchey MD, Goodman AB, Dias T, Twentyman E, Fuld J, et al. Potential Indirect Effects of the COVID-19 Pandemic on Use of Emergency Departments for Acute Life-Threatening Conditions - United States, January-May 2020. MMWR Morb Mortal Wkly Rep 2020 Jun 26;69(25):795-800 [FREE Full text] [doi: 10.15585/mmwr.mm6925e2] [Medline: 32584802]
10. CMS Releases Recommendations on Adult Elective Surgeries, Non-Essential Medical, Surgical, and Dental Procedures During COVID-19 Response. Centers for Medicare and Medicaid Services. URL: https://www.cms.gov/newsroom/press-releases/cms-releases-recommendations-adult-elective-surgeries-non-essential-medical-surgical-and-dental [accessed 2020-09-15]
11. Using Telehealth to Expand Access to Essential Health Services during the COVID-19 Pandemic. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/coronavirus/2019-ncov/hcp/telehealth.html [accessed 2020-09-15]
12. Koebnick C, Langer-Gould AM, Gould MK, Chao CR, Iyer RL, Smith N, et al. Sociodemographic characteristics of members of a large, integrated health care system: comparison with US Census Bureau data. Perm J 2012;16(3):37-41 [FREE Full text] [doi: 10.7812/tpp/12-031] [Medline: 23012597]

13. COVID-19: Telehealth Billing Correction, Nursing Home Recommendations, Billing for Multi-Function Ventilators, New ICD-10-CM Diagnosis Code. Centers for Medicare and Medicaid Services. 2020 Apr 03. URL: https://www.cms.gov/outreach-and-educationoutreachffsprovpartprogprovider-partnership-email-archive/2020-04-03-mlnc-se [accessed 2021-04-26]
14. Telehealth: Health care from the safety of our homes. Health Resources and Services Administration. URL: https://telehealth.hhs.gov/ [accessed 2021-04-26]
15. Understanding telehealth policy. Center for Connected Health Policy. URL: https://www.cchpca.org/telehealth-policy/current-state-laws-and-reimbursement-policies [accessed 2021-04-26]
16. Robinson J, Borgo L, Fennell K, Funahashi T. The Covid-19 Pandemic Accelerates the Transition to Virtual Care. NEJM Catalyst 2020;1(5). [doi: 10.1056/CAT.20.0399]

Abbreviations

CCI: Charlson comorbidity index
CDC: Centers for Disease Control and Prevention
CMS: Centers for Medicare and Medicaid Services
DID: difference in difference
ED: emergency department
EHR: electronic health record
KPSC: Kaiser Permanente Southern California

Edited by C Basch; submitted 16.12.20; peer-reviewed by A Baum, M Jeffery, X Wang; comments to author 08.01.21; revised version received 23.01.21; accepted 17.04.21; published 29.04.21.

Please cite as:
Xu S, Glenn S, Sy L, Qian L, Hong V, Ryan DS, Jacobsen S
Impact of the COVID-19 Pandemic on Health Care Utilization in a Large Integrated Health Care System: Retrospective Cohort Study
J Med Internet Res 2021;23(4):e26558
URL: https://www.jmir.org/2021/4/e26558
doi:10.2196/26558
PMID:33882020

Original Paper

Knowledge About COVID-19 Among Adults in China: Cross-sectional Online Survey

Fengyun Yu[1,2*], BSc; Pascal Geldsetzer[2,3*], MD, ScD; Anne Meierkord[4], MD; Juntao Yang[5], PhD; Qiushi Chen[6], PhD; Lirui Jiao[7], BA; Nadeem E Abou-Arraj[8], MD; An Pan[9], PhD; Chen Wang[10,11,12*], MD, PhD; Till Bärnighausen[2,10*], MD, ScD; Simiao Chen[2,10*], ScD

[1]Department of Industrial Engineering, Tsinghua University, Beijing, China

[2]Heidelberg Institute of Global Health, Faculty of Medicine and University Hospital, Heidelberg University, Heidelberg, Germany

[3]Division of Primary Care and Population Health, Department of Medicine, Stanford University School of Medicine, Stanford, CA, United States

[4]Faculty of Medicine, University of Southampton, Southampton, United Kingdom

[5]State Key Laboratory of Medical Molecular Biology, Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China

[6]The Harold and Inge Marcus Department of Industrial and Manufacturing Engineering, The Pennsylvania State University, University Park, PA, United States

[7]Reed College, Portland, OR, United States

[8]Department of Medicine, Stanford University School of Medicine, Stanford, CA, United States

[9]School of Public Health, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China

[10]Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China

[11]National Clinical Research Center for Respiratory Diseases, Beijing, China

[12]Department of Pulmonary and Critical Care Medicine, Center of Respiratory Medicine, China–Japan Friendship Hospital, Beijing, China

[*]these authors contributed equally

Corresponding Author:
Simiao Chen, ScD
Heidelberg Institute of Global Health
Faculty of Medicine and University Hospital
Heidelberg University
Im Neuenheimer Feld 130.3
Heidelberg
Germany
Phone: 49 15207926273
Email: simiao.chen@uni-heidelberg.de

Related Article:

This is a corrected version. See correction statement: https://www.jmir.org/2021/5/e30100

Abstract

Background: A detailed understanding of the public's knowledge and perceptions of COVID-19 could inform governments' public health actions in response to the pandemic.

Objective: The aim of this study was to determine the knowledge and perceptions of COVID-19 among adults in China and its variation among provinces and by sociodemographic characteristics.

Methods: Between May 8 and June 8, 2020, we conducted a cross-sectional online survey among adults in China who were registered with the private survey company KuRunData. We set a target sample size of 10,000 adults, aiming to sample 300-360 adults from each province in China. Participants were asked 25 questions that tested their knowledge about COVID-19, including measures to prevent infection, common symptoms, and recommended care-seeking behavior. We disaggregated responses by age; sex; education; province; household income; rural–urban residency; and whether or not a participant had a family member, friend, or acquaintance who they know to have been infected with SARS-CoV-2. All analyses used survey sampling weights.

Results: There were 5079 men and 4921 women who completed the questionnaire and were included in the analysis. Out of 25 knowledge questions, participants answered a mean and median of 21.4 (95% CI 21.3-21.4) and 22 (IQR 20-23) questions correctly, respectively. A total of 83.4% (95% CI 82.7%-84.1%) of participants answered four-fifths or more of the questions correctly. For at least one of four ineffective prevention measures (using a hand dryer, regular nasal irrigation, gargling mouthwash, and taking antibiotics), 68.9% (95% CI 68.0%-69.8%) of participants answered that it was an effective method to prevent a SARS-CoV-2 infection. Although knowledge overall was similar across provinces, the percent of participants who answered the question on recommended care-seeking behavior correctly varied from 47.0% (95% CI 41.4%-52.7%) in Tibet to 87.5% (95% CI 84.1%-91.0%) in Beijing. Within provinces, participants who were male, were middle-aged, were residing in urban areas, and had higher household income tended to answer a higher proportion of the knowledge questions correctly.

Conclusions: This online study of individuals across China suggests that the majority of the population has good knowledge of COVID-19. However, a substantial proportion still holds misconceptions or incorrect beliefs about prevention methods and recommended health care–seeking behaviors, especially in rural areas and some less wealthy provinces in Western China. This study can inform the development of tailored public health policies and promotion campaigns by identifying knowledge areas for which misconceptions are comparatively common and provinces that have relatively low knowledge.

(J Med Internet Res 2021;23(4):e26940) doi:10.2196/26940

KEYWORDS

COVID-19; knowledge; perception; risk; public health; China; cross-sectional; survey

Introduction

COVID-19 has taken a large toll on public health and economic growth worldwide [1-4]. Assessing the perception and knowledge among the public during infectious disease outbreaks is essential to inform public health campaigns. Research has shown that governmental policies can have a substantial impact on community transmission of SARS-CoV-2 [5-8]. It is likely that the more detailed an understanding governments have of their population's knowledge and perceptions of COVID-19, the more effectively they can design policies to contain COVID-19 in their population, whether this is on the national, regional, or local level.

A cross-sectional survey conducted in late February 2020 that assessed the public's perceptions of COVID-19 in the United States and the United Kingdom found that a considerable proportion of adults had misconceptions about infection prevention methods and care-seeking behaviors [9]. For example, over a third of survey participants selected at least one of the following options when asked whether they are effective prevention measures: using a hand dryer, rinsing your nose with saline, taking antibiotics, or gargling with mouthwash. Reasons for these false beliefs are unclear but could be different for populations in East Asian countries like China or Singapore, which were affected by the severe acute respiratory syndrome outbreak in 2002-2004.

Assessing the population's perceptions and knowledge of COVID-19 is not only essential to understanding and comparing different behaviors and policy decisions retrospectively but also vital for informing postlockdown policies since it will be crucial that people follow infection prevention methods as they start to interact more [10,11]. As of November 2020, China appears to have been successful in containing the spread of COVID-19 and has reported low case numbers [12]. However, further waves of COVID-19 may emerge in parts of China over the coming months. Therefore, collecting data about the knowledge and perception of COVID-19 across China is imperative.

Several studies have assessed the perception of COVID-19 within China. However, they have either focused on specific subgroups, such as pregnant women [13] or patients with mental health disorders [14], or assessed risk perception within particular contexts such as tourism [15]. To our knowledge, our study is the first large-scale survey that assessed COVID-19 perception and knowledge among the public in all provinces of China. This study aims to inform Chinese policy makers on the knowledge and perceptions of their population with regard to COVID-19 to facilitate effective policy design during future waves of the pandemic.

Methods

Sampling Process

The survey was implemented by KuRunData, an online private survey company that maintains a database of potential survey participants and delivers surveys. KuRunData recruits members through its own platform [16], partnerships with other websites, and encouraging registered members to recruit new members through the popular mobile app Wechat Mini. KuRunData verifies that members have access to mobile phones and the internet, and are capable of navigating online surveys. For this study, we used KuRunData to sample 300-360 participants in each of China's 31 provincial-level administrative units, with the total sample size goal being 10,000 adults. Potential participants were unable to access the questionnaire as soon as this sample size goal was reached. Within each province, KuRunData aimed to sample a proportion of participants that was reflective of the demographic composition of the province's population (as per the 2019 China Statistical Yearbook [17]) by sex and urban-rural residence. Adults in the survey pool were invited to participate in the survey by KuRunData's own platform. They were informed that they would receive ¥5 (US $0.77) for completing the questionnaire. Before filling in the questionnaire, participants had to provide their informed written consent with signature confirmation. The informed consent page described the project's background and purpose, the possible risks, the payment after completing the questionnaire, and the

confidentiality of information and records. To be able to access the questionnaire, the participants must have read the informed consent description for at least 15 seconds and self-declared understanding of the purpose and risks of the study before signing. The survey was administered between May 8 and June 8, 2020.

Questionnaire

The questionnaire was built in the KuRunData platform and had 25 questions partitioned into the following sections: introduction, perceived risk of death from COVID-19, mode of transmission of COVID-19, recognizing and acting upon an infection, sociodemographic characteristics, and specific questions about possible misconceptions or falsehoods on COVID-19 prevention and symptoms that were drawn from the World Health Organization's "myth busters" website [18]. The questionnaire was written in Standard Chinese and is shown in Text A1 in Multimedia Appendix 1. Participants had to answer a question to reach the next question. Numerical entry questions did not allow for nonsensical inputs (eg, percentage questions were restricted to inputs between 0 and 100).

Data Quality Checks

Three types of data quality checks were performed. First, we verified the time taken to complete the questionnaire and excluded participants who took less than 2 minutes to complete the questionnaire under the assumption that these participants did not read the questions. Second, we plotted the distribution of the time taken to complete the questionnaire. If some respondents used random clicking to complete the questionnaire as fast as possible, then a bimodal distribution in the time taken to complete the survey might be expected (with one study population clicking as quickly as possible and one reading the questions). Third, participants were asked whether they looked up any answers online and, if so, for which questions. Those who self-reported having looked up the answer online for a particular question were excluded from the analysis for that question in the supplementary analyses shown in Multimedia Appendix 1.

Data Analysis

We excluded participants who answered less than half of the questions in the questionnaire. All analyses used sampling weights to account for the complex survey design. The sampling weights were the inverse of the probability of selecting participants given the following variables: gender, rural versus urban residence, and province. These probabilities were calculated using population counts from the 2019 China Statistical Yearbook within each province. For binary and categorical response options, we computed the percentage of participants who selected each response to summarize the survey findings. For binomial proportions, we constructed two-sided 95% CIs using the Wilson score interval. In addition, we computed a total score for participants, which consists of the number of COVID-19 knowledge questions that were answered correctly. We henceforth refer to this score as the *overall knowledge score*. To examine how knowledge and perceptions varied by participants' characteristics, we used ordinary least squares regression to regress this overall score and the response to each question onto age (10-year age group); sex; educational attainment; province; rural versus urban residence; vocation; household income; and whether or not a participant had a family member, friend, or acquaintance who they knew to have been infected with SARS-CoV-2. All regressions included only one of these variables plus a binary indicator for each province (province-level fixed effects). We show regression results that we additionally adjusted for 10-year age group and sex in Multimedia Appendix 1.

Ethics

This research was considered to not involve human participants by the institutional review board of the Heidelberg University Hospital because all authors only had access to deidentified data.

Results

Sample Characteristics

A total of 14,493 adults agreed to take the online survey. After excluding participants who did not complete the whole survey or who took less than 2 minutes to complete the questionnaire, 10,000 participants (all of whom completed all survey questions) were included in the analysis. There was no evidence of a bimodal distribution in the time taken to complete the questionnaire (Multimedia Appendix 1 Figure A1). A total of 3643 participants reported looking up the answer online on a median of 2 questions (IQR 1-3).

There were 5079 males and 4921 females from 31 provinces that completed the questionnaire. Their sociodemographic characteristics are shown in Table 1. Around one-tenth of the 10,000 participants (n=900, 9.0%) were aged 18 or 19 years, 16.5% (n=1645) were aged 20-29 years, 19.0% (n=1895) were aged 30-39 years, and 16.8% (n=1675) were 60 years or older. A total 37.3% (n=3733) of the participants had received high school or technical secondary school education, and one-third (n=3369, 33.7%) had completed an undergraduate degree. Only 4.4% (n=438) and 4.8% (n=475) of participants had never been to school or had been to elementary school only, respectively. The majority of participants (n=5935, 59.3%) lived in urban areas. The number of participants per province ranged from 300 to 360. About half (n=5007, 47.0%) of participants reported to have an annual total household income between ¥60,000 (US $9180) and ¥119,999 (US $18,360).

Table 1. Sample characteristics.

Characteristic	Proportion of participants (weighted[a]), %	Participants (not weighted), n (%)	Population of China, %[b]
Sex			
Female	56.5	4921 (49.2)	48.8
Age group (years)			
<20	10.6	900 (9.0)	6.9
20-29	16.9	1645 (16.5)	20.8
30-39	18.4	1895 (19.0)	18.2
40-49	18.5	1890 (18.9)	22.1
50-59	17.9	1820 (18.2)	16.5
>60	17.7	1675 (16.8)	15.4
Education			
Never been to school	4.1	438 (4.4)	5.4
Elementary school	4.3	475 (4.8)	25.3
Middle school	16.3	1779 (17.8)	37.8
High school/technical secondary school	35.7	3733 (37.3)	17.6
College/undergraduate	37.3	3369 (33.7)	13.4
Graduate and above	2.2	206 (2.0)	0.6
Ethnicity			
Han	95.1	9381 (93.8)	95.0
Man	0.5	149 (1.5)	0.7
Hui	0.1	109 (1.1)	0.8
Zang	1.6	103 (1.0)	0.5
Zhuang	1.5	152 (1.5)	1.2
Other	1.1	106 (1.1)	1.8
Province of current residence			
Anhui	4.4	360 (3.6)	4.5
Beijing	1.8	360 (3.6)	1.5
Chongqing	2.3	360 (3.6)	2.2
Fujian	2.8	300 (3.0)	2.8
Gansu	1.8	300 (3.0)	1.9
Guangdong	8.6	360 (3.6)	8.1
Guangxi	3.3	300 (3.0)	3.5
Guizhou	2.4	300 (3.0)	2.6
Hainan	0.7	300 (3.0)	0.7
Hebei	5.3	360 (3.6)	5.4
Heilongjiang	2.7	300 (3.0)	2.7
Henan	6.3	360 (3.6)	6.9
Hubei	4.2	360 (3.6)	4.2
Hunan	4.8	300 (3.0)	4.9
Jiangsu	6.1	360 (3.6)	5.8
Jiangxi	3.2	300 (3.0)	3.3
Jilin	1.9	300 (3.0)	1.9

Characteristic	Proportion of participants (weighted[a]), %	Participants (not weighted), n (%)	Population of China, %[b]
Liaoning	3.3	340 (3.4)	3.1
Neimengol	1.8	300 (3.0)	1.8
Ningxia	0.5	300 (3.0)	0.5
Qinghai	0.4	300 (3.0)	0.4
Shaanxi	2.7	360 (3.6)	2.8
Shandong	7.3	360 (3.6)	7.2
Shanghai	2.0	300 (3.0)	1.7
Shanxi	2.6	300 (3.0)	2.7
Sichuan	5.7	360 (3.6)	6.0
Tianjin	1.3	360 (3.6)	1.1
Tibet	1.7	300 (3.0)	0.2
Xinjiang	0.2	300 (3.0)	1.8
Yunnan	3.2	300 (3.0)	3.5
Zhejiang	4.3	360 (3.6)	4.1
Rural–urban residency			
Urban	69.5	5935 (59.3)	59.6
Works as a health care provider			
No	96.0	9597 (96.0)	99.0
Nurse	0.5	55 (0.6)	0.3
Physician	0.8	84 (0.8)	0.5
Community health worker	1.5	157 (1.6)	<0.1
Pharmacist	0.2	17 (0.2)	<0.1
Other health care provider	1.0	90 (0.9)	0.1
Annual household income, ¥ (US $)			
<30,000 (3835)	5.7	560 (5.6)	—[c]
30,000-59,999 (3835-7670)	14.1	1670 (16.7)	—
60,000-89,999 (7670-11,505)	20.4	2303 (23.0)	—
90,000-119,999 (11,506-15,341)	25.0	2704 (24.0)	—
120,000-149,999 (15,341-19,176)	15.5	1211 (12.1)	—
150,000-199,999 (19,175-25,568)	12.2	974 (9.7)	—
≥200,000 (25,568)	7.2	578 (5.8)	—

[a]Weighted using survey sampling weights.

[b]As per the 2019 China Statistical Yearbook [17].

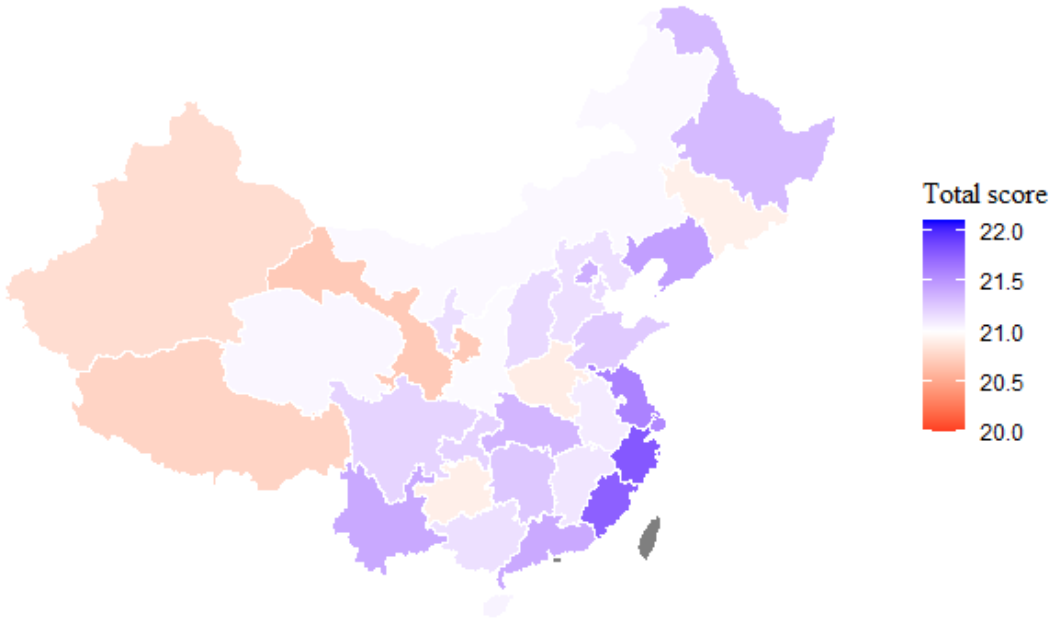
[c]Data not available.

Overall Knowledge Score

A total 83.4% (95% CI 82.7%-84.1%) of participants answered 80% or more (ie, 20 or more out of 25 questions) of the questions correctly, and almost all (98.4%, 95% CI 98.1%-98.6%) participants answered more than 60% of the questions correctly (Figure 1). The mean and median overall knowledge score was 21.4 (95% CI 21.3-21.4) and 22 (IQR 20-23), respectively. The distribution of the overall knowledge score is shown in Figure A2 in Multimedia Appendix 1.

Participants residing in the eastern provinces tended to have marginally higher overall knowledge scores. For instance, the mean knowledge score in the eastern province of Fujian was 21.9 (95% CI 21.6-22.1), whereas it was 20.9 (95% CI 20.6-21.2) in the western province of Gansu.

Figure 1. Map showing the mean overall knowledge score by province.



Perceived Risk of Death From a SARS-CoV-2 Infection

Survey participants' median estimate of the infection-fatality rate of COVID-19 was 3.2% (IQR 1.0%-3.6%; Table 2). When asked to estimate the percentage of patients infected with the common flu who die from the flu, participants' median response was 0.75% (IQR 0.10%-1.00%). Almost all (96.4%, 95% CI 96.0%-96.8%) participants identified that older adults were the age group most likely to die from COVID-19.

Table 2. Summary of survey findings.

Survey question and response[a]	Proportion or median estimate[b]
Perceived risk of death from a SARS-CoV-2 infection	
"What percent of individuals infected with the new coronavirus experience a fatal disease course?" (%), median (IQR)	
Continuous variable	3.2 (1.0-3.6)
"When they have been infected, what age groups are most likely to die from the illness caused by the new coronavirus?" % (95% CI)	
Children	25.0 (24.2-25.9)
Young adults	12.0 (11.4-12.6)
Older adults	96.4 (96.1-96.8)
"Are those with other health problems more likely to die from an infection with the new coronavirus disease than those without any other health problems?" % (95% CI)	
Yes	92.5 (92.0-93.0)
"What percent of people who get infected with the common flu end up dying from the common flu?" (%), median (IQR)	
Continuous variable	0.75 (0.10-1.00)
Transmission of SARS-CoV-2, % (95% CI)	
"Only older adults can become infected with the new coronavirus."	
False	98.3 (98.0-98.5)
"Is there currently a vaccine available that protects against infection with the new coronavirus?"	
No	78.9 (78.1-79.7)
"Which of the following actions help prevent catching an infection with the new coronavirus?"	
Selected all of the following: avoiding touching eyes, nose, and mouth with unwashed hands; washing your hands; and avoiding close physical contact with people who are sick	86.3 (85.6-86.9)
Selected at least one of the following: using a hand dryer, regularly rinsing your nose with saline, taking antibiotics, and gargling mouthwash	68.9 (68.0-69.8)
"Consistently wearing a face mask is highly effective in protecting you from getting infected with the new coronavirus."	
True	89.5 (88.9-90.1)
"What is the main way in which people are currently getting infected with the new coronavirus?"	
Droplets of saliva that land in the mouths or noses of people who are nearby when an infected person sneezes or coughs	82.2 (81.5-83.0)
Symptoms of COVID-19 and recommended health care–seeking behavior, % (95% CI)	
"What are common signs or symptoms of an infection with the new coronavirus?"	
Nose bleeds	9.7 (9.1-10.2)
Cough	98.2 (97.9-98.5)
Fever	99.5 (99.4-99.6)
Skin rash	8.7 (8.2-9.3)
Constipation	7.4 (6.9-7.9)
Shortness of breath	91.5 (91.0-92.1)
Frequent urination	4.5 (4.1-4.9)
"If you have a fever and new persistent cough that started today, what would you do?" % (95% CI)	
Go directly to a hospital	36.9 (35.9-37.8)
Call the official hotline	38.6 (37.7-39.6)
Continue my daily routine	10.9 (10.3-11.6)

[a]Response options are grouped to summarize categorical variables into a dichotomous measure.

[b]For dichotomous outcomes, data are expressed as percentage with correct response (95% CI). For continuous outcomes, data are expressed as median (IQR).

Transmission of SARS-CoV-2

A total 98.3% (95% CI 98.0%-98.5%) of participants correctly identified that COVID-19 does not only afflict older adults. When asked to identify the primary mode of transmission from a list of multiple choices, 82.2% (95% CI 81.5%-83.0%) of participants correctly selected a description of droplet transmission. A total 89.5% (95% CI 88.9%-90.1%) of participants answered that wearing a face mask was highly effective in protecting against infection, and 86.3% (95% CI 85.6%-86.9%) of participants selected correct behavioral prevention measures. However, 68.9% (95% CI 68.0%-69.8%) of participants answered that at least one of the following measures helps prevent SARS-CoV-2 infection: using a hand dryer, regular nasal irrigation, gargling mouthwash, and taking antibiotics.

Symptoms of COVID-19 and Recommended Health Care–Seeking Behavior

Most participants correctly identified the common symptoms of COVID-19: fever (99.5% answered correctly, 95% CI 99.4%-99.6%), cough (98.2% answered correctly, 95% CI 97.9%-98.5%), and shortness of breath (91.5% answered correctly, 95% CI 91.0%-92.1%). Conversely, when asked about symptoms that are not characteristic of COVID-19, participants correctly answered that these were not expected symptoms of the disease: relatively few participants believed that nose bleeds (9.7%, 95% CI 9.1%-10.2%), skin rash (8.7%, 95% CI 8.2%-9.3%), constipation (7.4%, 95% CI 6.9%-7.9%), or frequent urination (4.5%, 95% CI 4.1%-4.9%) were symptoms of COVID-19.

When asked what they should do if they developed new symptoms of fever and cough, 36.9% (95% CI 35.9%-37.8%) stated they would go directly to a hospital or contact their community health worker or other official contact person, 38.6% (95% CI 37.7%-39.6%) said they would call the official hotline, and 10.9% (95% CI 10.3%-11.6%) reported that they would continue their usual daily routines.

Variation in Knowledge and Perceptions of COVID-19 by Sociodemographic Characteristics

Participants who were male, were middle-aged (we found the highest knowledge score among the age group 40-49 years), were living in an urban area, and had a higher annual household income were more likely to answer knowledge questions correctly (Table 3). On average, participants living in urban areas answered an additional 1.33 (95% CI 1.22-1.44) questions correctly compared to participants living in rural areas. Men answered an additional 0.42 (95% CI 0.32-0.52) questions correctly compared to women.

Although there was relatively little variation in knowledge between provinces for prevention methods, common misconceptions, and the main mode of SARS-CoV-2 transmission (Figure 2), a higher proportion of participants in eastern coastal provinces answered the question on recommended care-seeking behavior correctly than in western inland provinces. The range in the proportion of correct responses for this question varied from 47.0% (95% CI 41.4%-52.7%) in Tibet to 87.5% (95% CI 84.1%-91.0%) in Beijing.

Table 3. Variation in the overall knowledge score by sociodemographic characteristics.[a]

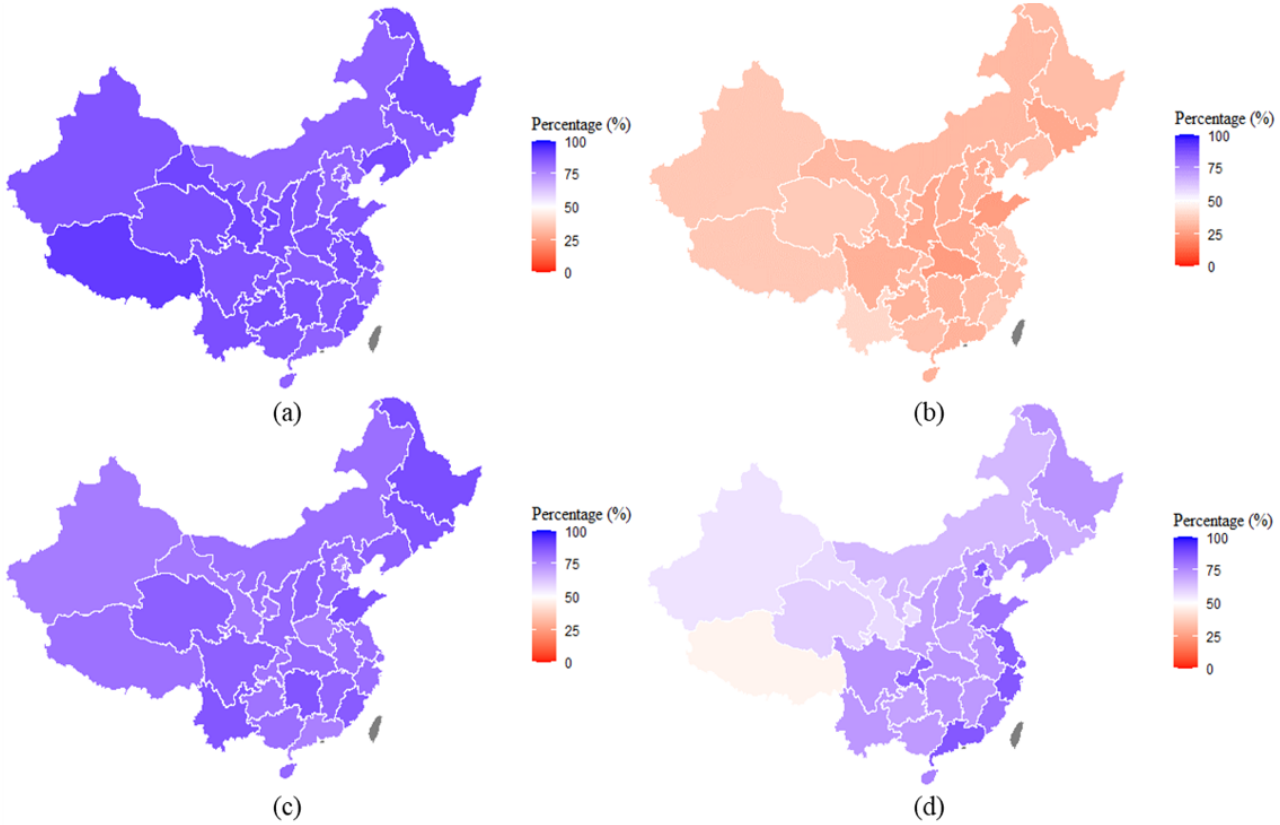
Characteristic	Absolute difference in the number of questions that were answered correctly (95% CI)	*P* value
Sex		
Male	0 (reference)	N/A[b]
Female	–0.42 (–0.52 to –0.32)	<.001
Age (years)		
18-19	0 (reference)	N/A
20-29	0.68 (0.46 to 0.9)	<.001
30-39	1.03 (0.8 to 1.25)	<.001
40-49	1.17 (0.95 to 1.38)	<.001
50-59	0.68 (0.46 to 0.9)	<.001
>60	0.75 (0.53 to 0.97)	<.001
Education		
Never been to school	0 (reference)	N/A
Elementary school	–0.18 (–0.53 to 0.16)	.30
Middle school	–0.36 (–0.63 to –0.08)	.01
High school/technical secondary school	–0.47 (–0.72 to –0.21)	<.001
College/undergraduate	0.20 (–0.05 to 0.46)	.12
Graduate and above	–0.11 (–0.52 to 0.3)	.60
Place of residence		
Rural	0 (reference)	N/A
Urban	1.33 (1.22 to 1.44)	<.001
Works as a health care provider		
No	0 (reference)	N/A
Nurse	–0.34 (–1.12 to 0.45)	.40
Physician	–0.59 (–1.48 to 0.3)	.20
Community health worker	–1.13 (–1.75 to –0.51)	<.001
Pharmacist	0.06 (–1.44 to 1.57)	.93
Other health care provider	–0.94 (–1.58 to –0.29)	.004
Annual household income, ¥ (US $)		
<30,000 (3835)	0 (reference)	N/A
30,000-59,999 (3835-7670)	0.76 (0.45 to 1.08)	<.001
60,000-89,999 (7670-11,505)	0.92 (0.62 to 1.22)	<.001
90,000-119,999 (11,506-15,341)	1.27 (0.97 to 1.57)	<.001
120,000-149,999 (15,341-19,176)	1.59 (1.27 to 1.9)	<.001
150,000-199,999 (19,175-25,568)	1.61 (1.29 to 1.93)	<.001
≥200,000 (25,568)	1.56 (1.21 to 1.92)	<.001
Knows someone with a confirmed SARS-CoV-2 infection		
No	0 (reference)	N/A
Self	2.09 (–3.04 to 7.23)	.43
Family member	–2.66 (–9.26 to 3.93)	.43
Neighbor	–0.76 (–2.76 to 1.24)	.45
Coworker	0.97 (–0.56 to 2.51)	.21

Characteristic	Absolute difference in the number of questions that were answered correctly (95% CI)	*P* value
Friend	–2.55 (–6.7 to 1.59)	.23

[a]All regressions included only one of the variables (sex; age group; education; place of residence; income; vocation; whether or not a participant has a family member, friend, or acquaintance who they know to have been infected with SARS-CoV-2) shown in the table and a binary indicator for each province (province-level fixed effects).

[b]N/A: not applicable.

Figure 2. The proportion of the population by province with correct responses to questions about (a) prevention methods, (b) common misconceptions, (c) transmission channels, and (d) recommended actions after infection.



Discussion

On average, participants in our survey answered 21.4 (95% CI 21.3-21.4) out of 25 questions correctly. Higher knowledge scores were associated with being middle-aged, higher household income, male sex, and living in urban areas. Knowledge about prevention methods, common misconceptions, and transmission modes of SARS-CoV-2 infection did not vary markedly across provinces. However, knowledge about appropriate measures upon the appearance of suspicious symptoms showed a decreasing trend from the eastern wealthier coastal provinces to the western less wealthy inland provinces.

A significant proportion of participants in all provinces held misconceptions about prevention methods and recommended health care–seeking behaviors, such as beliefs that regular nasal saline rinses or gargling mouthwash are effective at preventing a SARS-CoV-2 infection. Moreover, 10.0% (95% CI 9.4%-10.6%) of participants reported that, if they had symptoms suggestive of a possible SARS-CoV-2 infection, they would continue with their daily routine. It is important for these false and potentially dangerous beliefs to be addressed in public health campaigns. These findings could inform campaigns focused on the dissemination of misinformation on the internet and on social media, both of which have been identified as platforms to circulate such misinformation [19]. It may also be useful to inform Chinese health care professionals about common misconceptions held by the public so that these misconceptions can be addressed in consultations.

Although the majority of participants demonstrated a high proportion of correct responses to knowledge questions, there was nonetheless important variation in knowledge by knowledge domain. Questions that assessed knowledge of who is at increased risk of severe COVID-19 and common symptoms of SARS-CoV-2 infection were answered correctly over 90% of the time. Questions about how SARS-CoV-2 is primarily transmitted and what precautions are effective against transmission had correct response rates between 81% and 84%. The areas of lowest knowledge were identifying ineffective measures for, or misconceptions about, prevention of SARS-CoV-2 transmission and recommended care-seeking behavior when developing a fever or new persistent cough. This suggests that public health campaigns in China can most efficiently improve the public's knowledge about COVID-19 by dispelling misconceptions about ineffective prevention

measures and developing clear, consistent, and widely distributed instructions on what individuals should do if they develop symptoms suggestive of COVID-19.

Our study results are comparable with those of studies from other countries that assessed knowledge and perceptions of COVID-19. In comparison to a similar survey in the United States and the United Kingdom [9], the majority of the results were similar across the three countries. However, we also observed differences in the beliefs about the effectiveness of masks between these countries. Nearly nine-tenths of Chinese participants reported to believe wearing a face mask was effective in preventing a SARS-CoV-2 infection, whereas only about 40% and 30% of the US and UK participants, respectively, expressed this belief, suggesting that Chinese participants were more likely to accept and select masks as a prevention measure against infections. This comparison is limited, however, as it is important to note that beliefs in the effectiveness of mask wearing in the United States and the United Kingdom have likely increased since that survey was conducted in February-March 2020. In Australia, an online survey undertaken between March 18-24, 2020, found that understanding and adoption of hygiene-related behaviors was high [20]. Surveys conducted in Nepal, Egypt, and Malaysia found that participants had good general knowledge about COVID-19, including modes of transmission and recommended prevention measures [21]. However, several misconceptions were common, such as avoidance of eating poultry and meat to reduce the chance of SARS-CoV-2 infection or that wearing masks in public does not reduce infections [21-23]. The prevalence of misinformation and misconceptions about COVID-19 worldwide underscores our recommendation that public health authorities should continue campaigns that dispel widespread misinformation and focus on consistent messaging about the importance of following a combination of evidence-based infection prevention methods and behaviors.

A key strength of our study is that we sampled a large number of individuals across China's provinces, allowing us to assess how knowledge and perceptions vary across regions, and specifically to compare perceptions in provinces that have been less impacted by COVID-19 to those in more impacted areas like Hubei, the most severely affected province in China so far.

This study, however, also has several limitations. First, participants were selected from a pool of adults who were registered with KuRunData and who agreed to take the survey. It is possible that these individuals differ in important ways from the general population of China, as ability and willingness to register with an online survey company and participate in this survey might be related to literacy, education level, and perspectives on or experience with COVID-19, to name a few. Second, an important demographic limitation is the relatively small number of older adults in our sample. Although we used sampling weights to adjust for this pattern, our estimate for the oldest age groups had a small sample size and thus a large degree of uncertainty. Third, as answers were self-reported responses to hypothetical situations, they might not reflect what individuals would do in reality. Similarly, social desirability bias could have influenced participants to answer in ways that they thought they should, rather than how they truly felt. Finally, it must be noted that this survey was administered between May 8 and June 8, 2020, and consequently reflects knowledge and perceptions at that time.

In conclusion, although general knowledge about COVID-19 was comparatively high across China, a substantial proportion of the population was found to hold important misconceptions about some prevention methods and recommended health care–seeking behaviors. Chinese policy makers should address this misinformation, as it will be important that people have accurate knowledge and practice correct prevention measures to avoid a resurgence of infections.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary material.
[DOCX File , 121 KB - jmir_v23i4e26940_app1.docx]

References

1. Wang C, Horby PW, Hayden FG, Gao GF. A novel coronavirus outbreak of global health concern. Lancet 2020 Feb 15;395(10223):470-473 [FREE Full text] [doi: 10.1016/S0140-6736(20)30185-9] [Medline: 31986257]
2. Roberton T, Carter ED, Chou VB, Stegmuller AR, Jackson BD, Tam Y, et al. Early estimates of the indirect effects of the COVID-19 pandemic on maternal and child mortality in low-income and middle-income countries: a modelling study. Lancet Glob Health 2020 Jul;8(7):e901-e908 [FREE Full text] [doi: 10.1016/S2214-109X(20)30229-1] [Medline: 32405459]
3. Cutler DM, Summers LH. The COVID-19 pandemic and the $16 trillion virus. JAMA 2020 Oct 20;324(15):1495-1496 [FREE Full text] [doi: 10.1001/jama.2020.19759] [Medline: 33044484]
4. Do D, Sarker M, Chen S, Lenjani A, Tikka P, Bärnighausen T, et al. Healthcare worker attendance during the early stages of the COVID-19 pandemic: a longitudinal analysis of fingerprint-verified data from all public-sector secondary and tertiary care facilities in Bangladesh. J Glob Health 2020 Dec;10(2):020509. [doi: 10.7189/jogh.10.020509] [Medline: 33110592]

5. Chen S, Yang J, Yang W, Wang C, Bärnighausen T. COVID-19 control in China during mass population movements at New Year. Lancet 2020 Mar 07;395(10226):764-766 [FREE Full text] [doi: 10.1016/S0140-6736(20)30421-9] [Medline: 32105609]
6. Chen S, Zhang Z, Yang J, Wang J, Zhai X, Bärnighausen T, et al. Fangcang shelter hospitals: a novel concept for responding to public health emergencies. Lancet 2020 Apr 18;395(10232):1305-1314 [FREE Full text] [doi: 10.1016/S0140-6736(20)30744-3] [Medline: 32247320]
7. Chen S, Chen Q, Yang W, Xue L, Liu Y, Yang J, et al. Buying time for an effective epidemic response: the impact of a public holiday for outbreak control on COVID-19 epidemic spread. Engineering (Beijing) 2020 Oct;6(10):1108-1114. [doi: 10.1016/j.eng.2020.07.018] [Medline: 32983582]
8. Chen S, Jin Z, Bloom DE. Act early to prevent infections and save lives: causal impact of diagnostic efficiency on the COVID-19 pandemic. SSRN. 2020 Sep. URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3704132#references-widget [accessed 2020-10-05]
9. Geldsetzer P. Use of rapid online surveys to assess people's perceptions during infectious disease outbreaks: a cross-sectional survey on COVID-19. J Med Internet Res 2020 Apr 02;22(4):e18790 [FREE Full text] [doi: 10.2196/18790] [Medline: 32240094]
10. Dickens BL, Koo JR, Lim JT, Park M, Quaye S, Sun H, et al. Modelling lockdown and exit strategies for COVID-19 in Singapore. Lancet Reg Health 2020 Aug;1:100004. [doi: 10.1016/j.lanwpc.2020.100004]
11. Chen S, Chen Q, Yang J, Lin L, Li L, Jiao L, et al. Curbing the COVID-19 pandemic with facility-based isolation of mild cases: a mathematical modeling study. J Travel Med 2021 Feb 23;28(2):taaa226 [FREE Full text] [doi: 10.1093/jtm/taaa226] [Medline: 33274387]
12. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. Lancet Infect Dis 2020 May;20(5):533-534 [FREE Full text] [doi: 10.1016/S1473-3099(20)30120-1] [Medline: 32087114]
13. Lee T, Zhong Y, Zhou J, He X, Kong R, Ji J. The outbreak of coronavirus disease in China: risk perceptions, knowledge, and information sources among prenatal and postnatal women. Women Birth 2020 May 29:1 [FREE Full text] [doi: 10.1016/j.wombi.2020.05.010] [Medline: 32534906]
14. Gao Y, Sun F, Jiang W, Fang Y, Yue L, Lin X, et al. Beliefs towards the COVID-19 pandemic among patients with emotional disorders in China. Gen Psychiatr 2020;33(3):e100231 [FREE Full text] [doi: 10.1136/gpsych-2020-100231] [Medline: 32574346]
15. Zhu H, Deng F. How to influence rural tourism intention by risk knowledge during COVID-19 containment in China: mediating role of risk perception and attitude. Int J Environ Res Public Health 2020 May 18;17(10):3514 [FREE Full text] [doi: 10.3390/ijerph17103514] [Medline: 32443430]
16. How it works. Kurundata. URL: https://www.kurundata.com/91/ [accessed 2021-04-17]
17. The urban-rural composition of the population by region and the birth rate, death rate, and natural growth rate (2018). China Statistical Yearbook. 2019. URL: http://www.stats.gov.cn/tjsj/ndsj/2019/indexch.htm [accessed 2020-10-05]
18. Coronavirus disease (COVID-19) advice for the public: mythbusters. World Health Organization. 2020. URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/advice-for-public/myth-busters [accessed 2020-10-05]
19. Ball P, Maxmen A. The epic battle against coronavirus misinformation and conspiracy theories. Nature 2020 May;581(7809):371-374. [doi: 10.1038/d41586-020-01452-z] [Medline: 32461658]
20. Seale H, Heywood AE, Leask J, Sheel M, Thomas S, Durrheim DN, et al. COVID-19 is rapidly changing: examining public perceptions and behaviors in response to this evolving pandemic. PLoS One 2020;15(6):e0235112 [FREE Full text] [doi: 10.1371/journal.pone.0235112] [Medline: 32574184]
21. Abdelhafiz AS, Mohammed Z, Ibrahim ME, Ziady HH, Alorabi M, Ayyad M, et al. Knowledge, perceptions, and attitude of Egyptians towards the novel coronavirus disease (COVID-19). J Community Health 2020 Oct;45(5):881-890 [FREE Full text] [doi: 10.1007/s10900-020-00827-7] [Medline: 32318986]
22. Singh DR, Sunuwar DR, Karki K, Ghimire S, Shrestha N. Knowledge and perception towards universal safety precautions during early phase of the COVID-19 outbreak in Nepal. J Community Health 2020 Dec;45(6):1116-1122 [FREE Full text] [doi: 10.1007/s10900-020-00839-3] [Medline: 32405905]
23. Azlan AA, Hamzah MR, Sern TJ, Ayub SH, Mohamad E. Public knowledge, attitudes and practices towards COVID-19: a cross-sectional study in Malaysia. PLoS One 2020;15(5):e0233668 [FREE Full text] [doi: 10.1371/journal.pone.0233668] [Medline: 32437434]

Edited by R Kukafka, G Eysenbach; submitted 04.01.21; peer-reviewed by Z Jin; comments to author 05.02.21; revised version received 16.02.21; accepted 11.04.21; published 29.04.21.

Please cite as:
Yu F, Geldsetzer P, Meierkord A, Yang J, Chen Q, Jiao L, Abou-Arraj NE, Pan A, Wang C, Bärnighausen T, Chen S
Knowledge About COVID-19 Among Adults in China: Cross-sectional Online Survey
J Med Internet Res 2021;23(4):e26940
URL: https://www.jmir.org/2021/4/e26940
doi:10.2196/26940
PMID:33844637

Original Paper

Evaluation of an Intrahospital Telemedicine Program for Patients Admitted With COVID-19: Mixed Methods Study

Sean Legler[1], MD; Matthew Diehl[1], MD; Brian Hilliard[1], MD; Andrew Olson[1], MD; Rebecca Markowitz[1], MD; Christopher Tignanelli[2], MD; Genevieve B Melton[2], MD, PhD; Alain Broccard[1], MD; Jonathan Kirsch[1], MD; Michael Usher[1], MD, PhD

[1]Division of General Internal Medicine, Department of Medicine, University of Minnesota Medical School, Minneapolis, MN, United States
[2]Department of Surgery, University of Minnesota Medical School, Minneapolis, MN, United States

Corresponding Author:
Michael Usher, MD, PhD
Division of General Internal Medicine
Department of Medicine
University of Minnesota Medical School
420 Delaware St SE, MMC 741
Minneapolis, MN, 55455
United States
Phone: 1 734 223 3450
Email: mgusher@umn.edu

Abstract

Background: The increasing incidence of COVID-19 infection has challenged health care systems to increase capacity while conserving personal protective equipment (PPE) supplies and minimizing nosocomial spread. Telemedicine shows promise to address these challenges but lacks comprehensive evaluation in the inpatient environment.

Objective: The aim of this study is to evaluate an intrahospital telemedicine program (virtual care), along with its impact on exposure risk and communication.

Methods: We conducted a natural experiment of virtual care on patients admitted for COVID-19. The primary exposure variable was documented use of virtual care. Patient characteristics, PPE use rates, and their association with virtual care use were assessed. In parallel, we conducted surveys with patients and clinicians to capture satisfaction with virtual care along the domains of communication, medical treatment, and exposure risk.

Results: Of 137 total patients in our primary analysis, 43 patients used virtual care. In total, there were 82 inpatient days of use and 401 inpatient days without use. Hospital utilization and illness severity were similar in patients who opted in versus opted out. Virtual care was associated with a significant reduction in PPE use and physical exam rate. Surveys of 41 patients and clinicians showed high rates of recommendation for further use, and subjective improvements in communication. However, providers and patients expressed limitations in usability, medical assessment, and empathetic communication.

Conclusions: In this pilot natural experiment, only a subset of patients used inpatient virtual care. When used, virtual care was associated with reductions in PPE use, reductions in exposure risk, and patient and provider satisfaction.

(J Med Internet Res 2021;23(4):e25987) doi:10.2196/25987

KEYWORDS

telemedicine; hospital medicine; COVID-19; telehealth; hospital; mixed methods; evaluation; impact; exposure; risk; communication

Introduction

The COVID-19 pandemic presents unprecedented challenges to infection control in hospitals worldwide. These challenges are particularly urgent given the potential exposure risk for both health care personnel and uninfected patients amid shortages of personal protective equipment (PPE).

The health care–associated spread of COVID-19 to patients and health care workers is well documented. In Wuhan, China, presumed hospital-related transmission of COVID-19 was suspected in 41% of patients [1]. In Italy, 20% of health care professionals in one health system became infected with COVID-19 [2]. In Washington state, lack of sufficient PPE use in a skilled nursing facility is thought to have contributed to the

spread of infection to 101 residents, 34 staff members, and 14 visitors [3]. As of September 2, 2020, nearly 570,000 health care providers in the Pan-American region have been infected with COVID-19 and more than 2500 have died [4]. Although health care providers may be infected outside their job duties, health care–associated COVID-19 infections are potentially avoidable, cause direct morbidity and mortality, and lead to a reduction in workforce capacity as many health care workers are quarantined. Addressing these risks, the World Health Organization and Centers for Disease Control and Prevention both recognize PPE as one of the most effective preventive measures to reduce transmission of COVID-19. Unfortunately, there is currently a global shortage due to a surge in demand and disruptions to the global supply chain [5,6]. Although strategies have been developed to mitigate these shortages and infection risks, they will continue to present challenges as the pandemic continues.

Telemedicine has been proposed as a potential innovative solution to reduce the infection risk associated with COVID-19 [7-9]. Telemedicine, including delivery in a hospital-based setting, has the capacity to facilitate communication and a visual examination of a patient without entering the room. This in turn has the potential to decrease PPE use and health care–associated COVID-19 transmission. More than 50 US health systems already have existing telemedicine programs [7], which have been used in ambulatory [10], triage [11], inpatient [12], and intensive care unit settings [13]. These programs are associated with high patient and provider satisfaction and noninferior clinical outcomes [12,14,15]. Furthermore, telemedicine has been shown to be feasible in the context of the COVID-19 pandemic [16-19]. To date, however, telemedicine in the hospital-based setting has typically centered on using this technology to increase coverage to critical access hospitals—often for specialist services not available locally (eg, intensivists)—or for consultants as opposed to a more ubiquitous care modality for inpatient care.

To our knowledge, telemedicine in the hospital has not been extensively used and studied with the explicit purpose of reducing infection exposure and PPE use. If intrahospital telemedicine (virtual care) is an option, providers and patients have a choice on whether to conduct an in-person evaluation or virtual evaluation for a given interaction. This creates a unique environment to test the impact of a novel use for telemedicine technology on communication, in addition to evaluating its impact on exposure risk and PPE use and patient experience.

We built an in-house virtual care system at a large academic medical center in the midwestern United States with a dedicated COVID-19 service. The primary objective of this study was to evaluate if virtual care for hospitalized patients with confirmed COVID-19 could reduce overall exposures to patients with COVID-19. Secondary outcomes included patient and provider satisfaction.

Methods

The study was a prospective cohort study conducted in three general medicine units at an urban university hospital solely dedicated to treating patients with a confirmed positive test for COVID-19. The chart review portion of the study was approved by the University of Minnesota Institutional Review Board.

Patients

All patients were confirmed to have a positive polymerase chain reaction test result indicating infection with COVID-19 prior to transfer to the COVID-19–specific hospital to one of three medicine units. Only adult patients over 18 years of age were included in the evaluation.

Intervention

We conducted a natural experiment of an intrahospital telemedicine program (termed virtual care) on three medical units in a dedicated COVID-19 hospital. All rooms were equipped with an iPad (Apple Inc) with third-party videoconferencing software (Polycom) assigned to the patient's room. Patients and providers were given the option to opt out of the use of this virtual technology on a daily basis if deemed inappropriate based on mental status, illness severity, technical issues, or individual preference. If accepted for use, providers and nurses could conference into the patient's room using a device at the nursing station or an enterprise-associated phone or tablet. The virtual evaluation process required the patient to opt in by either leaving a virtual room open or turning on the camera and microphone. We present an analysis of a convenience sample of consecutively collected patients during a 6-week evaluation period.

Measures

Age, race, gender, interpreter use, and comorbidities were extracted from the electronic health record (EHR). Comorbidities were characterized by the Elixhauser Comorbidity Index using diagnosis codes collected in the year prior to hospitalization [20]. In addition, iPad use was reported in a standardized EHR note template, which was tracked on a daily basis prospectively. We measured PPE use on even number days by extracting daily PPE usage logs that were placed on each patient's door and filled out by all staff entering the room. Admission days, discharge days, and days where the patient was transferred between units were excluded. We specifically report face shield, mask, and gown use rates. We also report PPE breach rates, which were recorded on the same sheet. Finally, we report physical exam rates, which is the number of in-person exams conducted by physicians and nurses documented in the chart or sign-in sheet.

Survey Deployment

We developed a multidisciplinary workgroup to assess patient, nurse, and physician satisfaction. The goals were to assess comfort with the technology, usefulness in supporting communication, disruptiveness, and perceptions of safety using a modified Likert scale. Questions were designed to adhere loosely to the Technology Acceptance Model [21], assessing perceived usefulness, perceived ease of use, and attitude toward use. Clinicians serving dedicated COVID-19 units were contacted with both paper and electronic surveys. Patients were contacted virtually via the iPad or by phone if they had at least one documented use of virtual care and were felt to be close to discharge by the patient's medical team. Surveys were continually applied to patients as they approached discharge,

and were distributed to providers and nurses at week 3 and week 6. At week 6, an additional survey was distributed to providers and nurses to address usability and capture reasons inpatient virtual care was not used.

Statistical Analysis

Continuous variables are displayed as median and IQR for skewed variables or otherwise mean and SD and binary or categorical variables as a count and percentage. An ordinal variable was generated related to survey responses as measured by 5-point Likert scale, with 0 being strongly disagree and 4 being strongly agree. Mean and SD for each group (patients, clinicians) were calculated. Between-group comparisons between patients who opted in and opted out were performed using Mann-Whitney and chi-square tests where appropriate.

To adjust for repeated measures, daily PPE use rates were converted into panel data. Gown, face shield, and mask use were evaluated using Poisson regression adjusting for clustering by patient. Physical exam rate was evaluated by negative binomial regression. Two comparisons were performed. First, we evaluated the impact of virtual care by evaluating per-patient daily PPE use against all control patient days. Since patients were not reported to use virtual care every day, we performed a subsequent sensitivity analysis comparing PPE use on days where virtual care was used against days it was not, among a cohort that used it at least once. All statistical analysis was completed in STATA (version 16; StataCorp LLC).

Results

Overview

A convenience sample of 137 patients admitted to the COVID-19 inpatient unit was evaluated, accounting for a total of 483 patient days during the 6-week evaluation period. This included 43 patients who used virtual care at least once (Table 1). On average, patients who opt out versus opt in to virtual care are not statistically different with respect to age (mean 66.6 years versus 65.1 years, P=.62), gender (46% male versus 49% male, P=.79), racial make-up (62% White versus 49% White, P=.15), or comorbidity rates. Although proportionally more White patients opted out of virtual care than Black or Asian patients, there was not a statistically significant difference. Use of interpreters was lower in the opt-in population (12% versus 20%, P=.09), but this did not reach statistical significance. On average, those patients who opted in had a longer length of stay (11.8 days versus 9.9 days, P=.09) but this was also not statistically significant. Similarly, there were no significant differences in mortality in those who opted in versus opted out (7% versus 10%, P=.62) or maximum oxygen requirements (maximum oxygen requirement 4.5 liters per minute versus 6 liters per minute, P=.68).

Table 1. Patient demographics, comorbidities, and hospital utilization among patients with and without virtual care.

Variables	Opt-out group (n=94)	Opt-in group (n=43)	*P* value
Demographics			
Age years, mean (SD)	66.6 (18)	65.1 (19)	.62
Male, n (%)	46 (49)	20 (46)	.79
White, n (%)	58 (62)	21 (49)	.16
Black, n (%)	14 (15)	8 (19)	.58
Asian, n (%)	10 (11)	4 (9)	.81
Other/missing, n (%)	12 (13)	10 (23)	.12
Comorbidities			
Congestive heart failure, n (%)	26 (28)	7 (16)	.08
Chronic obstructive pulmonary disease, n (%)	15 (16)	5 (12)	.50
Obesity, n (%)	26 (28)	16.9 (37)	.26
Depression, n (%)	27 (29)	16 (37)	.32
Hypertension, n (%)	68 (72)	30 (70)	.76
Diabetes mellitus, n (%)	29 (31)	17 (40)	.32
Elixhauser Comorbidity Index score, median (IQR)	5 (3-8)	5 (3-7)	.49
Hospital utilization			
Maximum oxygen requirement (liters per minute), median (IQR)	6 (3-10)	4.5 (3-7)	.68
Discharged, n (%)	70 (74)	31 (72)	.77
Length of stay (days, if discharged), median (IQR)	9.9 (6.5-14.9)	11.8 (7.0-24.9)	.09
In-hospital mortality, n (%)	9 (10)	3 (7)	.62
Interpreter use, n (%)	19 (20)	5 (12)	.09

Use of PPE

Distribution of daily PPE use is displayed in Figure S1 in Multimedia Appendix 1. As shown in Table 2, daily gown use (median 5, IQR 4-5 versus median 3, IQR 2-3; *P*<.001), face shield use (median 5, IQR 4-5 versus median 3, IQR 2-3; *P*<.001, *P*=.001), and mask use (median 5, IQR 4-5 versus median 3, IQR 2-3; *P*<.001) were all reduced on days where virtual care was reported to be used. Similarly, physical exam rates (median 1, IQR 1-2 versus median 0, IQR 0-1; *P*<.001) were reduced on days where virtual care was used. There was no significant difference in the recorded PPE breach rate; breaches were rare for both groups. Similar differences were found when comparing days where virtual care was used and not used among patients who opted in (Table 2).

Table 2. Daily personal protective equipment use rates among patients with and without virtual care.

Variables	Opt-out group	Opt-in group	Coefficient (95% CI)	*P* value
All patients				
Number of days	401	82	N/A[a]	N/A
Face shields, median (IQR)	5 (4-6)	3 (2-3)	–0.54 (–0.65 to –0.43)[b]	<.001
Masks, median (IQR)	5 (4-6)	3 (2-3)	–0.54 (–0.65 to –0.43)[b]	<.001
Gowns, median (IQR)	5 (5-6)	3 (2-3)	–0.54 (-0.65 to –0.43)[b]	<.001
Physical exam, median (IQR)	1 (1-2)	0 (0-1)	–0.84 (–1.12 to –0.57)[c]	.001
Personal protective equipment breach, n (%)	4 (1)	3 (4)	1.65 (0.33 to 8.30)[d]	.55
Patients who opted in at least once				
Number of days	116	82	N/A	N/A
Face shields, median (IQR)	5 (4-6)	3 (2-3)	–0.54 (–0.69 to –0.41)[b]	<.001
Masks, median (IQR)	5 (4-6)	3 (2-3)	–0.54 (–0.69 to –0.41)[b]	<.001
Gowns, median (IQR)	5 (5-6)	3 (2-3)	–0.54 (–0.69 to –0.41)[b]	<.001
Physical exam, median (IQR)	1 (1-2)	0 (0-1)	–0.85 (–1.31 to –0.56)[c]	.001
Personal protective equipment breach, n (%)	2 (2)	5 (6)	1.65 (0.33 to 8.30)[d]	.55

[a]N/A: not applicable.

[b]Poisson regression adjusting for clustering by patient.

[c]Negative binomial regression adjusting for clustering by patient.

[d]Logistic regression adjusting for clustering by patient.

Patient Experience

We then investigated patient attitudes and perceptions on use and usability. The overall response rate was 40% (42/105), and 65% (27/41) for providers. Among patients surveyed, 8 of 15 (53%) reported common use of videoconferencing and 5 (33%) reported not using it at all. Overall, the patients' perceptions of virtual care were positive (Table 3). Furthermore, 8 (53%) respondents reported virtual care improved communication with the care team; however, only 7 (46%) felt emotionally supported through virtual communication. Patients reported improvements in sense of isolation with use (11/15, 73%), felt that it reduced exposure risk (n=15, 100%), and felt that overall continuing the project was a good idea (n=14, 92%). Importantly, patients largely felt that providers remained responsive to patient needs (n=14, 92%), and 10 (67%) disagreed with the idea that virtual care would cause doctors to miss something.

Table 3. Patient satisfaction: means and distributions of survey responses of patients with COVID-19.

Survey questions	Mean (SD)[a]
I use Facetime or other video chat programs in my daily life	2.13 (1.50)
The purpose and use of the iPad was explained to me	2.53 (1.12)
My doctor and nurses primarily communicated with me virtually	1.73 (1.10)
Virtual visits improved my ability to communicate with my care team	2.73 (0.96)
I felt that use of virtual communication reduced the risk of exposing my medical team to COVID-19	3.60 (0.51)
My care team was able to emotionally support me through virtual communication	2.27 (0.96)
Being able to use virtual communication allowed me to feel less isolated	2.93 (0.80)
My care team was responsive to my needs including visiting me in person when I needed it	3.67 (0.62)
I feel that my care team was more likely to miss something when relying on virtual communication	1.33 (0.82)
I think continuing to use virtual visits in the hospital is a good idea	3.13 (0.52)

[a]Modified Likert scale was converted to a 0-4 scale, with 0 denoting strongly disagree and 4 denoting strongly agree. Scores above 2 are significantly more positive.

Health Care Worker Experience

Results of the provider survey are shown in Table 4. Two areas identified as potential barriers for provider acceptance of this technology were lack of prior use of telemedicine with patients (Likert mean 1.65, SD 1.44) and concern that virtual care does not allow for adequate assessment of patient disease severity (Likert mean 1.92, SD 1.02; responses less than 2 indicated disagreement). Benefits of virtual care included reduced exposure risk and PPE use for providers (Likert mean 3.19, SD 0.85) and interest in expansion of virtual care visits (Likert mean 3.09, SD 0.7). Providers on average were less confident in virtual care's ability to allow effective evaluation of disease severity (n=7, 26% responding positively) and empathetic communication (n=11, 41% responding positively).

Table 4. Provider satisfaction: means and distribution of surveys to health care workers caring for patients with COVID-19.

Survey questions	Mean (SD)[a]
I am familiar with telemedicine technology	2.69 (0.93)
I had personally used telemedicine with a patient prior to this initiative	1.65 (1.44)
Using the iPad/remote assessment improves my job effectiveness and performance	2.38 (1.06)
Using the iPad/remote assessment makes me feel safer	2.65 (0.98)
Using the iPad/remote assessment reduced my exposure risk and use of personal protective equipment	3.19 (0.85)
Using the iPad/remote assessment is disruptive to my normal workflow	1.84 (1.00)
Using the iPad/remote assessment improves my ability to communicate with patients under isolation	2.69 (0.97)
Using the iPad/remote assessment allows for adequate assessment of patient disease severity	1.92 (1.02)
Using the iPad/remote assessment allowed me to effectively show empathy to patients under isolation	2.15 (1.12)
iPad/Remote assessment should be expanded to other patients with COVID-19	2.84 (0.96)

[a]Modified Likert scale was converted to a 0-4 scale, with 0 denoting strongly disagree and 4 denoting strongly agree.

Finally, we investigated barriers among providers who were not successful in using virtual care (Table S1 in Multimedia Appendix 1). Sentiment for virtual health remained positive in regard to improving communication, family involvement, and reducing exposure risk (Table S2 in Multimedia Appendix 1) even among providers who reported difficulty with use. Respondents reported language (n=7, 58% reported frequent occurrence) and usability issues (n=8, 67% reported frequent occurrence) were the most common barriers to use. The virtual care also did not appear to detract significantly from providers' efficiency; most disagreed with the statement that the iPad/remote assessment was disruptive to normal workflow (Likert mean 1.84, SD 1.0). Patient and provider preference as a barrier to use was also rare (Table S1 in Multimedia Appendix 1).

Discussion

Principal Findings

In the era of COVID-19, hospitals have necessarily become testing grounds for innovation. Inpatient wards under duress need to adapt to patient surges, minimize health care worker exposure, prevent nosocomial transmission, and streamline patient flow. Despite these pressures, however, the human core of medical care—effective communication, empathy, and clinical expertise—should not be sacrificed. We took a mixed

methods approach to evaluate the degree to which an inpatient telemedicine program (virtual care) could reduce COVID-19 exposure risk and PPE use while maintaining effective communication. Although telehealth programs have rapidly expanded in the context of COVID-19, direct evaluation of patient and provider experience and overall utility remain understudied. This is particularly true for inpatients where choice between a virtual encounter and in-person encounter exists continuously. We provide several contributions to the literature that have important implications.

We provide a structured evaluation of virtual care for patients with COVID-19 and find important limitations. Patients and providers had the ability to opt out of use based on usability, preference, and medical appropriateness. As such, it was used a minority of the time even when it was available for all patients. Even among patients who used virtual care, it was relied upon less than half the days. Providers found use was of limited utility for risk assessment and providing empathetic communication.

Concerns regarding empathetic communication echo the findings that the digitization of health care can lead to a corresponding decrease in the expression of empathy by providers [22]. For example, in at least one study, empathy and praise utterances were observed less in telemedicine consultations compared to face-to-face consultations [23]. This is of particular importance because empathy is a fundamental determinant of quality in medical care [24]. Techniques have been suggested to overcome these barriers during the COVID-19 pandemic, such as finding a private place for videoconferencing, looking directly at the camera, ensuring proper lighting, and paying close attention to subtle comments and body language [16,25]. However, it is not yet clear whether these or other measures can fully bridge the digital empathy gap.

These observations highlight the importance of patient and provider engagement in the implementation of new technology, especially during a crisis. In this natural study, implementation was guided by efficiency to address a rapidly emerging health crisis. Development of a more patient-centered telemedicine program could build upon this work, improving accessibility by integrating translation services, addressing the needs of those who are hard of hearing or have vision loss, and integrating other health services to promote multidisciplinary virtual rounding. Without a patient-centered approach, there is also a risk that the digital health divide could exacerbate health disparities even within the confines of the hospital [26]. It is estimated that in the United States, not only may 13 million older adults have trouble accessing telemedical services during the COVID-19 pandemic, but also a disproportionate number of those may be among the already disadvantaged [27]. This study thus provides a first step toward understanding what is needed for a more patient-centered system.

Despite these limitations, providers and patients were generally positive regarding their experience with inpatient virtual care in two domains: supporting patient-provider communication and reducing risk of exposure. Both patients and providers supported the value of a virtual rounding program for COVID-19, with all patients and 73% (19/26) providers surveyed agreeing with program continuance. This is consistent with prior literature, which has found adequate patient satisfaction with telemedicine interventions, including during the COVID-19 pandemic [9,11,15,16,25,28,29].

Importantly, virtual care was also associated with significantly reduced PPE usage and potential exposure. Given critical shortfalls in PPE supply and distribution, efficient management of “burn rate” is an important factor in maintaining safety in our health care system [5,6]. This observation was corroborated by survey data in which both providers and patients agree that virtual care reduced exposure and PPE use. Understanding predictors of use and identifying factors that can improve efficiency of PPE use will be an important future area of study in adapting to the COVID-19 pandemic [30]. It is not known, however, whether the reduced exposure alone led to a subsequent decrease in health care–associated COVID-19 transmission because health care–associated infections were not reported as part of this study.

There are several further limitations to this study as a single-center natural experiment on the use of inpatient telemedicine. First, the sample size is small and as such we may not be able to capture the full variation in PPE use pattern or patient experience. Future work with larger sample sizes should elucidate in which ways patient satisfaction varies based on demographic groups. Second, this was not a randomized trial and there are numerous possible sources for bias. Associations can be confounded by unobserved factors such as policy changes in PPE use (such as approaches to reuse) and patient characteristics. One alternative explanation for our findings, for example, is that the differences in PPE use could be related to changes in illness severity during times in which inpatient telemedicine was felt to not be medically appropriate. Although a survey of clinicians who reported usability barriers suggested this was a minority of episodes, such a relationship could still be present. A critical next step will be understanding how PPE use rates correlate with patient factors and how they change temporally, while also determining how they correlate with outcomes.

Similarly, as this experiment used new technology, lack of use may be highly practitioner dependent and less dependent on patient-related factors. Thus, observed PPE differences and patient experiences could be attributed to the providers that tended to use virtual care more often. More structured evaluation of use to eliminate some of these inherent biases with our observational experience, such as a cluster randomized controlled trial with structured use protocols, would be better designed to overcome these shortfalls. Finally, we only surveyed patients who used virtual care at least once. The primary reason for this design was to get perspectives based on direct experience. As such, we did not obtain perceptions for patients who did not choose to use virtual care, resulting in an additional source of potential bias, and limitation to generalizability of our study findings.

Importantly, this program was built upon pre-existing technology available to our health care system, including iPads, an enterprise network, and Health Insurance Portability and Accountability Act–compliant videoconferencing technology. To other health care systems without existing telemedicine

infrastructure, this intervention may be more expensive and difficult to put into place. Furthermore, in the setting of this study, a direct in-person evaluation was always immediately available. In such circumstances, intrahospital telemedicine may not be directly comparable to a fully remote telemedicine model where in-person consultation is not possible.

Indeed, the psychological effects of virtual care possibilities in light of COVID-19 warrant further investigation. We found gaps in the ability to deliver empathetic care, but there were also anecdotal stories of success. For example, an additional use case that was used was the ability to hold remote family conversations for patients in isolation. Additionally, we had multiple family members hospitalized in separate rooms in this study who could stay continually connected through this technology, for which they expressed gratitude during our survey collection.

Ultimately, we suggest that an individualized, prudent selection process for the use of virtual care may offer an appropriate use model in response to COVID-19. This can be based on evolving and unique patient and provider characteristics. In times of clinical stability and when used in the right population, virtual care may enhance communication, reduce PPE use, and reduce exposure risk. However, during periods where diagnostic evaluation, direct communication, or empathy are paramount, in-person care remains preferable. The future of telemedicine must be individualized and patient centered.

Conclusions

In summary, our experience suggests that intrahospital telemedicine is feasible and well received. It can be used by a select subpopulation and effectively reduces infectious exposure and PPE usage, and can support communication, though with the caveat of a more limited capability for empathetic support and need to improve usability in a patient-centered manner.

Acknowledgments

Grant UL1TR002494 from the National Institutes of Health's National Center for Advancing Translational Sciences. Additional funding was obtained from the Agency for Healthcare Research and Quality (AHRQ HS26732-01).

Authors' Contributions

SL, MD, BH, AB, JK, and MU contributed to the concept and design. SL, BH, JK, and MU contributed to the acquisition, analysis, or interpretation of the data. SL, MD, BH, AO, RM, CT, GBM, AB, JK, and MU critically revised the manuscript for important intellectual content. MU performed the statistical analysis. MU, GBM, AO, and RM provided administrative, technical, or material support. MU provided supervision.

Conflicts of Interest

Author CT is supported by the Agency for Healthcare Research and Quality (AHRQ), Patient-Centered Outcomes Research Institute (PCORI), grant K12HS026379 and the National Institutes of Health's National Center for Advancing Translational Sciences, grant KL2TR002492. The content is solely the responsibility of the authors and does not necessarily represent the official views of AHRQ, PCORI, or Minnesota Learning Health System Mentored Career Development Program (MN-LHS). The other authors have no conflicts to declare.

Multimedia Appendix 1
Supplemental data.
[DOCX File , 57 KB - jmir_v23i4e25987_app1.docx]

References

1. Wang D, Hu B, Hu C, Zhu F, Liu X, Zhang J, et al. Clinical Characteristics of 138 Hospitalized Patients With 2019 Novel Coronavirus-Infected Pneumonia in Wuhan, China. JAMA 2020 Mar 17;323(11):1061-1069 [FREE Full text] [doi: 10.1001/jama.2020.1585] [Medline: 32031570]
2. Remuzzi A, Remuzzi G. COVID-19 and Italy: what next? The Lancet 2020 Apr;395(10231):1225-1228. [doi: 10.1016/s0140-6736(20)30627-9]
3. McMichael TM, Currie DW, Clark S, Pogosjans S, Kay M, Schwartz NG, et al. Epidemiology of Covid-19 in a Long-Term Care Facility in King County, Washington. N Engl J Med 2020 May 21;382(21):2005-2011. [doi: 10.1056/nejmoa2005412]
4. Pan American Health Organization. COVID-19 has infected some 570,000 health workers and killed 2,500 in the Americas, PAHO Director says. URL: https://www.paho.org/en/news/2-9-2020-covid-19-has-infected-some-570000-health-workers-and-killed-2500-americas-paho [accessed 2021-01-13]
5. World Health Organization. Rational use of personal protective equipment for coronavirus disease (COVID-19): interim guidance, 27 February 2020. URL: https://apps.who.int/iris/bitstream/handle/10665/331215/WHO-2019-nCov-IPCPPE_use-2020.1-eng.pdf [accessed 2021-04-23]

6. Centers for Disease Control and Prevention. Coronavirus Disease 2019 (COVID-19): Interim Infection Prevention and Control Recommendations. URL: https://www.cdc.gov/coronavirus/2019-ncov/infection-control/control-recommendations.html [accessed 2020-03-31]
7. Hollander JE, Carr BG. Virtually Perfect? Telemedicine for Covid-19. N Engl J Med 2020 Apr 30;382(18):1679-1681. [doi: 10.1056/nejmp2003539]
8. Greenhalgh T, Wherton J, Shaw S, Morrison C. Video consultations for covid-19. BMJ 2020 Mar 12;368:m998 [FREE Full text] [doi: 10.1136/bmj.m998] [Medline: 32165352]
9. Vidal-Alaball J, Acosta-Roja R, Pastor Hernández N, Sanchez Luque U, Morrison D, Narejos Pérez S, et al. Telemedicine in the face of the COVID-19 pandemic. Aten Primaria 2020 Jun;52(6):418-422 [FREE Full text] [doi: 10.1016/j.aprim.2020.04.003] [Medline: 32402477]
10. Downes MJ, Mervin MC, Byrnes JM, Scuffham PA. Telephone consultations for general practice: a systematic review. Syst Rev 2017 Jul 03;6(1):128 [FREE Full text] [doi: 10.1186/s13643-017-0529-0] [Medline: 28673333]
11. Boggan JC, Shoup JP, Whited JD, Van Voorhees E, Gordon AM, Rushton S, et al. Effectiveness of Acute Care Remote Triage Systems: a Systematic Review. J Gen Intern Med 2020 Jul 02;35(7):2136-2145. [doi: 10.1007/s11606-019-05585-4] [Medline: 31898116]
12. Kuperman E, Linson E, Klefstad K, Perry E, Glenn K. The Virtual Hospitalist: A Single-Site Implementation Bringing Hospitalist Coverage to Critical Access Hospitals. J Hosp Med 2018 Nov 01;13(11):759-763. [doi: 10.12788/jhm.3061] [Medline: 30255859]
13. Khunlertkit A, Carayon P. Contributions of tele-intensive care unit (Tele-ICU) technology to quality of care and patient safety. J Crit Care 2013 Jun;28(3):315.e1-315.12. [doi: 10.1016/j.jcrc.2012.10.005] [Medline: 23159139]
14. Ignatowicz A, Atherton H, Bernstein CJ, Bryce C, Court R, Sturt J, et al. Internet videoconferencing for patient-clinician consultations in long-term conditions: A review of reviews and applications in line with guidelines and recommendations. Digit Health 2019 Apr 23;5:2055207619845831 [FREE Full text] [doi: 10.1177/2055207619845831] [Medline: 31069105]
15. Backhaus A, Agha Z, Maglione ML, Repp A, Ross B, Zuest D, et al. Videoconferencing psychotherapy: a systematic review. Psychol Serv 2012 May;9(2):111-131. [doi: 10.1037/a0027924] [Medline: 22662727]
16. Calton B, Abedini N, Fratkin M. Telemedicine in the Time of Coronavirus. J Pain Symptom Manage 2020 Jul;60(1):e12-e14 [FREE Full text] [doi: 10.1016/j.jpainsymman.2020.03.019] [Medline: 32240756]
17. Chou E, Hsieh Y, Wolfshohl J, Green F, Bhakta T. Onsite telemedicine strategy for coronavirus (COVID-19) screening to limit exposure in ED. Emerg Med J 2020 Jun 03;37(6):335-337. [doi: 10.1136/emermed-2020-209645] [Medline: 32366616]
18. Loeb AE, Rao SS, Ficke JR, Morris CD, Riley LH, Levin AS. Departmental Experience and Lessons Learned With Accelerated Introduction of Telemedicine During the COVID-19 Crisis. J Am Acad Orthop Surg 2020;28(11):e469-e476. [doi: 10.5435/jaaos-d-20-00380]
19. Bryant MS, Fedson SE, Sharafkhaneh A. Using Telehealth Cardiopulmonary Rehabilitation during the COVID-19 Pandemic. J Med Syst 2020 May 28;44(7):1. [doi: 10.1007/s10916-020-01593-8]
20. van Walraven C, Austin P, Jennings A, Quan H, Forster A. A modification of the Elixhauser comorbidity measures into a point system for hospital death using administrative data. Med Care 2009 Jun;47(6):626-633. [doi: 10.1097/MLR.0b013e31819432e5] [Medline: 19433995]
21. Lee Y, Kozar K, Larsen K. The Technology Acceptance Model: Past, Present, and Future. CAIS 2003;12:1-31 [FREE Full text] [doi: 10.17705/1cais.01250]
22. Terry C, Cain J. The Emerging Issue of Digital Empathy. Am J Pharm Educ 2016 May 25;80(4):58 [FREE Full text] [doi: 10.5688/ajpe80458] [Medline: 27293225]
23. Liu X, Sawada Y, Takizawa T, Sato H, Sato M, Sakamoto H, et al. Doctor-patient communication: a comparison between telemedicine consultation and face-to-face consultation. Intern Med 2007;46(5):227-232 [FREE Full text] [doi: 10.2169/internalmedicine.46.1813] [Medline: 17329917]
24. Neumann M, Bensing J, Mercer S, Ernstmann N, Ommen O, Pfaff H. Analyzing the "nature" and "specific effectiveness" of clinical empathy: a theoretical overview and contribution towards a theory-based research agenda. Patient Educ Couns 2009 Mar;74(3):339-346. [doi: 10.1016/j.pec.2008.11.013] [Medline: 19124216]
25. Bokolo Anthony Jnr. Use of Telemedicine and Virtual Care for Remote Treatment in Response to COVID-19 Pandemic. J Med Syst 2020 Jun 15;44(7):132 [FREE Full text] [doi: 10.1007/s10916-020-01596-5] [Medline: 32542571]
26. Baur C. An Analysis of Factors Underlying E-Health Disparities. Camb Q Healthc Ethics 2008 Oct 01;17(4):417-428. [doi: 10.1017/s0963180108080547]
27. Lam K, Lu AD, Shi Y, Covinsky KE. Assessing Telemedicine Unreadiness Among Older Adults in the United States During the COVID-19 Pandemic. JAMA Intern Med 2020 Oct 01;180(10):1389-1391. [doi: 10.1001/jamainternmed.2020.2671] [Medline: 32744593]
28. Rahimpour M, Lovell NH, Celler BG, McCormick J. Patients' perceptions of a home telecare system. Int J Med Inform 2008 Jul;77(7):486-498. [doi: 10.1016/j.ijmedinf.2007.10.006] [Medline: 18023610]
29. Johansson A, Lindberg I, Söderberg S. Patients' Experiences with Specialist Care via Video Consultation in Primary Healthcare in Rural Areas. Int J Telemed Appl 2014;2014:143824 [FREE Full text] [doi: 10.1155/2014/143824] [Medline: 25243009]

30. Cook TM. Personal protective equipment during the coronavirus disease (COVID) 2019 pandemic - a narrative review. Anaesthesia 2020 Jul;75(7):920-927. [doi: 10.1111/anae.15071] [Medline: 32246849]

Abbreviations

EHR: electronic health record
PPE: personal protective equipment

Edited by G Eysenbach; submitted 23.11.20; peer-reviewed by M Elbattah, X He, H Mehdizadeh; comments to author 09.01.21; revised version received 30.01.21; accepted 11.04.21; published 29.04.21.

Please cite as:
Legler S, Diehl M, Hilliard B, Olson A, Markowitz R, Tignanelli C, Melton GB, Broccard A, Kirsch J, Usher M
Evaluation of an Intrahospital Telemedicine Program for Patients Admitted With COVID-19: Mixed Methods Study
J Med Internet Res 2021;23(4):e25987
URL: https://www.jmir.org/2021/4/e25987
doi:10.2196/25987
PMID:33872187

Original Paper

People's Willingness to Vaccinate Against COVID-19 Despite Their Safety Concerns: Twitter Poll Analysis

Fabian Eibensteiner[1,2*], MD; Valentin Ritschl[3,4*], MSc; Faisal A Nawaz[5], MBBSc; Sajjad S Fazel[6], MPH, PharmD; Christos Tsagkaris[7], MDc; Stefan Tino Kulnik[8], PhD; Rik Crutzen[8,9], PhD; Elisabeth Klager[2], MSc; Sabine Völkl-Kernstock[2], PhD; Eva Schaden[2,10], MD, PhD; Maria Kletecka-Pulker[2,11], PhD; Harald Willschke[2,10], MD, PhD; Atanas G Atanasov[2,12,13], PD, PhD

[1]Division of Pediatric Nephrology and Gastroenterology, Department of Pediatrics and Adolescent Medicine, Comprehensive Center for Pediatrics, Medical University of Vienna, Vienna, Austria

[2]Ludwig Boltzmann Institute for Digital Health and Patient Safety, Medical University of Vienna, Vienna, Austria

[3]Section for Outcomes Research, Center for Medical Statistics, Informatics, and Intelligent Systems, Medical University of Vienna, Vienna, Austria

[4]Ludwig Boltzmann Institute for Arthritis and Rehabilitation, Vienna, Austria

[5]College of Medicine, Mohammed Bin Rashid University of Medicine and Health Sciences, Dubai, United Arab Emirates

[6]Department of Oncology, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada

[7]Faculty of Medicine, University of Crete, Heraklion, Greece

[8]Ludwig Boltzmann Institute for Digital Health and Prevention, Salzburg, Austria

[9]Department of Health Promotion, Care and Public Health Research Institute, Maastricht University, Maastricht, Netherlands

[10]Department of Anaesthesia, Intensive Care Medicine and Pain Medicine, Medical University of Vienna, Vienna, Austria

[11]Institute for Ethics and Law in Medicine, University of Vienna, Vienna, Austria

[12]Institute of Genetics and Animal Biotechnology of the Polish Academy of Sciences, Jastrzebiec, Poland

[13]Department of Pharmacognosy, University of Vienna, Vienna, Austria

[*]these authors contributed equally

Corresponding Author:
Atanas G Atanasov, PD, PhD
Ludwig Boltzmann Institute for Digital Health and Patient Safety
Medical University of Vienna
Spitalgasse 23
Vienna, 1090
Austria
Phone: 43 6641929852
Email: Atanas.Atanasov@dhps.lbg.ac.at

Abstract

Background: On January 30, 2020, the World Health Organization's Emergency Committee declared the rapid, worldwide spread of COVID-19 a global health emergency. Since then, tireless efforts have been made to mitigate the spread of the disease and its impact, and these efforts have mostly relied on nonpharmaceutical interventions. By December 2020, the safety and efficacy of the first COVID-19 vaccines were demonstrated. The large social media platform Twitter has been used by medical researchers for the analysis of important public health topics, such as the public's perception on antibiotic use and misuse and human papillomavirus vaccination. The analysis of Twitter-generated data can be further facilitated by using Twitter's built-in, anonymous polling tool to gain insight into public health issues and obtain rapid feedback on an international scale. During the fast-paced course of the COVID-19 pandemic, the Twitter polling system has provided a viable method for gaining rapid, large-scale, international public health insights on highly relevant and timely SARS-CoV-2–related topics.

Objective: The purpose of this study was to understand the public's perception on the safety and acceptance of COVID-19 vaccines in real time by using Twitter polls.

Methods: We developed 2 Twitter polls to explore the public's views on available COVID-19 vaccines. The surveys were pinned to the Digital Health and Patient Safety Platform Twitter timeline for 1 week in mid-February 2021, and Twitter users and influencers were asked to participate in and retweet the polls to reach the largest possible audience.

Results: The adequacy of COVID-19 vaccine safety (ie, the safety of currently available vaccines; poll 1) was agreed upon by 1579 out of 3439 (45.9%) Twitter users. In contrast, almost as many Twitter users (1434/3439, 41.7%) were unsure about the safety of COVID-19 vaccines. Only 5.2% (179/3439) of Twitter users rated the available COVID-19 vaccines as generally unsafe. Poll 2, which addressed the question of whether users would undergo vaccination, was answered affirmatively by 82.8% (2862/3457) of Twitter users, and only 8% (277/3457) categorically rejected vaccination at the time of polling.

Conclusions: In contrast to the perceived high level of uncertainty about the safety of the available COVID-19 vaccines, we observed an elevated willingness to undergo vaccination among our study sample. Since people's perceptions and views are strongly influenced by social media, the snapshots provided by these media platforms represent a static image of a moving target. Thus, the results of this study need to be followed up by long-term surveys to maintain their validity. This is especially relevant due to the circumstances of the fast-paced pandemic and the need to not miss sudden rises in the incidence of vaccine hesitancy, which may have detrimental effects on the pandemic's course.

(J Med Internet Res 2021;23(4):e28973) doi:10.2196/28973

KEYWORDS

COVID-19; SARS-CoV-2; vaccine; vaccination; Twitter; survey; vaccination willingness; vaccination hesitancy; coronavirus; vaccine confidence; willingness; hesitancy; social media; safety; concern; public health; opinion; perception

Introduction

On January 30, 2020, the World Health Organization's Emergency Committee declared the rapid, worldwide spread of SARS-CoV-2 and COVID-19 a global health emergency [1]. Since then, tireless efforts have been undertaken in order to mitigate disease spread and its impacts on many different areas of public health, which range from the amount of patient and health care personnel to nationwide public health measures that mostly rely on nonpharmaceutical interventions [2-5]. Several of these measures have already been associated with the reduced transmission of COVID-19 in geographic, region-wide studies [6,7]. By December 2020, the first COVID-19 vaccine candidates were proven to be safe and efficacious in protecting against COVID-19 and were approved by regulators [8-11], and more vaccine candidates are still under development [12]. Consequently, medical societies and experts all over the world have advocated that vaccination against COVID-19 should be prioritized for high-risk groups, especially older people and people with underlying chronic medical conditions that place them at an increased risk of severe outcomes resulting from SARS-CoV-2 infection [13,14].

Since the beginning of the COVID-19 pandemic, traditional media coverage and information distribution via social media channels have been shaping public opinions and international public health strategies [15]. Although this may have positive effects on public health attitudes related to mitigation measures, one should also be aware of the detrimental effects of misinformation in media [16-20]. Misinformation should be of special consideration in the context of social media platforms such as Twitter, since false claims regarding COVID-19 appear to propagate faster on such platforms, as demonstrated in a recent study by Shahi et al [21]. However, even before the COVID-19 "infodemic," the spread of misinformation on social media platforms, on e-commerce platforms (eg, Amazon [22]), and by prominent celebrities in the United States has led to the emergence of an antivaccine movement, which has detrimental effects on national vaccine programs [23]. The resulting increased incidence of vaccination hesitancy is partly responsible for the re-emergence of measles in the United States almost 20 years after its elimination [24].

The potential implications of social media for public health are becoming increasingly clear, and the medical community was using Twitter as a tool for public health research long before the onset of the COVID-19 pandemic. The primary purposes of its use, as reported in previous health-related studies, are content analysis, surveillance, engagement, recruitment, intervention, and network analysis, and most related studies are being published in the areas of public health and infectious diseases [25]. Such Twitter-based analyses involve important public health topics that are often related to the themes of infectious diseases, such as the public's perception on antibiotic use and misuse or human papillomavirus vaccination [26,27].

Twitter—a social media platform with about 353 million monthly active users [28]—allows registered users (possible for anyone aged over 13 years) to share short, 280-character texts (also known as tweets) with other users. These tweets may be further categorized into different topics to start discussions. This is done by the use of tagging symbols, such as the hashtag symbol (#). Such discussions may be further facilitated by incorporating Twitter's built-in and anonymous polling tool. This tool has the potential to obtain insight into public health topics and real-time feedback on an international scale [29].

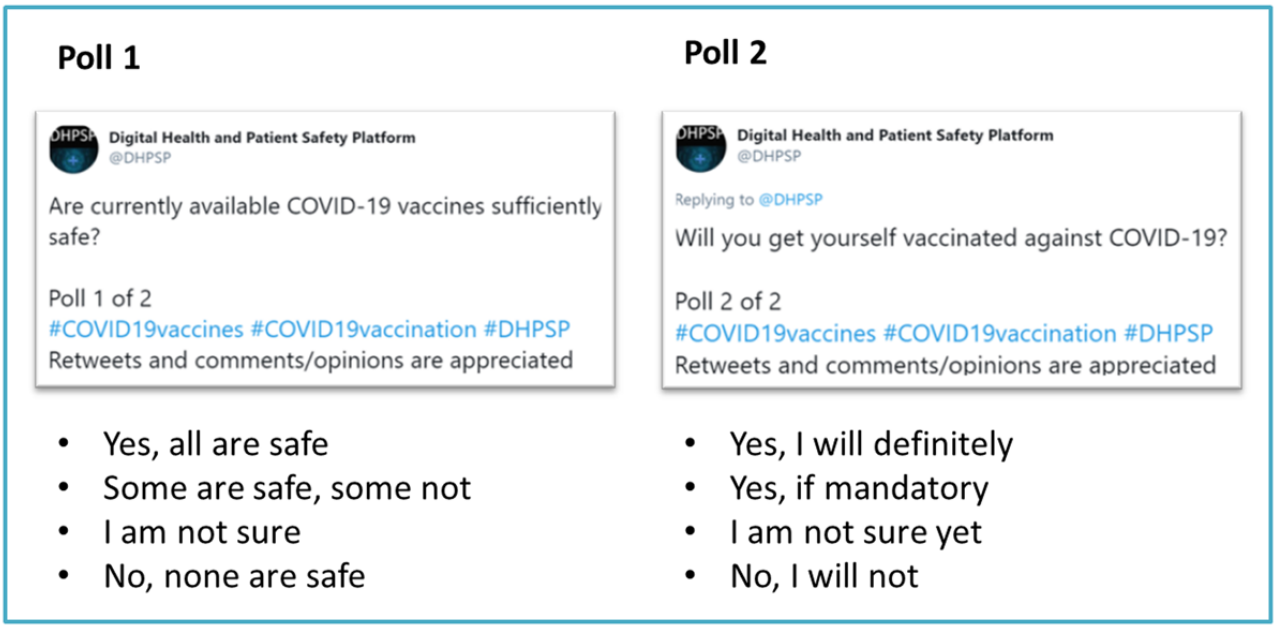
Surveying public attitudes toward COVID-19 vaccination is of high importance, since it might provide a better understanding of the reasons behind vaccine hesitancy and how to better design vaccine awareness strategies. Given the fast-paced dynamic of the COVID-19 pandemic, the Twitter polling tool seems to be a reasonable instrument for gaining immediate, large-scale, international public health insights on SARS-CoV-2–related topics. We therefore used this opportunity to rapidly collect and analyze international public health data on COVID-19 vaccines by using the Twitter polling tool. Through these efforts, we aimed to explore the potential benefits and limitations of using such a highly relevant digital health tool to investigate a highly important and broadly discussed topic.

Methods

In 2019, the Ludwig Boltzmann Institute for Digital Health and Patient Safety in Vienna, Austria was launched with the major aim of empowering patients and health care professionals with digital tools and promoting innovative research and the development of digital health and patient safety tools. One of these innovations was the initiation of the Digital Health and Patient Safety Platform (DHPSP) [30], which provides subscribers with an overview of recent scientific publications regarding digital health and patient safety.

For this study, we used the Twitter account of the DHPSP (Twitter handle: @DHPSP) to distribute two polls regarding COVID-19 vaccine safety and acceptance. The polls were developed by the authors and posted on Twitter under the @DHPSP Twitter handle between February 12 and February 19, 2021. Poll 1 addressed the perceived safety of the available (at the time of polling) COVID-19 vaccines ("Are currently available COVID-19 vaccines sufficiently safe?"), whereas poll 2 addressed the confidence or hesitancy of the respondents with regard to undergoing vaccination against COVID-19 ("Will you get yourself vaccinated against COVID-19?"). Both polls were linked (poll 2 was posted as a comment below the tweet for poll 1) and pinned at the top of the DHPSP Twitter timeline during the polling period. The poll questions, which included hashtags for categorization, were limited (by Twitter) to 280 characters. Twitter allows up to 4 answers with a limit of 25 characters (including spaces) for each poll. Therefore, both polls included 4 answers that ranged from total agreement ("Yes, all are safe" and "Yes, I will definitely") to total disagreement ("No, none are safe" and "No, I will not") and were presented in the manner of a 4-point response scale. Both polls were categorized using the following hashtags to promote better visibility and facilitate analysis: #COVID19vaccines, #COVID19vaccination, and #DHPSP. Figure 1 displays the detailed construction of both polls on Twitter.

Figure 1. Structure of the two Twitter polls.



After launching the polls, the first people who could see them on their Twitter timelines were the DHPSP Twitter followers. Twitter poll votes are anonymous and do not allow for the evaluation of respondents' characteristics (eg, gender). Therefore, in order to at least obtain some data on the characteristics of the audience that was first exposed to the polls, we aimed to analyze the follower characteristics of @DHPSP via the web-based tool Followerwonk [31] on February 20, 2021 (Table 1). In total, the @DHPSP Twitter account had 526 followers at the time of the analysis. Of these, 121 (23%) were male, 69 (13.1%) were female, and 336 (63.9%) did not state their gender on Twitter. A total of 66 (12.6%) @DHPSP followers had more than 5000 followers, 152 (28.9%) had between 500 and 5000 followers, and 308 (58.6%) had less than 499 followers.

Table 1. @DHPSP's Twitter follower characteristics (N=526).

Characteristics	Value, n (%)
Gender	
Male	121 (23)
Female	68 (13)
Not stated	337 (64)
Follower count	
<499	305 (58)
500-5000	153 (29)
>5000	68 (13)
Account age (years)	
<1	95 (18)
1-5	158 (30)
>5	273 (52)
Language	
English	310 (59)
Spanish	21 (4)
Other	195 (37)

The polls' body message encouraged people to retweet the polls ("Retweets and comments/opinions are appreciated"; Figure 1), and with each new retweet, the polls gained a bigger audience (consisting of the followers of the retweeting accounts). Moreover, to achieve greater visibility, members and email list subscribers of the DHPSP [30] were asked to support the polls by voting; retweeting the polls; and disseminating them via diverse networking approaches, including direct emails or direct social media messages. The website of the DHPSP and diverse social media accounts of DHPSP members were also used to post hyperlinks to the polls. Additionally, information for the polls was shared through the DHPSP Facebook [32] and LinkedIn [33] accounts.

To characterize the population of users that retweeted the studied polls, we performed a hashtag analysis by using the web-based tool Symplur Signals [34]. We therefore analyzed the number of retweets, users, locations, and languages of all tweets that contained our unique combination of hashtags (#COVID19vaccines, #COVID19vaccination, and #DHPSP) by the end of this study. This was done on the day after the polls were closed (February 20, 2021). To ensure accuracy and to limit interference bias from other Twitter discussions related to our topic, we conducted a Twitter search prior to launching the polls (February 11, 2021), and we confirmed that our combination of hashtags was never used.

Ethical approval was not required for this study, as it was outside the scope of the medical ethics law in Austria. Participation in the studied polls was completely anonymous. Therefore, the collected data were outside the scope of the General Data Protection Regulation [35]. Excluding the poll votes, analyzed parameters such as the number of followers and retweets were based on data that were publicly available on the internet.

Results

Both of our Twitter polls were pinned to the Twitter timeline of the @DHPSP Twitter account for 7 days, beginning on February 12, 2021. Pinning a tweet permanently places it at the top of a Twitter user's account. Therefore, any new visitors will see this tweet at the top of the visited user's timeline.

Poll 1 ("Are currently available COVID-19 vaccines sufficiently safe?") received a total of 3439 votes (194,695 views), whereas poll 2 ("Will you get yourself vaccinated against COVID-19?") received a total of 3457 votes (246,814 views). The analysis of the poll retweets that contained our unique combination of hashtags (#COVID19vaccines, #COVID19vaccination, and #DHPSP) revealed a total of 930 tweets from 375 users. Overall, 262 (69.9%) users posted 1 retweet, 67 (17.9%) users posted 2 retweets, and 46 (12.3%) posted ≥3 retweets. The polls, including all retweets, had a total of 15,446,703 views on Twitter. The top 3 locations of Twitter users who retweeted the polls were the United States of America (n=64, 17.1%), the United Kingdom (n=19, 5.1%), and Canada (n=16, 4.2%). A summary of these details is provided in Table 2. The other top locations (rank 4-10) were India (13 users), Mexico (8 users), Argentina (5 users), Spain (5 users), Australia (4 users), United Arab Emirates (4 users), and Italy (3 users).

Table 2. Analysis of poll retweets that contained our unique combination of hashtags (#COVID19vaccines, #COVID19vaccination, and #DHPSP).

Characteristics	Value, n (%)
Top locations[a,b]	
United States of America	64 (17.1)
United Kingdom	19 (5.1)
Canada	16 (4.2)
Number of retweets[b]	
1	262 (69.9)
2[c]	67 (17.9)
≥3[c]	46 (12.3)
Top languages[d,e]	
English	802 (86.2)
Spanish	6 (0.6)
Indonesian	5 (0.5)

[a]Determined based on data derived from the users who indicated their location in their account information on Twitter. During the interpretation of the data, readers should be aware that 50.5% (189/375) of Twitter users did not provide location information on their profiles.

[b]Percentage is based on the number of Twitter users who retweeted the polls (N=375).

[c]Includes regular retweets, retweets with comments, and quote retweets (in which a hyperlink to the original tweet is inserted in a newly composed tweet).

[d]Only the most used languages are indicated. All other tweet languages each accounted for less than 0.5% (5/930) of the retweets.

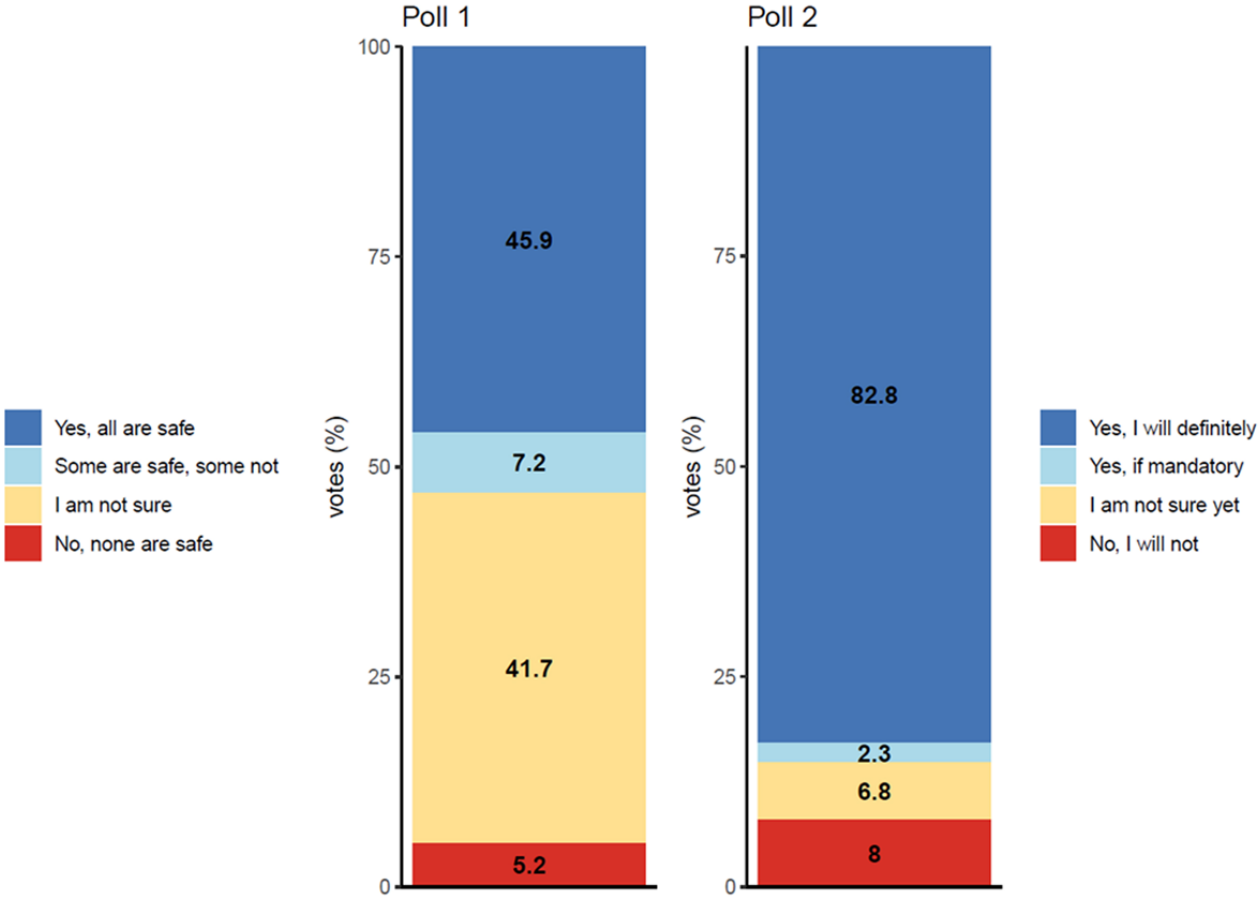
[e]Percentage is based on the number of retweets (N=930).

In total, 45.9% (1579/3439) of Twitter users who responded to poll 1 (“Are currently available COVID-19 vaccines sufficiently safe?”) voted with total agreement (“Yes, all are safe”), meaning that the users considered all currently available COVID-19 vaccines to be safe. However, almost as many Twitter users (1434/3439, 41.7%) were not sure about the safety of the available COVID-19 vaccines (voted with “I am not sure”). Interestingly, only 5.2% (179/3439) of the respondents in this poll felt that the available COVID-19 vaccines were generally unsafe (“No, none are safe”). In addition, a total of 7.2% of Twitter users (248/3439) advocated for the safety of some vaccines but felt that not all of them were safe.

Poll 2 explored Twitter users’ confidence or hesitancy toward undergoing vaccination against COVID-19 (“Will you get yourself vaccinated against COVID-19?”). A majority (2862/3457, 82.8%) of the respondents stated, “[y]es, I definitely will [get vaccinated].” In total, 6.8% (235/3457) of Twitter respondents were not yet sure (voted with “I am not sure yet”) about undergoing vaccination, and 8% (277/3457) categorically rejected vaccination at the time of polling (“No, I will not”). Only a minor percentage (80/3457, 2.3%) of Twitter users stated that they would undergo vaccination if it was mandatory.

A detailed summary of the answers to both polls is provided in Figure 2.

Figure 2. Twitter users' answers to poll 1 ("Are currently available COVID-19 vaccines sufficiently safe?"; respondents: n=3439) and poll 2 ("Will you get yourself vaccinated against COVID-19?"; respondents: n=3457).



Discussion

In the context of the fast-paced dynamic of the COVID-19 pandemic, we used the rapid, progressive environment of social media (ie, Twitter) to gain international insights into the public's opinion on COVID-19 vaccination. We followed a methodological approach that was outlined in a previous study on public attitudes toward telemedicine, which was conducted by Vidal-Alaball et al [29]. They suggested that the Twitter polling tool for quick surveys on timely topics should be used to obtain prompt feedback for new questionnaires before their validation. In this study, by using the DHPSP's Twitter handle and gaining the support of the retweeting accounts, we were able to validate the Twitter polling approach on a large scale. We obtained a 30-fold higher poll response rate and view rate (impressions) and an 18-fold higher retweet rate compared to those of Vidal-Alaball et al [29]. In the previously mentioned study regarding attitudes toward telemedicine, Vidal-Alaball et al [29] only used the Twitter handle of one of the authors, whereas in our study, an established Twitter network (the DHPSP) was used to achieve a greater reach and higher response rates. Therefore, we were able to not only validate the previously published approach on a larger scale but also demonstrate the definitive advantage of an established user network that is less dependent on single users for achieving a wider reach and higher response rates.

Although this study involves the first scientific Twitter poll analysis of the perceived safety of available (at the time of polling) COVID-19 vaccines and the confidence or hesitancy of respondents with regard to undergoing vaccination against COVID-19, it is not the first survey on this topic in medical literature. In this study, despite the insecurities about the sufficient safety of the available (at the time of polling) COVID-19 vaccines (BNT162b2 by Pfizer-BioNTech, mRNA-1273 by Moderna, ChAdOx1 by AstraZeneca, and Gam-COVID-Vac by the Gamaleya Research Institute of Epidemiology and Microbiology of the Russian Federation), which was observed in 54.1% (1861/3439) of the poll's respondents, a surprisingly large group of respondents (2863/3457, 82.8%) voted that they would definitely undergo vaccination.

Although this is a positive result, due to the Twitter poll's anonymous nature, there is no reassurance that the Twitter users who answered the first poll also answered the second poll. In contrast to this study, in a large-scale international analysis of 13,426 participants from 19 countries that was conducted via multiple international, web-based panel providers (Dynata, Opinion Access, Survey Monkey, and Amazon MTurk), Lazarus et al [36] found that only 46.8% of participants agreed to accept a COVID-19 vaccine if it was generally available; 24.7% somewhat agreed; and 14% and 34.1% completely and somewhat agreed to accept a COVID-19 vaccine if it was recommended by their employer, respectively. However, the data from the Lazarus et al [36] study were collected in June

2020. At that time, none of the currently available COVID-19 vaccines were approved by regulatory authorities. Other studies that were conducted during an earlier pandemic phase reported higher rates of willingness to undergo vaccination once a vaccine against COVID-19 became available, ranging from 59% to 75% [37].

Our study may be limited in terms of interpretability. This is due to the fact that the visibility of the polls was widely promoted by the DHPSP, which has a follower base that consists of highly educated individuals with scientific backgrounds or strong interests in science. Since the DHPSP account exclusively posts science-based content, it is reasonable to assume that it attracted followers with interests in science. This assumption is in line with a recent study by Schwarzinger et al [38], which revealed that COVID-19 vaccination hesitancy was highly prevalent among people with low educational levels in the French population. These findings are in line with other, more recent studies that were conducted in France, the United States of America, and Australia [39-41]. As reported and discussed by Kreps et al [42], several factors associated with willingness, hesitancy, advice, and recommendations to vaccinate against COVID-19 are consistent with those in past studies on other vaccines, whereas other factors may be more complex due to the fast-paced dynamic and unpredictable course of the pandemic as well as difficult political and public health actions and communications.

The number of surveys on confidence and hesitancy toward vaccinating against COVID-19 that have been conducted over the course of the pandemic has demonstrated the importance of regular reviews on the public's opinion toward the effectiveness and safety of vaccines. Such reviews are needed in order to instill public confidence [36]. Lazarus et al [36] also concluded that one of the most important factors for initiating positive health behaviors is credible and culturally informed health communication. With this in mind, health authorities could reach out to the public via rapid, low-barrier, and easy-to-access media platforms, such as Twitter, in order to monitor and instill positive health behaviors through the provision of clear information and credible sources that are tailored to the cultural backgrounds of target populations. As Twitter polls provide the necessary anonymity for confident and large-scale participation and allow for rapid and concise questioning, we propose that this tool is useful for reaching out to the public and addressing public health issues.

The strengths of this tool and this study lie in rapid assessment, the large-scale dissemination of information, and the expeditious retrieval of concise information. The possible strengths of using a preformed network to disseminate information and surveys in order to reach a broad target audience are shown in this study, especially in our comparison of single-user promotion and our promotion method. The very tight restrictions of Twitter polling, including question and answer character limits as well as limits on the total number of answers, might serve as strengths for constraining poll creators to the development of concise and well-formulated surveys, which might result in higher response rates than those of traditional surveys. However, these restraints may also impede the formulation of more complex questions and the clarification of questions and answers. This might interfere with poll results, as a lack of clarity may result in different interpretations among poll participants. Therefore, Twitter polling is better suited for clear, concise, and close-ended questions instead of open-ended and semistructured questions that leave room for interpretation.

The pinning and promoting of Twitter polls by specific accounts may also interfere with sample selection and result in biases that may be challenging to mitigate because of the lack of data on the baseline characteristics (eg, gender, age, and socioeconomic status) of participants. However, this challenge can be easily overcome by promoting Twitter polls for a longer period of time across multiple accounts and groups. Nevertheless, it should be noted that complete anonymity can result in the manipulation of votes in the polling tool due to people exploiting multiple usership, as outlined in the study by Vidal-Alaball et al [29]. Even though the analysis of Twitter network followers is made possible by third-party web-based tools, the geographical distribution of the @DHPSP follower network (Figure 3) might not be representative of our polls' responders, as random Twitter users were able to participate in these polls. The Twitter polls did not undergo a formal validation process. Future studies (eg, those that pilot polling tools) should address the validation of the questions used in this study. However, we were able to gain a better perspective of the users who responded to the polls by using a novel approach, which involved a unique combination of hashtags and a hashtag analysis, to gain insights on the population of users who retweeted the Twitter polls.

In conclusion, despite the high levels of uncertainty regarding the safety of available COVID-19 vaccines in the study sample, the respondents had a high willingness to undergo vaccination. The public's perceptions and views on health issues are strongly influenced by social media. This underscores the importance of using social media polling tools to understand public health perspectives in real time. Such information can be used to inform public health messaging and communication efforts. Regular surveys on public health issues that use social media platforms may aid in the early discovery of sudden rises in the incidence of COVID-19 vaccination hesitancy among the public before the detrimental effects of the pandemic can manifest.

Figure 3. Main locations of the DHPSP's Twitter followers. These data cover only a fraction of the DHPSP followers who indicated their location in their account information on Twitter. DHPSP: Digital Health and Patient Safety Platform.

Conflicts of Interest

None declared.

References

1. Velavan TP, Meyer CG. The COVID-19 epidemic. Trop Med Int Health 2020 Mar;25(3):278-280 [FREE Full text] [doi: 10.1111/tmi.13383] [Medline: 32052514]
2. Anderson M, Mckee M, Mossialos E. Covid-19 exposes weaknesses in European response to outbreaks. BMJ 2020 Mar 18;368:m1075. [doi: 10.1136/bmj.m1075] [Medline: 32188590]
3. Anderson RM, Heesterbeek H, Klinkenberg D, Hollingsworth TD. How will country-based mitigation measures influence the course of the COVID-19 epidemic? Lancet 2020 Mar 21;395(10228):931-934 [FREE Full text] [doi: 10.1016/S0140-6736(20)30567-5] [Medline: 32164834]
4. Chu DK, Akl EA, Duda S, Solo K, Yaacoub S, Schünemann HJ, COVID-19 Systematic Urgent Review Group Effort (SURGE) study authors. Physical distancing, face masks, and eye protection to prevent person-to-person transmission of SARS-CoV-2 and COVID-19: a systematic review and meta-analysis. Lancet 2020 Jun 27;395(10242):1973-1987 [FREE Full text] [doi: 10.1016/S0140-6736(20)31142-9] [Medline: 32497510]
5. Eibensteiner F, Ritschl V, Stamm T, Cetin A, Schmitt CP, Ariceta G, et al. Countermeasures against COVID-19: how to navigate medical practice through a nascent, evolving evidence base - a European multicentre mixed methods study. BMJ Open 2021 Feb 17;11(2):e043015 [FREE Full text] [doi: 10.1136/bmjopen-2020-043015] [Medline: 33597140]
6. Pan A, Liu L, Wang C, Guo H, Hao X, Wang Q, et al. Association of public health interventions with the epidemiology of the COVID-19 outbreak in Wuhan, China. JAMA 2020 May 19;323(19):1915-1923 [FREE Full text] [doi: 10.1001/jama.2020.6130] [Medline: 32275295]
7. Cowling BJ, Ali ST, Ng TWY, Tsang TK, Li JCM, Fong MW, et al. Impact assessment of non-pharmaceutical interventions against coronavirus disease 2019 and influenza in Hong Kong: an observational study. Lancet Public Health 2020 May;5(5):e279-e288 [FREE Full text] [doi: 10.1016/S2468-2667(20)30090-6] [Medline: 32311320]
8. Baden LR, El Sahly HM, Essink B, Kotloff K, Frey S, Novak R, COVE Study Group. Efficacy and Safety of the mRNA-1273 SARS-CoV-2 Vaccine. N Engl J Med 2021 Feb 04;384(5):403-416 [FREE Full text] [doi: 10.1056/NEJMoa2035389] [Medline: 33378609]
9. Logunov DY, Dolzhikova IV, Shcheblyakov DV, Tukhvatulin AI, Zubkova OV, Dzharullaeva AS, Gam-COVID-Vac Vaccine Trial Group. Safety and efficacy of an rAd26 and rAd5 vector-based heterologous prime-boost COVID-19 vaccine: an interim analysis of a randomised controlled phase 3 trial in Russia. Lancet 2021 Feb 20;397(10275):671-681 [FREE Full text] [doi: 10.1016/S0140-6736(21)00234-8] [Medline: 33545094]
10. Polack FP, Thomas SJ, Kitchin N, Absalon J, Gurtman A, Lockhart S, C4591001 Clinical Trial Group. Safety and efficacy of the BNT162b2 mRNA Covid-19 Vaccine. N Engl J Med 2020 Dec 31;383(27):2603-2615 [FREE Full text] [doi: 10.1056/NEJMoa2034577] [Medline: 33301246]
11. Voysey M, Clemens SAC, Madhi SA, Weckx LY, Folegatti PM, Aley PK, Oxford COVID Vaccine Trial Group. Safety and efficacy of the ChAdOx1 nCoV-19 vaccine (AZD1222) against SARS-CoV-2: an interim analysis of four randomised

controlled trials in Brazil, South Africa, and the UK. Lancet 2021 Jan 09;397(10269):99-111 [FREE Full text] [doi: 10.1016/S0140-6736(20)32661-1] [Medline: 33306989]
12. Sadoff J, Le Gars M, Shukarev G, Heerwegh D, Truyers C, de Groot AM, et al. Interim Results of a Phase 1-2a Trial of Ad26.COV2.S Covid-19 Vaccine. N Engl J Med. Epub ahead of print. 2021 Jan 13 [FREE Full text] [doi: 10.1056/NEJMoa2034201] [Medline: 33440088]
13. Gharpure R, Patel A, Link-Gelles R. First-Dose COVID-19 vaccination coverage among skilled nursing facility residents and staff. JAMA. Epub ahead of print. 2021 Feb 24. [doi: 10.1001/jama.2021.2352] [Medline: 33625464]
14. Windpessl M, Bruchfeld A, Anders HJ, Kramer H, Waldman M, Renia L, et al. COVID-19 vaccines and kidney disease. Nat Rev Nephrol. Epub ahead of print. 2021 Feb 08 [FREE Full text] [doi: 10.1038/s41581-021-00406-6] [Medline: 33558753]
15. Su Y. It doesn't take a village to fall for misinformation: Social media use, discussion heterogeneity preference, worry of the virus, faith in scientists, and COVID-19-related misinformation beliefs. Telematics and Informatics 2021 May;58:101547. [doi: 10.1016/j.tele.2020.101547]
16. Al-Dmour H, Masa'deh R, Salman A, Abuhashesh M, Al-Dmour R. Influence of social media platforms on public health protection against the COVID-19 pandemic via the mediating effects of public health awareness and behavioral changes: Integrated model. J Med Internet Res 2020 Aug 19;22(8):e19996 [FREE Full text] [doi: 10.2196/19996] [Medline: 32750004]
17. Depoux A, Martin S, Karafillakis E, Preet R, Wilder-Smith A, Larson H. The pandemic of social media panic travels faster than the COVID-19 outbreak. J Travel Med 2020 May 18;27(3):taaa031 [FREE Full text] [doi: 10.1093/jtm/taaa031] [Medline: 32125413]
18. Garrett L. COVID-19: the medium is the message. Lancet 2020 Mar 21;395(10228):942-943 [FREE Full text] [doi: 10.1016/S0140-6736(20)30600-0] [Medline: 32171075]
19. Gottlieb M, Dyer S. Information and disinformation: Social media in the COVID-19 crisis. Acad Emerg Med 2020 Jul;27(7):640-641 [FREE Full text] [doi: 10.1111/acem.14036] [Medline: 32474977]
20. Liao Q, Yuan J, Dong M, Yang L, Fielding R, Lam WWT. Public engagement and government responsiveness in the communications about COVID-19 during the early epidemic stage in China: Infodemiology study on social media data. J Med Internet Res 2020 May 26;22(5):e18796 [FREE Full text] [doi: 10.2196/18796] [Medline: 32412414]
21. Shahi GK, Dirkson A, Majchrzak TA. An exploratory study of COVID-19 misinformation on Twitter. Online Soc Netw Media 2021 Mar;22:100104 [FREE Full text] [doi: 10.1016/j.osnem.2020.100104] [Medline: 33623836]
22. Shin J, Valente T. Algorithms and health misinformation: A case study of vaccine books on Amazon. J Health Commun 2020 May 03;25(5):394-401. [doi: 10.1080/10810730.2020.1776423] [Medline: 32536257]
23. Dubé E, Vivion M, MacDonald NE. Vaccine hesitancy, vaccine refusal and the anti-vaccine movement: influence, impact and implications. Expert Rev Vaccines 2015 Jan;14(1):99-117. [doi: 10.1586/14760584.2015.964212] [Medline: 25373435]
24. Hotez PJ, Nuzhath T, Colwell B. Combating vaccine hesitancy and other 21st century social determinants in the global fight against measles. Curr Opin Virol 2020 Apr;41:1-7. [doi: 10.1016/j.coviro.2020.01.001] [Medline: 32113136]
25. Sinnenberg L, Buttenheim AM, Padrez K, Mancheno C, Ungar L, Merchant RM. Twitter as a tool for health research: A systematic review. Am J Public Health 2017 Jan;107(1):e1-e8. [doi: 10.2105/AJPH.2016.303512] [Medline: 27854532]
26. Kendra RL, Karki S, Eickholt JL, Gandy L. Characterizing the discussion of antibiotics in the Twittersphere: What is the bigger picture? J Med Internet Res 2015 Jun 19;17(6):e154 [FREE Full text] [doi: 10.2196/jmir.4220] [Medline: 26091775]
27. Massey PM, Leader A, Yom-Tov E, Budenz A, Fisher K, Klassen AC. Applying multiple data collection tools to quantify human papillomavirus vaccine communication on Twitter. J Med Internet Res 2016 Dec 05;18(12):e318 [FREE Full text] [doi: 10.2196/jmir.6670] [Medline: 27919863]
28. das Statistik-Portal: Statistiken, Marktdaten & Studien. Statista GmbH. URL: https://de.statista.com/ [accessed 2021-03-04]
29. Vidal-Alaball J, Fernandez-Luque L, Marin-Gomez FX, Ahmed W. A new tool for public health opinion to give insight into telemedicine: Twitter poll analysis. JMIR Form Res 2019 May 28;3(2):e13870 [FREE Full text] [doi: 10.2196/13870] [Medline: 31140442]
30. Digital Health and Patient Safety Platform. Digital Patient Safety. URL: https://digitalpatientsafety.com/ [accessed 2021-03-19]
31. Tools for Twitter analytics, bio search and more. Followerwonk. URL: https://followerwonk.com/ [accessed 2021-02-20]
32. Digital Health and Patient Safety Platform. Facebook. URL: https://www.facebook.com/DHPSP [accessed 2021-03-19]
33. Ludwig Boltzmann Institute for Digital Health and Patient Safety. LinkedIn. URL: https://at.linkedin.com/company/ludwig-boltzmann-institute-for-digital-health-and-patient-safety [accessed 2021-03-19]
34. Healthcare analytics products for social and real-world data. Symplur. URL: https://www.symplur.com/ [accessed 2021-02-20]
35. Crutzen R, Peters GJY, Mondschein C. Why and how we should care about the General Data Protection Regulation. Psychol Health 2019 Nov;34(11):1347-1357. [doi: 10.1080/08870446.2019.1606222] [Medline: 31111730]
36. Lazarus JV, Ratzan SC, Palayew A, Gostin LO, Larson HJ, Rabin K, et al. A global survey of potential acceptance of a COVID-19 vaccine. Nat Med 2021 Feb;27(2):225-228 [FREE Full text] [doi: 10.1038/s41591-020-1124-9] [Medline: 33082575]

37. Reiter PL, Pennell ML, Katz ML. Acceptability of a COVID-19 vaccine among adults in the United States: How many people would get vaccinated? Vaccine 2020 Sep 29;38(42):6500-6507 [FREE Full text] [doi: 10.1016/j.vaccine.2020.08.043] [Medline: 32863069]
38. Schwarzinger M, Watson V, Arwidson P, Alla F, Luchini S. COVID-19 vaccine hesitancy in a representative working-age population in France: a survey experiment based on vaccine characteristics. Lancet Public Health 2021 Apr;6(4):e210-e221 [FREE Full text] [doi: 10.1016/S2468-2667(21)00012-8] [Medline: 33556325]
39. Hacquin AS, Altay S, de Araujo E, Chevallier C, Mercier H. Sharp rise in vaccine hesitancy in a large and representative sample of the French population: reasons for vaccine hesitancy. PsyArXiv. Preprint posted online on November 17, 2020. [FREE Full text] [doi: 10.31234/osf.io/r8h6z]
40. Lin C, Tu P, Beitsch LM. Confidence and receptivity for COVID-19 vaccines: A rapid systematic review. Vaccines (Basel) 2020 Dec 30;9(1):16 [FREE Full text] [doi: 10.3390/vaccines9010016] [Medline: 33396832]
41. Rhodes A, Hoq M, Measey MA, Danchin M. Intention to vaccinate against COVID-19 in Australia. Lancet Infect Dis. Epub ahead of print. 2020 Sep 14 [FREE Full text] [doi: 10.1016/S1473-3099(20)30724-6] [Medline: 32941786]
42. Kreps S, Prasad S, Brownstein JS, Hswen Y, Garibaldi BT, Zhang B, et al. Factors associated with US adults' likelihood of accepting COVID-19 vaccination. JAMA Netw Open 2020 Oct 01;3(10):e2025594 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.25594] [Medline: 33079199]

Abbreviations

DHPSP: Digital Health and Patient Safety Platform

Edited by G Eysenbach; submitted 20.03.21; peer-reviewed by J Vidal-Alaball, S Martin; comments to author 12.04.21; revised version received 17.04.21; accepted 17.04.21; published 29.04.21.

Please cite as:
Eibensteiner F, Ritschl V, Nawaz FA, Fazel SS, Tsagkaris C, Kulnik ST, Crutzen R, Klager E, Völkl-Kernstock S, Schaden E, Kletecka-Pulker M, Willschke H, Atanasov AG
People's Willingness to Vaccinate Against COVID-19 Despite Their Safety Concerns: Twitter Poll Analysis
J Med Internet Res 2021;23(4):e28973
URL: https://www.jmir.org/2021/4/e28973
doi:10.2196/28973
PMID:33872185

Original Paper

A Peer-to-Peer Live-Streaming Intervention for Children During COVID-19 Homeschooling to Promote Physical Activity and Reduce Anxiety and Eye Strain: Cluster Randomized Controlled Trial

Yingfeng Zheng[1,2,3*], MD, PhD; Wei Wang[1*], MD, PhD; Yuxin Zhong[1], BA; Fengchun Wu[4,5], MD; Zhuoting Zhu[1], MD, PhD; Yih-Chung Tham[6], MD, PhD; Ecosse Lamoureux[6], MD, PhD; Liang Xiao[7], BSc; Erta Zhu[7], BSc; Haoning Liu[8], BSc; Ling Jin[1], MSc; Linyi Liang[1], MD, PhD; Lixia Luo[1], MD, PhD; Mingguang He[1], MD, PhD; Ian Morgan[9], PhD; Nathan Congdon[10*], MD, PhD; Yizhi Liu[1*], MD, PhD

[1]State Key Laboratory of Ophthalmology, Zhongshan Ophthalmic Center, Sun Yat-sen University, Guangzhou, China

[2]Guangzhou Regenerative Medicine and Health Guangdong Laboratory, Guangzhou, China

[3]Research Units of Ocular Development and Regeneration, Chinese Academy of Medical Sciences, Guangzhou, China

[4]Department of Psychiatry, Affiliated Brain Hospital of Guangzhou Medical University (Guangzhou Huiai Hospital), Guangzhou, China

[5]Department of Psychiatry, Guangdong Engineering Technology Research Center for Translational Medicine of Mental Disorders, Guangzhou, China

[6]Singapore Eye Research Institute, Singapore National Eye Centre, Singapore, Singapore

[7]Zhaoqing Education Bureau, Zhaoqing, China

[8]Duanzhou District Education Bureau, Zhaoqing, China

[9]Research School of Biology, Australian National University, Canberra, Australia

[10]Centre for Public Health, Queen's University Belfast, Belfast, Ireland

[*]these authors contributed equally

Corresponding Author:
Yingfeng Zheng, MD, PhD
State Key Laboratory of Ophthalmology
Zhongshan Ophthalmic Center
Sun Yat-sen University
No 7 Jinsui Rd, Room 509
Zhujiang New Town
Guangzhou, 510060
China
Phone: 86 13922286455
Email: yingfeng.zheng@qq.com

Abstract

Background: The COVID-19 pandemic has led to worldwide school closures, with millions of children confined to online learning at home. As a result, children may be susceptible to anxiety and digital eye strain, highlighting a need for population interventions.

Objective: The objective of our study was to investigate whether a digital behavior change intervention aimed at promoting physical activity could reduce children's anxiety and digital eye strain while undergoing prolonged homeschooling during the COVID-19 pandemic.

Methods: In this cluster randomized controlled trial, homeschooled grade 7 students at 12 middle schools in southern China were recruited through local schools and randomly assigned by the school to receive (1:1 allocation): (1) health education information promoting exercise and ocular relaxation, and access to a digital behavior change intervention, with live streaming and peer sharing of promoted activities (intervention), or (2) health education information only (control). The primary outcome was change in self-reported anxiety score. Secondary outcomes included change in self-reported eye strain and sleep quality.

Results: On March 16, 2020, 1009 children were evaluated, and 954 (94.5%) eligible children of consenting families were included in the intention-to-treat analysis. Children in the intervention (n=485, 6 schools) and control (n=469, 6 schools) groups

were aged 13.5 (SD 0.5) years, and 52.3% (n=499) were male. The assigned interventions were completed by 896 children (intervention: n=467, 96.3%; control: n=429, 91.5%). The 2-week change in square-root–transformed self-reported anxiety scores was greater in the intervention (–0.23, 95% CI –0.27 to –0.20) vs control group (0.12, 95% CI 0.09-0.16; unadjusted difference –0.36, 95% CI –0.63 to –0.08; *P*=.02). There was a significant reduction in square-root–transformed eye strain in the intervention group (–0.08, 95% CI –0.10 to 0.06) compared to controls (0.07, 95% CI 0.05-0.09; difference –0.15, 95% CI –0.26 to –0.03; *P*=.02). Change in sleep quality was similar between the two groups.

Conclusions: This digital behavior change intervention reduced children's anxiety and eye strain during COVID-19–associated online schooling.

Trial Registration: ClinicalTrials.gov NCT04309097; http://clinicaltrials.gov/ct2/show/NCT04309097

(J Med Internet Res 2021;23(4):e24316) doi:10.2196/24316

KEYWORDS

homeschooling; children; anxiety, digital eye strain; peer to peer; live streaming; digital health; intervention; health information; physical activity; COVID-19; online learning; behavior; app; mobile phone

Introduction

COVID-19 has spread worldwide, with more than 24 million confirmed cases as of August 29, 2020 [1]. UNESCO (United Nations Educational, Scientific and Cultural Organization) announced that at least 188 countries have closed schools nationwide as of August 2020, which has resulted in the unprecedented adoption of online homeschooling [2]. An estimated 1.5 billion children are learning at home without direct access to school activities [2].

Prolonged adoption of homeschooling has important implications for children's mental health. Although the long-term mental health consequences of the COVID-19 pandemic on children have not been systematically explored [3-5], it has been suggested that quarantined children have an average posttraumatic stress score 4 times higher than nonquarantined children [6]. Given that school isolation and stay-at-home orders could further worsen anxiety and other mental health problems, there is a critical need for novel interventions to safeguard the well-being of students in this regard [7-9].

In addition to its psychological impacts, the COVID-19 pandemic can take a toll on children's vision as many schools move online. The expanded use of digital devices and increased screen time may result in worsening symptoms of eye strain and pose greater risk for myopic progression [10]. The American Academy of Ophthalmology has recommended the 20-20-20 rule, which calls for taking a visual break every 20 minutes by looking at an object 20 feet away for 20 seconds. While it is important to deliver such health information, it is equally necessary to study effective ways to promote behavior change, motivate eye relaxation exercises, and test effectiveness in reducing eye strain.

Digital behavior change interventions have been considered as an effective approach to promote physical activity and/or reduce sedentary behavior [11]. Furthermore, physical activity may be a useful target for strategies to handle stress, anxiety, sleeping problem, and eye relaxation, given previous reports regarding the associations of a lower level of physical activity with anxiety [12], depression [13], sleep disorder [14], and dry eye disease [15]. We have therefore hypothesized that a digital behavior intervention aimed at promoting physical activity could improve outcomes in anxiety, sleeping, and eye strain. This issue is critically important, given the magnitude of the COVID-19 pandemic. While this manuscript was under review, it was reported that individuals with inadequate physical activity reported a greater level of psychological distress during the COVID-19 pandemic [16].

The advent of digital technology has provided new opportunities to deliver digital behavior interventions on a population scale [17]. More than 90% of American teenagers are online every day (averaging more than 6.5 hours daily), where they spend much of their time interacting with friends and family through social media [18]. More than 40% of adolescents in grades 7 to 9 in China report at least 2 hours of exposure to digital screens per day [19]. In particular, live-streaming apps have become a popular form of social entertainment for children. However, despite the increasing adoption of live-streaming apps among children and teenagers, their use in public health interventions has yet to receive wide attention in clinical trials.

In light of this gap, we developed a novel digital behavior change intervention that encourages children to engage in regular physical activity and relaxation of accommodation (near focusing) during online school recess periods. The aim of this study is to evaluate the effectiveness of this digital intervention in reducing anxiety syndrome (main outcome) and eye strain, compared to a conventional educational intervention among Chinese children during a recent period of home confinement occasioned by the COVID-19 pandemic.

Methods

Study Design and Participants

This cluster randomized controlled trial was conducted in the Duanzhou district of Zhaoqing City, Guangdong Province, in southern China. This setting was chosen because Zhaoqing, one of southern China's major cities, is readily accessible from Guangzhou, the provincial capital. The 2016 national census shows that Zhaoqing has a relatively stable population of 4,084,600, which is representative of the Chinese national urban population in terms of demographic and socioeconomic characteristics [20]. Smartphone use is very widespread in this area, as throughout the rest of China, where the number of

mobile phone subscribers in December 2019 (1.6 billion) exceeded the total population of 1.43 billion [21]. China has mandated 9 years of compulsory education since 1986, and the enrollment rate for the 7th grade (12-14 years) is nearly 100% in Zhaoqing. Thus, the school-going cohort is representative of children in the general population.

The research protocol was approved by the Institutional Review Board of the Zhongshan Ophthalmic Center, Sun Yat-sen University, and the study was performed in accordance with the Declaration of Helsinki. Written informed consent was obtained from at least one parent or guardian of all participating children. The date of registration of the clinical trial was March 16, 2020.

The COVID-19 outbreak was recognized in China during the nationwide winter school vacation. On January 27, 2020, China's Ministry of Education announced the postponement of the 2020 spring semester, and all schools were asked to recommend that students stay at home, avoiding group activities and large gatherings. As the outbreak continued, the Guangdong Education Department announced in February 2020 that secondary schools would begin formal online education beginning March 2, 2020. Teachers at the 14 public secondary schools in Duanzhou district (Figure S1 and Table S1, Multimedia Appendix 1) formulated an online curriculum for homeschooling, with standardized content and schedules for all students in the area (Table S2, Multimedia Appendix 1).

After excluding 2 secondary schools owing to an insufficient number of students per cluster, permission to conduct the study was requested from the remaining public secondary schools (n=12) in the district (Figure S1, Multimedia Appendix 1). Inclusion criteria were as follows: (1) grade 7 (12-13 years old) students in Duanzhou district and (2) under home confinement and enrolled in online learning courses during the COVID-19 outbreak. Exclusion criteria included the presence of disorders such as autism, pervasive developmental delay, and schizophrenia, which might interfere with participation in the intervention. The study followed the CONSORT (Consolidated Standards of Reporting Trials) guideline [22] (Multimedia Appendix 2). The reporting of the mobile-phone–based questionnaires followed the CHERRIES checklist (Checklist for Reporting Results of Internet E-Surveys) [23] (Multimedia Appendix 3).

Randomization, Concealment, and Masking

Cluster randomization with schools as the cluster was adopted in order to maximize peer-to-peer support while avoiding contamination in the use of social media between study groups. A total of 12 schools with a block size of 4 were assigned (1:1) to receive either health education + download of the peer-to-peer live-streaming app (intervention group) or health education only (control). The randomization sequence was generated by an independent statistician using an online random number generator. In each school, 2 classes (on average, 40 students per class) were further randomly selected for participation.

Due to the nature of the intervention, students were not masked. However, masking of investigators was achieved through the exclusive use of electronic, self-administered questionnaires and masking the statistician to group allocation until completion of all analyses. In addition, participating students were informed that their responses to questionnaires would not be made available to their parents or teachers. Qualitative feedback on the app was not obtained from participants.

Procedures

Recruitment and enrollment were assisted by trained teachers, who described the study in detail to participants and their parents over the phone. Participants and their parents were offered the opportunity to ask questions about participation. Before randomization, online seminars for parents were conducted through all 12 participating schools, during which the investigators answered questions and collected consent forms. Teachers were trained and were provided with a manual of operations, which included the study protocol and standard operating procedures for the study.

During the period of home confinement, the Chinese government had already issued a policy recommending a specific schedule for recess and physical activity breaks for students. In the control group, teachers delivered an online health information session covering the following topics and health advice to students: (1) an outline was provided on the recommended 20-20-20 rule during study and viewing of on-screen content; (2) during recess (15 mins for each recess; 4 times per day), participants in the control group received SMS text message prompts (≤50 characters) to participate in broadcast exercise programs at home, eye relaxation, or to stretch for 10 minutes. Students had access to at-home workout videos developed by exercise physiologists. Breaks were part of the online curriculum, and students were instructed by teachers (who were not aware of the study allocation) to rest and take exercise breaks according to government-issued recommendations.

Students in the intervention group received the identical health information session, online curriculum, workout videos, and breaks as described above. Additionally, at the beginning of the study, students in the intervention group were asked to log on and download a peer-to-peer live-streaming app (the Recess and Exercise Advocacy Program [REAP]). REAP is a live-streaming platform that allows users to capture short videos and photographs with their smartphones related to their physical exercise or eye relaxation activities (eg, looking outdoors through the window). The app has been optimized for both iOS (Apple) and Android operating systems. During each recess from homeschooling (15 mins for each recess; 4 times per day), participants in the intervention were prompted by SMS text messaging (≤50 characters) to log in to the REAP app to participate in live streaming or posting their workouts and outdoor videos and photographs (Supplementary Method S1, Multimedia Appendix 1).

The outcomes were measured by a self-assessed survey through mobile-phone–based questionnaires after signing the informed consent form. The questionnaires were surveyed at the beginning of the study and at the 2-week follow-up. At the start of the questionnaire, participants were given the following information: the purpose of the study, the expected length of time of the mobile phone survey, the names of the investigators, and the way and duration the data were stored. To ensure completeness

checks, the participants were asked to complete the mandatory items before the questionnaire was submitted. The study investigators ensured that the confidentiality of the participants' data was preserved. Individual participant data will not be disclosed to outside personnel and will not appear in any publications. Data assistants entered the questionnaires into a database and pseudonymized the data before it passed to the study statistician.

Outcome Measures

The primary study outcome was a change in self-reported anxiety score between baseline and the 2-week follow-up, as measured using the 45-item Chinese version of the Spence Children's Anxiety Scale (SCAS). In addition to the total SCAS score, this measure assesses 6 domains of anxiety: panic/agoraphobia, separation anxiety, social phobia, fears of physical injury, obsessive compulsive problems, and generalized anxiety. Details relating to the SCAS can be found in Multimedia Appendix 1 [24]. This instrument has been well validated in Chinese children [25]. A higher total score (range 0-114) summing the point values of all items indicates greater anxiety.

Secondary outcomes included changes from baseline to the 2-week follow-up in the following items: children's self-reported eye strain, sleep quality, and time spent on different near-work activities. Eye strain was measured with the self-reported Computer Vision Syndrome Questionnaire (CVS - Q), which evaluates the frequency (never, occasionally, or often/always) and intensity (moderate or intense) of 16 symptoms. Details relating to the CVS-Q scale are available in Multimedia Appendix 1 [26]. The 4-item Patient-Reported Outcomes Measurement Information System (PROMIS) pediatric sleep disturbance questionnaire was used to assess sleep disturbance (Multimedia Appendix 1) [27]. Self-reported information regarding the average daily time spent on near-work activities was collected, including each of the following: reading; writing; and use of computers/tablets, smartphones, television, and video games (Multimedia Appendix 1).

In addition, parentally reported anxiety scores on behalf of children were obtained using the SCAS for Parent (SCAS-P) questionnaire (Multimedia Appendix 1). This variable was not prespecified but was used to assess the reliability of self-reported scores from children.

Statistical Analysis

The sample size was calculated based on a cluster randomized design, assuming that children's self-reported raw anxiety score in the intervention group would be reduced by 2.5 points, with a standard deviation of 5 points, and the anxiety score of the control group would not change. The average cluster size was about 80, and an intracluster correlation coefficient of 0.02 was estimated based on the Refractive Error Studies in Children [28]. A sample size of 8 schools (4 in each group) at a two-sided significance level of .05 would give a power of 90%. Assuming a participation rate of 90% and attrition of 20%, a total of 12 schools was required. The sample size was calculated using PASS 16.0 (NCSS, LLC).

The distribution of baseline characteristics was presented as the mean (SD) for continuous variables and frequency (percentage) for categorical variables. The unadjusted mean differences between study groups in 2-week change and the 95% CI for primary and secondary outcomes were calculated using linear regression. The adjusted intervention effect on the primary outcome and 95% CI were estimated using linear regression models, adjusting for baseline measures. The study group and all variables with $P<.20$ in univariable regression analyses were included in the multivariable regression analysis. Histograms and quantile-quantile plots were used to verify the normality assumptions of the *t* test and linear regression models. Square-root transformation was applied to all outcomes due to lack of normality. To satisfy intention-to-treat criteria, we conducted missing data imputation. Multiple imputation was used to create 20 copies of the data, and final results were obtained by averaging these data sets using the rules developed by Rubin [29]. A two-sided *P* value less than .05 was considered statistically significant. All analyses were performed using Stata 15.0 (Stata Corp). The study protocol was registered at ClinicalTrials.gov (NCT04309097) prior to enrollment of the first participant and is available for online access.

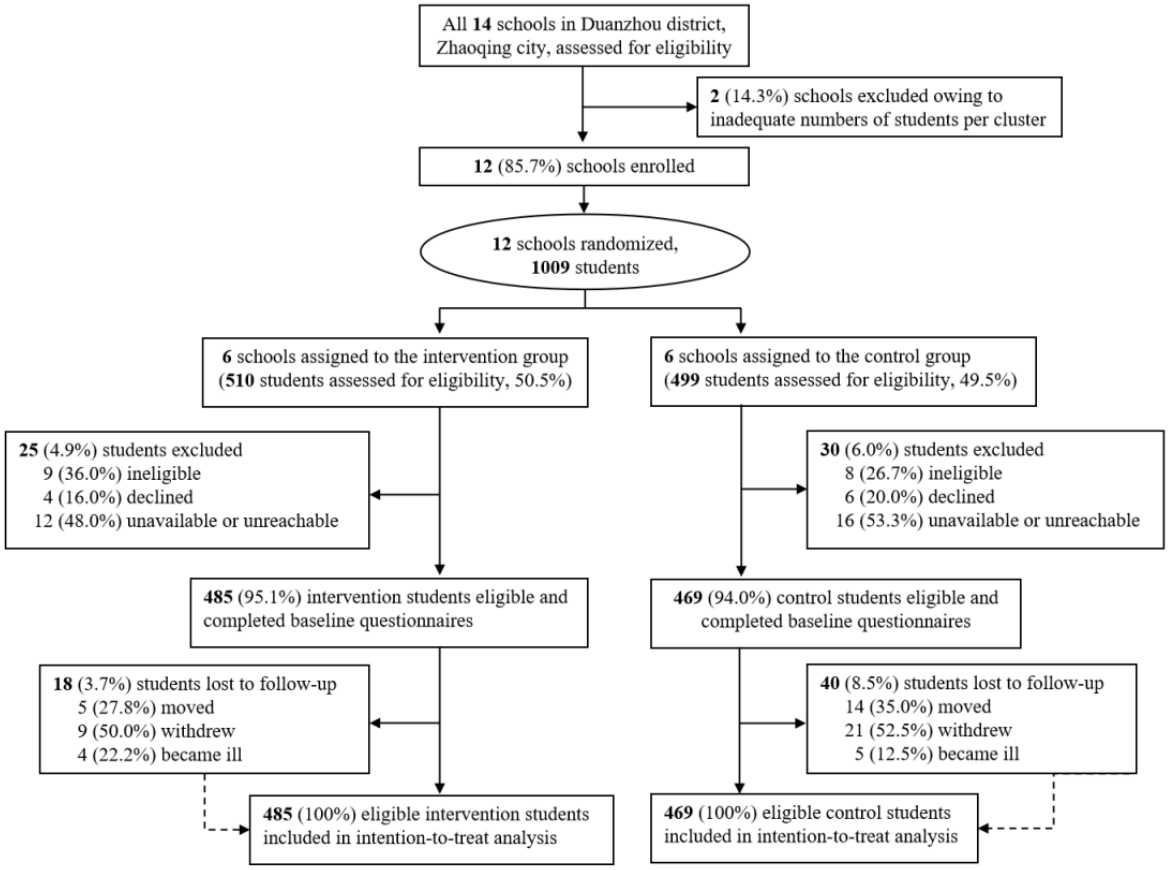
Data-Sharing Statement

Requests for anonymized individual participant data and study documents will be considered on a case-by-case basis and on scientific merit by the principal investigators and approved by the relevant institutional review boards.

Results

A total of 12 schools were randomized (6 to the intervention group and 6 to the control group; Figure 1). Of the 1009 grade 7 students assessed for eligibility (intervention group: n=510, 50.5%; control group: n=499, 49.5%), 55 (5.45%) were unreachable, ineligible, or declined to participate, leaving 954 (94.5%) eligible students, with 485 (95.1%) in the intervention group and 469 (94.0%) as controls. Rolling recruitment occurred on March 16, 2020. During the 2-week follow-up period, 18 (3.7%) and 40 (8.5%) students withdrew from the intervention and control groups, respectively (Figure 1).

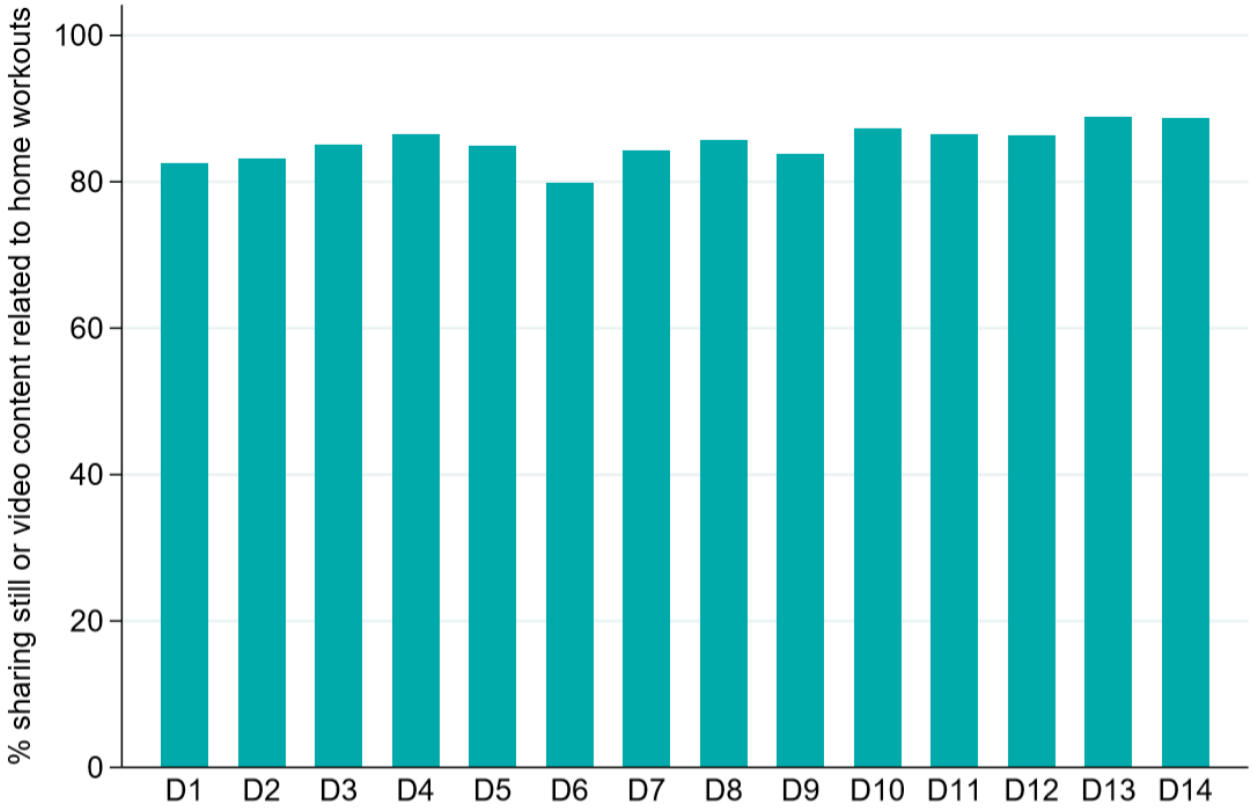
Figure 1. Flow diagram for the trial.



The baseline characteristics of students in the study groups were similar (Supplementary Table S4, Multimedia Appendix 1). The mean age was 13.5 (SD 0.50) years in both groups, and there were no significant baseline differences in sex, use of glasses, parental education and smoking, and family income.

Intervention compliance was monitored using the Cloud platform of the live-streaming app. An average of 1.91 (SD 0.32) videos and photographs of stay-at-home workouts and eye relaxation activities per intervention group student per day were uploaded to the live-streaming app (1.63, SD 0.31, during weekdays; 2.14, SD 0.80, on weekends). All students uploaded at least one video or photograph per day in the intervention group. The app includes the essential feature of a data-monitoring system, and detailed utilization data are presented in Figure 2.

Figure 2. The proportion of participants sharing photos or video content related to home workouts in the intervention group.



The square-root–transformed self-reported anxiety score (main study outcome) fell by –0.23 (95% CI –0.27 to –0.20) in the intervention group and rose (worsened) by 0.12 (95% CI 0.09-0.16) in the controls by the end of the study (Table 1). The change in anxiety score was significantly greater in the intervention group compared to the controls (difference –0.36, 95% CI –0.63 to –0.08; *P*=.02). Significant associations were also found between changes in many SCAS subscale scores and the intervention (Supplementary Table S9, Multimedia Appendix 1). Change in self-reported eye strain was also significantly greater in the intervention vs the control group (intervention group: –0.08, 95% CI –0.10 to 0.06; control group: 0.07, 95% CI 0.05-0.09; difference –0.15, 95% CI –0.26 to –0.03; *P*=.02). The changes in sleep disturbance score (*P*=.23), screen time (*P*=.84), and reading time (*P*=.47) during the 2-week follow-up did not differ significantly between study groups (Table 1).

Table 1. Comparisons of change between study groups in self-reported anxiety, eye strain, sleep score, and time spent on different near-work activities between baseline and the 2-week follow-up (intention-to-treat analysis).

Variable	Intervention group (n=485), mean (95% CI)	Control group (n=469), mean (95% CI)	Difference in change between groups, mean (95% CI)	*P* value[a]
Anxiety score[b,c]				
Baseline	3.72 (3.69 to 3.76)	3.67 (3.64 to 3.70)		
2-week follow-up	3.49 (3.46 to 3.52)	3.79 (3.76 to 3.83)		
Change (follow-up – baseline)	–0.23 (–0.27 to –0.20)	0.12 (0.09 to 0.16)	–0.36 (–0.63 to –0.08)	.02[d]
Eye strain score[b,c]				
Baseline	1.21 (1.19 to 1.23)	1.08 (1.06 to 1.10)		
2-week follow-up	1.13 (1.11 to 1.15)	1.15 (1.12 to 1.18)		
Change (follow-up – baseline)	–0.08 (–0.10 to 0.06)	0.07 (0.05 to 0.09)	–0.15 (–0.26 to –0.03)	.02[d]
Sleep disturbance score[b,c]				
Baseline	2.51 (2.50 to 2.52)	2.53 (2.53 to 2.54)		
2-week follow-up	2.57 (2.56 to 2.58)	2.55 (2.54 to 2.56)		
Change (follow-up – baseline)	0.06 (0.05 to 0.07)	0.01 (0.002 to 0.02)	0.05 (–0.03 to 0.13)	.22
Average daily time spent in near work, hours				
Screen time[b]				
Baseline	2.68 (2.67 to 2.69)	2.69 (2.68 to 2.70)		
2-week follow-up	2.61 (2.60 to 2.62)	2.61 (2.60 to 2.63)		
Change (follow-up – baseline)	–0.07 (–0.08 to –0.05)	–0.08 (–0.09 to –0.06)	0.01 (–0.10 to 0.12)	.84
Reading time[b]				
Baseline	1.37 (1.35 to 1.38)	1.29 (1.28 to 1.30)		
2-week follow-up	1.34 (1.32 to 1.35)	1.21 (1.20 to 1.22)		
Change (follow-up – baseline)	–0.03 (–0.05 to –0.01)	–0.08 (–0.09 to –0.06)	0.05 (–0.09 to 0.18)	.47

[a]Linear regression models adjusting for cluster effects within schools.

[b]Square root transformed.

[c]Higher score indicates greater severity.

[d]Significant.

In linear regression models, randomization to receive the peer-to-peer live-streaming intervention was associated with a significant reduction in self-reported anxiety compared to the controls (β=–0.36, 95% CI –0.63 to –0.08; *P*=.02), after adjusting for sex and household income (Supplementary Table S5, Multimedia Appendix 1). Results based on parentally reported anxiety scores were consistent (Supplementary Table S6, Multimedia Appendix 1).

Discussion

Principal Results

Our findings from a geographically representative region in urban China showed that a digital behavior change intervention effectively reduced anxiety and eye strain during children's homeschooling, without increasing overall screen time. This digital tool might be useful during the COVID-19 pandemic, which has led to widespread school closures, with over a billion children around the world receiving online homeschooling, potentially increasing risk of anxiety and eye strain [10,30,31].

Comparison With Prior Work

A recent review suggested that children's mental and visual health may suffer during home confinement and online schooling [32]. These effects may result from reduced face-to-face social interaction, persistent and intense screen exposure, prolonged time spent on near-work activities, and the disruption of normal life rhythms, all of which may contribute to increased anxiety, eye strain, and sleep disturbance [33]. Children with preexisting mental health conditions such as depression may be especially susceptible to such problems [7].

The app only permitted users to post content related to exercise and activities promoting eye relaxation. This focus on activities highly relevant to both parents and children during the COVID-19 lockdown may explain the high study participation rate of over 95% and the large counts of children who shared photos or video content on a daily basis (Figure 2). We did not

record any adverse events related to our intervention, nor did we observe any significant increase in screen time in the intervention group (Table 1). This latter finding is significant in view of concerns that children's use of smartphones and social media may increase the risk of attention-deficit disorder, sleep disturbance, obesity, and other conditions [18,34,35]. Further studies are needed to investigate the long-term impacts of our digital intervention on children's social media behavior.

Strengths and Limitations

Strengths of the current study include the fact that participants were recruited from a population-representative sample of schools in a typical Chinese urban region, allowing our findings to be generalized to a wider range of Chinese children affected by the COVID-19 lockdown. Second, because of social media's reliance on peer support, our protocol followed a cluster randomized design, which minimized the risk of cross-contamination in social media use between study groups. Thirdly, children in both study groups received the same online courses, ensuring intergroup homogeneity except with respect to the intervention. Finally, the results of primary, secondary, and sensitivity analyses were in full agreement; for example, there was concordance between self- and parentally reported anxiety scores (Supplementary Table S7, Multimedia Appendix 1), indicating the robustness and reliability of our findings.

Study limitations must also be acknowledged. First, masking of participants was not feasible due to the nature of the intervention, so the possibility of placebo effects cannot be excluded. However, good agreement between parental and self-report of anxiety makes this somewhat less likely. Second, data for the main outcome were obtained by questionnaires and were thus based on an evaluation of symptoms, rather than a clinical diagnosis of anxiety disorder. Regardless, the reproducibility of this questionnaire has been well validated [21], and, as mentioned above, parental and self-reported scores were in accord (Supplementary Table S7, Multimedia Appendix 1). Third, the study design allowed for only 2 weeks of follow-up. Previous studies have shown that interventions of this length can be effective [36-38], and such brief periods of engagement may be particularly applicable to unpredictable, short-term school closures [39], which may become more common in the future with additional waves of the COVID-19 pandemic. For example, a brief (2-3 week) lay counsellor–delivered, problem-solving intervention was found to be effective for adolescents with diverse mental health problems [38]. Nevertheless, it is certain that a longer follow-up will provide additional information on the robustness and stability of this intervention effect. Fourth, only Chinese children were included in the study, and the effectiveness of live-streaming interventions to reduce children's anxiety and eye strain must be validated in other settings. Fifth, due to the nature of our behavior intervention, we could not exclude the possibility that the changes in children's anxiety and eye strain may have been caused by other factors (eg, the increased interactions between children and parents), rather than a change in the behavior itself. In addition, the questionnaires were not fully validated for online use, which is the only means that can be used during COVID-19–related home learning. Finally, the urban sample used in the study may not be representative of China, not only because of the lack of smartphones in less developed parts of China but also due to the higher probability of separation between children and their parents due to labor migration. Our results therefore need to be interpreted with caution.

Conclusions

In summary, this study demonstrated that a novel digital behavior change intervention can significantly reduce anxiety and eye strain in homeschooled children. Further work is needed to determine whether these results can be extended to other aspects of pediatric and adult health, and to different geographic and cultural settings.

Acknowledgments

We are grateful to the Clinical Research Center of the Zhongshan Ophthalmic Center, all staff at the Zhaoqing Education Bureau, and the involved schools for providing coordination and field workers for data collection.

This study was supported by the Construction Project of High-Level Hospitals in Guangdong Province (303020107, 303010303058); the National Natural Science Foundation of China (81530028, 81721003); the Clinical Innovation Research Program of Guangzhou Regenerative Medicine and Health Guangdong Laboratory (2018GZR0201001); the research units of Ocular Development and Regeneration, Chinese Academy of Medical Sciences (2019-I2M-5-005); the Local Innovative and Research Teams Project of the Guangdong Pearl River Talents Program (2017BT01S138); and the State Key Laboratory of Ophthalmology, Zhongshan Ophthalmic Center, Sun Yat-sen University. NC is supported by the Ulverscroft Foundation (United Kingdom).

The funding sources did not play any role in the design and conduct of the study; in the collection, management, analysis, or interpretation of the data; in the preparation, review, approval of the manuscript, or decision to submit the manuscript for publication. The views and opinions expressed in this paper are those of the authors and do not necessarily reflect those of the funders of the study.

Authors' Contributions

Y Zheng, NC, WW, and Y Zhong had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. YL, NC, and Y Zheng were responsible for concept and design. All authors were involved in acquisition, analysis, and interpretation of data. Y Zheng, NC, and WW drafted the manuscript. All authors were involved in

critical revision of the manuscript for important intellectual content. LJ conducted statistical analysis. YL was responsible for securing funding. Y Zhong, LX, ZZ, and HL provided administrative, technical, or material support. Y Zheng, NC, and YL were responsible for supervision. All authors approved the final version.

Conflicts of Interest

YL reports receiving grants from the National Natural Science Foundation of China. Y Zheng has served on the digital advisory board for Novartis and is the owner of the REAP app. NC is the director of research for Orbis International, a nongovernmental organization that carries out children's eye health work in China. The remaining authors have no disclosures to report.

Multimedia Appendix 1
Supplementary content, tables, and figures.
[DOC File , 1700 KB - jmir_v23i4e24316_app1.doc]

Multimedia Appendix 2
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 2157 KB - jmir_v23i4e24316_app2.pdf]

Multimedia Appendix 3
CHERRIES checklist.
[PDF File (Adobe PDF File), 105 KB - jmir_v23i4e24316_app3.pdf]

References

1. COVID-19 case tracker. Coronavirus Resource Center, Johns Hopkins University. URL: https://coronavirus.jhu.edu [accessed 2020-05-10]
2. COVID-19 and its implications for protecting children online. United Nations Educational, Scientific and Cultural Organization (UNESCO). 2020 Apr. URL: https://www.unicef.ca/sites/default/files/2020-04/PROTECTING_CHILDREN_ONLINE.pdf [accessed 2020-05-10]
3. Co-SPACE (COVID-19 Supporting Parents, Adolescents, and Children in Epidemics). Department of Experimental Psychology, University of Oxford. URL: https://oxfordxpsy.az1.qualtrics.com/jfe/form/SV_3VO130LTKOcloMd [accessed 2020-05-10]
4. Lim JM, Tun ZM, Kumar V, Quaye SED, Offeddu V, Cook AR, et al. Population anxiety and positive behaviour change during the COVID-19 epidemic: Cross-sectional surveys in Singapore, China and Italy. Influenza Other Respir Viruses 2021 Jan;15(1):45-55 [FREE Full text] [doi: 10.1111/irv.12785] [Medline: 32889784]
5. Hou F, Bi F, Jiao R, Luo D, Song K. Gender differences of depression and anxiety among social media users during the COVID-19 outbreak in China:a cross-sectional study. BMC Public Health 2020 Nov 04;20(1):1648 [FREE Full text] [doi: 10.1186/s12889-020-09738-7] [Medline: 33148202]
6. Sprang G, Silman M. Posttraumatic stress disorder in parents and youth after health-related disasters. Disaster Med Public Health Prep 2013 Feb;7(1):105-110. [doi: 10.1017/dmp.2013.22] [Medline: 24618142]
7. Wang G, Zhang Y, Zhao J, Zhang J, Jiang F. Mitigate the effects of home confinement on children during the COVID-19 outbreak. The Lancet 2020 Mar 21;395(10228):945-947 [FREE Full text] [doi: 10.1016/S0140-6736(20)30547-X] [Medline: 32145186]
8. Sun S, Folarin AA, Ranjan Y, Rashid Z, Conde P, Stewart C, RADAR-CNS Consortium. Using Smartphones and Wearable Devices to Monitor Behavioral Changes During COVID-19. J Med Internet Res 2020 Sep 25;22(9):e19992 [FREE Full text] [doi: 10.2196/19992] [Medline: 32877352]
9. Saha K, Torous J, Caine ED, De Choudhury M. Psychosocial Effects of the COVID-19 Pandemic: Large-scale Quasi-Experimental Study on Social Media. J Med Internet Res 2020 Nov 24;22(11):e22600 [FREE Full text] [doi: 10.2196/22600] [Medline: 33156805]
10. Jaiswal S, Asper L, Long J, Lee A, Harrison K, Golebiowski B. Ocular and visual discomfort associated with smartphones, tablets and computers: what we do and do not know. Clin Exp Optom 2019 Sep;102(5):463-477. [doi: 10.1111/cxo.12851] [Medline: 30663136]
11. Stockwell S, Schofield P, Fisher A, Firth J, Jackson SE, Stubbs B, et al. Digital behavior change interventions to promote physical activity and/or reduce sedentary behavior in older adults: A systematic review and meta-analysis. Exp Gerontol 2019 Jun;120:68-87. [doi: 10.1016/j.exger.2019.02.020] [Medline: 30836130]
12. Stubbs B, Koyanagi A, Hallgren M, Firth J, Richards J, Schuch F, et al. Physical activity and anxiety: A perspective from the World Health Survey. J Affect Disord 2017 Jan 15;208:545-552. [doi: 10.1016/j.jad.2016.10.028] [Medline: 27802893]
13. Merikangas KR, Swendsen J, Hickie IB, Cui L, Shou H, Merikangas AK, et al. Real-time Mobile Monitoring of the Dynamic Associations Among Motor Activity, Energy, Mood, and Sleep in Adults With Bipolar Disorder. JAMA Psychiatry 2019 Feb 01;76(2):190-198 [FREE Full text] [doi: 10.1001/jamapsychiatry.2018.3546] [Medline: 30540352]

14. Mead MP, Baron K, Sorby M, Irish LA. Daily Associations Between Sleep and Physical Activity. Int J Behav Med 2019 Oct;26(5):562-568. [doi: 10.1007/s12529-019-09810-6] [Medline: 31372835]
15. Kawashima M, Uchino M, Yokoi N, Uchino Y, Dogru M, Komuro A, et al. The Association between Dry Eye Disease and Physical Activity as well as Sedentary Behavior: Results from the Osaka Study. J Ophthalmol 2014;2014:943786 [FREE Full text] [doi: 10.1155/2014/943786] [Medline: 25485144]
16. Sfendla A, Hadrya F. Factors Associated with Psychological Distress and Physical Activity During the COVID-19 Pandemic. Health Secur 2020 Dec;18(6):444-453. [doi: 10.1089/hs.2020.0062] [Medline: 32946286]
17. The Lancet Digital Health. Child and adolescent health in the digital age. The Lancet Digital Health 2020 Mar;2(3):e102 [FREE Full text] [doi: 10.1016/S2589-7500(20)30029-7] [Medline: 33328079]
18. Odgers C. Smartphones are bad for some teens, not all. Nature 2018 Feb 22;554(7693):432-434 [FREE Full text] [doi: 10.1038/d41586-018-02109-8] [Medline: 29469108]
19. Wang H, Zhong J, Hu R, Fiona B, Yu M, Du H. Prevalence of high screen time and associated factors among students: a cross-sectional study in Zhejiang, China. BMJ Open 2018 Jun 19;8(6):e021493 [FREE Full text] [doi: 10.1136/bmjopen-2018-021493] [Medline: 29921687]
20. 2017 China Statistical Yearbook. National Bureau of Statistics of China. URL: http://www.zhaoqing.gov.cn/attachment/0/50/50130/838219.zip
21. Weissberger A. China has more than 1.6 billion mobile subscribers (>China's entire population!). IEEE Communications Society. 2020 Jan 23. URL: https://techblog.comsoc.org/2020/01/23/china-has-more-than-1-6-billion-mobile-subscribers-chinas-entire-population/ [accessed 2020-05-10]
22. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res 2011 Dec 31;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]
23. Eysenbach G. Improving the quality of Web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). J Med Internet Res 2004 Sep 29;6(3):e34 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
24. Spence SH. A measure of anxiety symptoms among children. Behav Res Ther 1998 May;36(5):545-566. [doi: 10.1016/s0005-7967(98)00034-5] [Medline: 9648330]
25. Wang M, Meng Q, Liu L, Liu J. Reliability and Validity of the Spence Children's Anxiety Scale for Parents in Mainland Chinese Children and Adolescents. Child Psychiatry Hum Dev 2016 Oct;47(5):830-839. [doi: 10.1007/s10578-015-0615-2] [Medline: 26667807]
26. Seguí MDM, Cabrero-García J, Crespo A, Verdú J, Ronda E. A reliable and valid questionnaire was developed to measure computer vision syndrome at the workplace. J Clin Epidemiol 2015 Jun;68(6):662-673. [doi: 10.1016/j.jclinepi.2015.01.015] [Medline: 25744132]
27. Forrest CB, Meltzer LJ, Marcus CL, de la Motte A, Kratchman A, Buysse DJ, et al. Development and validation of the PROMIS Pediatric Sleep Disturbance and Sleep-Related Impairment item banks. Sleep 2018 Jun 01;41(6). [doi: 10.1093/sleep/zsy054] [Medline: 29546286]
28. Negrel AD, Maul E, Pokharel GP, Zhao J, Ellwein LB. Refractive error study in children: sampling and measurement methods for a multi-country survey. American Journal of Ophthalmology 2000 Apr;129(4):421-426. [doi: 10.1016/s0002-9394(99)00455-9] [Medline: 10764848]
29. Royston P. Multiple Imputation of Missing Values. The Stata Journal 2018 Nov 19;4(3):227-241. [doi: 10.1177/1536867x0400400301]
30. Lee J. Mental health effects of school closures during COVID-19. The Lancet Child & Adolescent Health 2020 Jun;4(6):421 [FREE Full text] [doi: 10.1016/S2352-4642(20)30109-7] [Medline: 32302537]
31. Van Lancker W, Parolin Z. COVID-19, school closures, and child poverty: a social crisis in the making. The Lancet Public Health 2020 May;5(5):e243-e244 [FREE Full text] [doi: 10.1016/S2468-2667(20)30084-0] [Medline: 32275858]
32. Viner RM, Russell SJ, Croker H, Packer J, Ward J, Stansfield C, et al. School closure and management practices during coronavirus outbreaks including COVID-19: a rapid systematic review. The Lancet Child & Adolescent Health 2020 May;4(5):397-404 [FREE Full text] [doi: 10.1016/S2352-4642(20)30095-X] [Medline: 32272089]
33. Golberstein E, Wen H, Miller BF. Coronavirus Disease 2019 (COVID-19) and Mental Health for Children and Adolescents. JAMA Pediatr 2020 Sep 01;174(9):819-820. [doi: 10.1001/jamapediatrics.2020.1456] [Medline: 32286618]
34. Ra CK, Cho J, Stone MD, De La Cerda J, Goldenson NI, Moroney E, et al. Association of Digital Media Use With Subsequent Symptoms of Attention-Deficit/Hyperactivity Disorder Among Adolescents. JAMA 2018 Jul 17;320(3):255-263 [FREE Full text] [doi: 10.1001/jama.2018.8931] [Medline: 30027248]
35. Chen W, Adler JL. Assessment of Screen Exposure in Young Children, 1997 to 2014. JAMA Pediatr 2019 Apr 01;173(4):391-393 [FREE Full text] [doi: 10.1001/jamapediatrics.2018.5546] [Medline: 30776061]
36. Erickson SJ, Gerstle M, Feldstein SW. Brief interventions and motivational interviewing with children, adolescents, and their parents in pediatric health care settings: a review. Arch Pediatr Adolesc Med 2005 Dec;159(12):1173-1180. [doi: 10.1001/archpedi.159.12.1173] [Medline: 16330743]

37. Caldwell DM, Davies SR, Hetrick SE, Palmer JC, Caro P, López-López JA, et al. School-based interventions to prevent anxiety and depression in children and young people: a systematic review and network meta-analysis. The Lancet Psychiatry 2019 Dec;6(12):1011-1020 [FREE Full text] [doi: 10.1016/S2215-0366(19)30403-1] [Medline: 31734106]
38. Michelson D, Malik K, Parikh R, Weiss HA, Doyle AM, Bhat B, et al. Effectiveness of a brief lay counsellor-delivered, problem-solving intervention for adolescent mental health problems in urban, low-income schools in India: a randomised controlled trial. The Lancet Child & Adolescent Health 2020 Aug;4(8):571-582 [FREE Full text] [doi: 10.1016/S2352-4642(20)30173-5] [Medline: 32585185]
39. Ali ST, Cowling BJ, Lau EH, Fang VJ, Leung GM. Mitigation of Influenza B Epidemic with School Closures, Hong Kong, 2018. Emerg Infect Dis 2018 Nov;24(11):2071-2073 [FREE Full text] [doi: 10.3201/eid2411.180612] [Medline: 30334723]

Abbreviations

CHERRIES: Checklist for Reporting Results of Internet E-Surveys
CONSORT: Consolidated Standards of Reporting Trials
CVS-Q: Computer Vision Syndrome Questionnaire
PROMIS: Patient-Reported Outcomes Measurement Information System
REAP: Recess and Exercise Advocacy Program
SCAS: Spence Children's Anxiety Scale
SCAS-P: Spence Children's Anxiety Scale for Parent
UNESCO: United Nations Educational, Scientific and Cultural Organization

Edited by G Eysenbach; submitted 25.09.20; peer-reviewed by S Doan, Z Jin; comments to author 24.01.21; revised version received 31.01.21; accepted 14.04.21; published 30.04.21.

Please cite as:
Zheng Y, Wang W, Zhong Y, Wu F, Zhu Z, Tham YC, Lamoureux E, Xiao L, Zhu E, Liu H, Jin L, Liang L, Luo L, He M, Morgan I, Congdon N, Liu Y
A Peer-to-Peer Live-Streaming Intervention for Children During COVID-19 Homeschooling to Promote Physical Activity and Reduce Anxiety and Eye Strain: Cluster Randomized Controlled Trial
J Med Internet Res 2021;23(4):e24316
URL: https://www.jmir.org/2021/4/e24316
doi:10.2196/24316
PMID:33882021

Original Paper

Emotions of COVID-19: Content Analysis of Self-Reported Information Using Artificial Intelligence

Achini Adikari[1*], BSc; Rashmika Nawaratne[1*], BSc, PhD; Daswin De Silva[1*], BSc, PhD; Sajani Ranasinghe[1*], BSc; Oshadi Alahakoon[2*], BSc, PhD; Damminda Alahakoon[1*], BSc, PhD

[1]Research Centre for Data Analytics and Cognition, La Trobe University, Melbourne, Australia

[2]College of Engineering and Science, Victoria University, Melbourne, Australia

[*]all authors contributed equally

Corresponding Author:
Achini Adikari, BSc
Research Centre for Data Analytics and Cognition
La Trobe University
Kingsbury Drive
Melbourne
Australia
Phone: 61 394793109
Email: A.Adikari@latrobe.edu.au

Abstract

Background: The COVID-19 pandemic has disrupted human societies around the world. This public health emergency was followed by a significant loss of human life; the ensuing social restrictions led to loss of employment, lack of interactions, and burgeoning psychological distress. As physical distancing regulations were introduced to manage outbreaks, individuals, groups, and communities engaged extensively on social media to express their thoughts and emotions. This internet-mediated communication of self-reported information encapsulates the emotional health and mental well-being of all individuals impacted by the pandemic.

Objective: This research aims to investigate the human emotions related to the COVID-19 pandemic expressed on social media over time, using an artificial intelligence (AI) framework.

Methods: Our study explores emotion classifications, intensities, transitions, and profiles, as well as alignment to key themes and topics, across the four stages of the pandemic: declaration of a global health crisis (ie, prepandemic), the first lockdown, easing of restrictions, and the second lockdown. This study employs an AI framework comprised of natural language processing, word embeddings, Markov models, and the growing self-organizing map algorithm, which are collectively used to investigate social media conversations. The investigation was carried out using 73,000 public Twitter conversations posted by users in Australia from January to September 2020.

Results: The outcomes of this study enabled us to analyze and visualize different emotions and related concerns that were expressed and reflected on social media during the COVID-19 pandemic, which could be used to gain insights into citizens' mental health. First, the topic analysis showed the diverse as well as common concerns people had expressed during the four stages of the pandemic. It was noted that personal-level concerns expressed on social media had escalated to broader concerns over time. Second, the emotion intensity and emotion state transitions showed that *fear* and *sadness* emotions were more prominently expressed at first; however, emotions transitioned into *anger* and *disgust* over time. Negative emotions, except for *sadness*, were significantly higher (P<.05) in the second lockdown, showing increased frustration. Temporal emotion analysis was conducted by modeling the emotion state changes across the four stages of the pandemic, which demonstrated how different emotions emerged and shifted over time. Third, the concerns expressed by social media users were categorized into profiles, where differences could be seen between the first and second lockdown profiles.

Conclusions: This study showed that the diverse emotions and concerns that were expressed and recorded on social media during the COVID-19 pandemic reflected the mental health of the general public. While this study established the use of social media to discover informed insights during a time when physical communication was impossible, the outcomes could also contribute toward postpandemic recovery and understanding psychological impact via emotion changes, and they could potentially inform health care decision making. This study exploited AI and social media to enhance our understanding of human behaviors in global emergencies, which could lead to improved planning and policy making for future crises.

(J Med Internet Res 2021;23(4):e27341) doi:10.2196/27341

KEYWORDS

COVID-19; pandemic; lockdown; human emotions; affective computing; human-centric artificial intelligence; artificial intelligence; AI; machine learning; natural language processing; language modeling; infodemiology; infoveillance

Introduction

Overview

The COVID-19 pandemic continues to devastate the world, with more than 80 million infections and more than 1.7 million deaths [1]. It is an unprecedented public health emergency that has prompted most governments to enforce hard borders, strict social distancing, and rigorous quarantine restrictions. The impact and aftermath of these public health emergency measures have resulted in a severe psychological burden for all individuals. The economic and social fallout has affected individuals and communities alike, resulting in mental health disorders and emotional distress that adds to already overwhelmed health care systems and services worldwide [2-4]. Although researchers and authorities are readily investing in vaccinations, social distancing regulations, and health care facilities to eliminate the virus, long-term mental health impacts will also require undivided attention and action to minimize the detrimental effects. Given this context, developing an understanding of the psychological and emotional burden will aid and accelerate postpandemic recovery and enable policy making for future emergencies of this scale.

Social restrictions and physical distancing measures during this pandemic have led to increased use of social media as a medium of communication by individuals and communities [5]. Expressions on social media during such crises are critical and representative of public opinion, as it has become the primary medium of communication for the exchange of information, experiences, and emotions with others who are facing similar challenges [6-8]. Identification of such opinions in the form of topics enables identification of people's past and ongoing concerns throughout the pandemic [9]. Recent studies have shown a rapid growth of social media content focused on informational and emotional sharing as means of emotion regulation, avoiding mental health issues, and adjustments to the quality of life in lockdown [10,11]. This phenomenon has been described as the *social sharing of emotion*, which postulates that individuals who experience emotions are often eager to share and talk about their emotions [12]. Crises similar to the current pandemic may lead to the amplification of expression of emotions within a community or group of people [13-15]. The extent and frequency of sharing depend on the intensity of the emotional episode, and such social sharing occurs as a means of regulating one's emotions. Thereby, the social sharing of emotions during a crisis creates a spreading and escalation of emotions within the group or community impacted by the event, which, in turn, reflect the mental health status of the community.

Given the volume and variety of content being shared on social media, online behaviors have been investigated as a *proxy* for offline human behaviors, with several studies reporting conclusive results and outcomes [16-18]. A massive number of social media conversations represents a cross-section of society that encompasses people's opinions across different demographic dimensions. These emotions and opinions create a pool of untapped self-reported information. Therefore, it is relevant and useful to study the emotions and concerns voiced over such internet-mediated communication to better understand the mental health and emotional well-being during a time of crisis. Several recent studies have investigated the use of social media during the COVID-19 pandemic to discover topics, emotions, and sentiments from social media conversations [19-21] as well as different information sharing behaviors [22]. While the use of social media to understand public opinion is well established, studies suggest that more focus should be given to showcase how social media can be used to improve health knowledge [23].

On this premise, this study aims to investigate concerns and emotions expressed over social media to infer insights on people's mental health as the pandemic progresses. It has been shown that emotions and concerns expressed on social media represent and relate to people's underlying mental health [24-27]. This establishes social media as a lens through which to comprehend people's mental health during a constrained situation such as the current pandemic, where conducting clinical trials is challenging.

We selected Twitter as the test bed social media platform for our experiments, given that it provides fast-paced, frequent, current affairs–focused end-user engagement, in comparison to other social media platforms [28]. We extracted tweets related to the pandemic that were posted from January to September 2020 and specifically focused on an Australian context based on the following reasons. First, Australia was one of the countries where distinct phases of the lockdown and their impact were clearly visible. This enables the temporal analysis of emotions and concerns in each phase to gauge insights into how citizens react and change their emotion-related behaviors as the pandemic progresses. The four phases are (1) prepandemic—when the outbreaks were yet to be declared a global pandemic and no positive cases have been reported in Australia, (2) the first lockdown—social restrictions following the first wave of positive cases locally, (3) easing restrictions—relaxation of social restrictions when the case numbers were brought under control, and (4) the second lockdown—following the emergence of new cases. Next, as Australia is one of the few countries with a sizeable population to successfully suppress both first and second waves of the pandemic [29,30], it was pertinent to study the emotions related to COVID-19 as experienced by the Australian public. Despite the successful management, a recent study mentions that mental health problems were widespread among Australians during the lockdown periods, where one-quarter of the participants in the study showcased mild to moderate symptoms of anxiety or depression [31,32]. The use of social media among the

Australian public is well established for being representative of a broad cross-section of society [33-35], especially during emergencies as a powerful communication tool [36,37]. This allows us to investigate the emotions expressed via virtual platforms in order to determine similar behaviors related to Australian citizens' mental health.

In this study, we focus on the application of artificial intelligence (AI) approaches that have been validated across several studies [18,27,38] to analyze human emotions. The algorithms are used to quantify emotion intensities, detect emotion state transitions, and identify profiles of impacted individuals based on the concerns they have expressed on social media to provide a holistic view of people's mental health. This approach can be described as an ensemble of machine learning algorithms and natural language understanding (NLU) techniques that establish an end-to-end pipeline to extract, process, normalize, analyze, and aggregate self-reported information in its unstructured format into topics and emotions expressed during the key phases of the pandemic. We have also compared the two lockdown periods and the social media expressions in response to restrictions imposed during each lockdown to capture changes in temporal emotions and related concerns.

The rest of the paper is organized as follows. The Methods section presents an overview of the methodology used in this study and the Results section presents the outcomes of the analysis. The paper concludes with the Discussion section, which outlines the implications of the study, limitations, and potential avenues for future research.

Background

The current literature reports many research studies that detect sentiment and emotions from social media using a variety of techniques. Sentiment analysis has been explored using lexical and statistical approaches [39-42] and recent research studies have extended this by using advanced deep learning models that can detect the bipolarity of social media conversations [43-46]. While sentiment analysis classifies posts into negative and positive categories, emotion detection extends this categorization by detecting granular variants of emotions [47]. Many emotion models, such as Ekman's basic emotion model [48], the valence and arousal model [49], and Plutchik's emotion model [50], are being used as the basis for emotion detection. Most of the approaches of emotion detection rely on prelabeled data via crowdsourced annotations and semantics to develop machine learning models [51-53]. Current research studies report the use of many deep learning models, such as word embedding models [46,54], convolutional neural networks [55], and recurrent neural networks [56-58], for emotion classification. Emoticons and emojis have also been used in studies to infer emotions from social media as they represent user-annotated labels for emotion classification [56,59]. In addition to the emotion classification, emotion intensity is also captured via the annotated data sets [52]. A more recent research study reports on the use of the stacked ensemble method for emotion intensity detection that enables the identification of expressions that denote the emotion intensity [60].

However, these approaches are focused on detecting sentiment or emotion from a social media post but do not focus on emotion modeling, which makes it possible to explore underlying changes of intensity, shifts, and patterns of change in emotions and to generate insights by linking to associated human behaviors [61]. The lack of labeled data for emotion detection is also a major drawback in using supervised machine learning approaches. Therefore, it requires a robust method to explore emotions in unstructured, voluminous social media data that does not require prior knowledge about the data. In this study, we use an ensemble of word embedding and NLU techniques to derive emotions from unlabeled social media conversations. As the foundation of the emotion analysis, we use Plutchik's model [50], which comprises eight basic emotions: anger, sadness, joy, trust, anticipation, fear, disgust, and surprise. The emotion classification is based on these eight emotions, and their intensities are mapped to show the strength of each emotion.

Methods

Data

Twitter conversations that were posted from January to September 2020 within Australia or an Australian context were extracted. Popular hashtags related to COVID-19 in Australia were used to query the data sources. Content by news channels and bots was identified by examining an unusual volume of data generated by each user and was eliminated from the analysis. This process yielded 73,000 public conversations by Twitter users in Australia. The extracted data were cleaned, preprocessed, and anonymized before being used in the analysis.

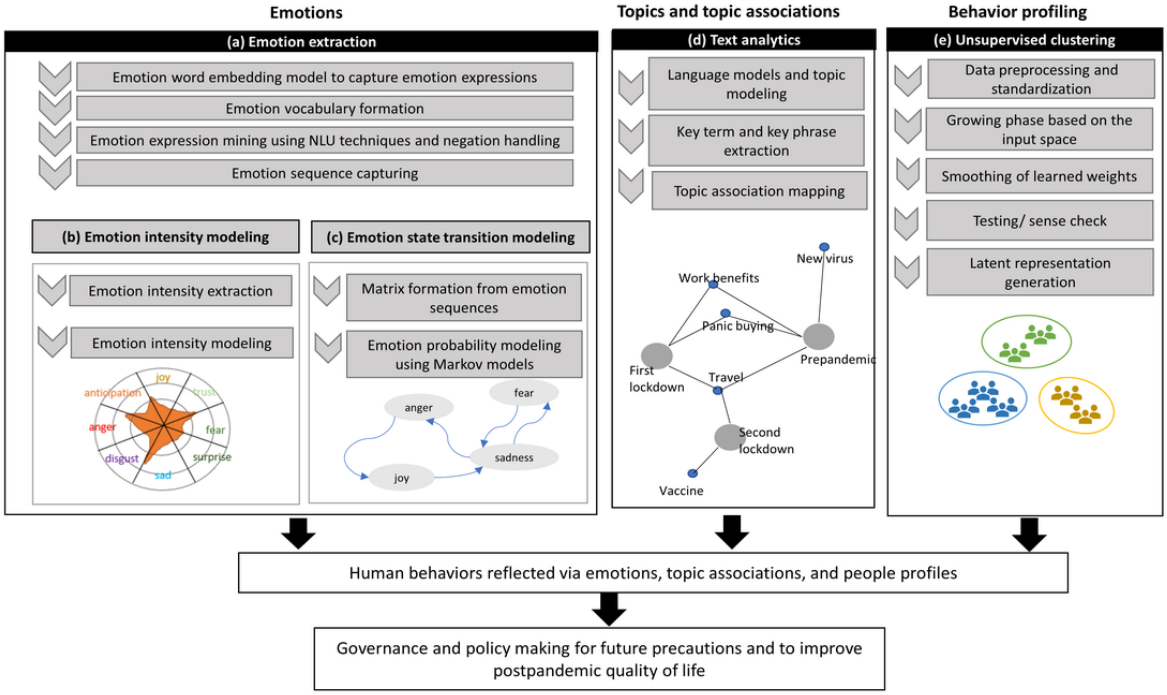
In order to identify different phases of the pandemic, the stages from Table 1 were formed to align with the timeline of the pandemic experience in Australia [62].

Figure 1 presents the high-level view and the workflow of the framework with the respective components.

Table 1. The four main phases of the COVID-19 pandemic from January to September 2020.

Time period (in 2020)	Phase
January to February	Prepandemic
March to May	First lockdown
June	Easing restrictions
July to September	Second lockdown

Figure 1. The high-level view and workflow of the artificial intelligence framework. NLU: natural language understanding.



Emotion Analysis

The emotion extraction, intensity generation, and emotion transition modeling process is shown in the emotions component of the architecture diagram in Figure 1 (a). The emotion analysis was conducted to represent the emotions of each stage as well as the temporal changes in emotions over time. The emotion extraction was carried out on chronologically ordered social media conversations in order to represent emotions over a conversation as a sequence of individual, single-post emotions. First, we trained a word2vec [63] word embedding model, which is a deep learning technique to create vector representations of textual data. This allows the positioning of semantically similar terms together, thereby enabling the finding of closely associated terms for each basic emotion. This querying process yielded a rich vocabulary for representing each emotion, from the basic emotions, which was used as the basis for emotion extraction.

Next, emotion negation was handled to fine-tune the extracted emotions. This process resulted in a sequence of emotions that denoted the chronologically ordered emotions and change of emotions over a conversation. The intensity of emotions in the four stages was used to model an intensity profile for each stage. We have used a modified intensity-capturing algorithm presented in Adikari et al [16] that accommodates the frequency and the presence of expressions that increase or inhibit the valence of the emotion. This allowed for the identification of prominent emotions and their differences in each stage, as seen in Figure 1 (b).

The temporal emotion analysis was conducted by modeling emotion transition models, which models the probabilities of emotion state changes over a course of time. Affective computing research suggests that basic emotions at a particular instance—represented by one post of a tweet in this application—can be regarded as discrete states, and the interactions among these can be modeled to represent the likelihood of the change in the state of emotions [64]. For this purpose, a mathematical model can be used to model the interactions between the emotion states. We have used the sequence of emotions at each stage to generate respective emotion transition models using Markov models [65,66], as seen in Figure 1 (c). The use of emotion transitions to capture individual and group emotions using social media conversations has been successful in a smart city context [17] and for analysis of mental health via online data sources [16]. We have applied this to model the emotion state transitions of each stage, which reflect the flow of emotions as the pandemic progressed.

Topic and Topic Association Analysis

The topic analysis was conducted based on a series of NLU techniques to capture topics of discussions from social media data, as seen in Figure 1 (d). The topics represent people's concerns and, therefore, encompass useful insights on their discourse patterns related to COVID-19 concerns. Outcomes of this analysis help in better interpreting the outcomes of the emotion analysis by providing contextual information for different emotion changes.

To conduct the topic analysis, the data set was first cleaned and processed in order to remove weblinks, stop words, news links, and repeating content generated by bots. This cleaned data set was then split into each stage and taken for the topic analysis. Given the large amount of content, an unsupervised topic modeling technique was used, which groups similar terms based on latent Dirichlet allocation [67]. Once the prominent topic clusters were identified, a trained word2vec embedding model and an automated keyphrase extraction algorithm [68] were used to identify similar topics and subtopics. The outcomes generated the most frequently mentioned terms as well as semantically prominent terms, which allowed us to understand a discrete topic from the terms. The analysis was conducted

separately for the four stages to understand which topics were prominent in each stage; thereafter, a topic association map was created to identify common topics across stages to demonstrate people's continuing concerns.

Behavior Profiling

Behavior profiling was conducted to identify different groupings of citizens in the community based on their emotions and concerns expressed over social media. Identification of such profiles can form a visualization of different profiles that exist in society. This representation is useful to better understand different citizen needs and can inform decision making for relevant authorities.

For profiling, we used an improved variant of the growing self-organizing map (GSOM) algorithm [69]. The GSOM presents a map topography that self-structures by adapting its size and shape based on the attributes and variations of input data. The GSOM algorithm utilizes competition and correlative learning based on the data provided. As input data are presented, nodes of the network compete among each other for ownership of the input, and the winners strengthen their relationships with this input. The competitive learning process is repeated for the complete data set for several cycles, and ultimately the map associates output nodes with patterns in the input data set. The GSOM acts as a knowledge map and can discard outdated information and overfitting of knowledge during the knowledge acquisition process [70,71] that allows the retention of relevant information only. The primary reason to incorporate such mature self-structuring AI was to discover hidden representations of behavior profiles based on their discourse and emotion patterns, which is otherwise challenging to interpret from a large amount of data. The ability of the GSOM algorithm to uncover patterns among data without prior training or supervision makes it possible to detect different groups of behaviors present among the citizens.

Ethics Statement

We have obtained ethics approval for this research from the La Trobe University Human Ethics Committee.

Results

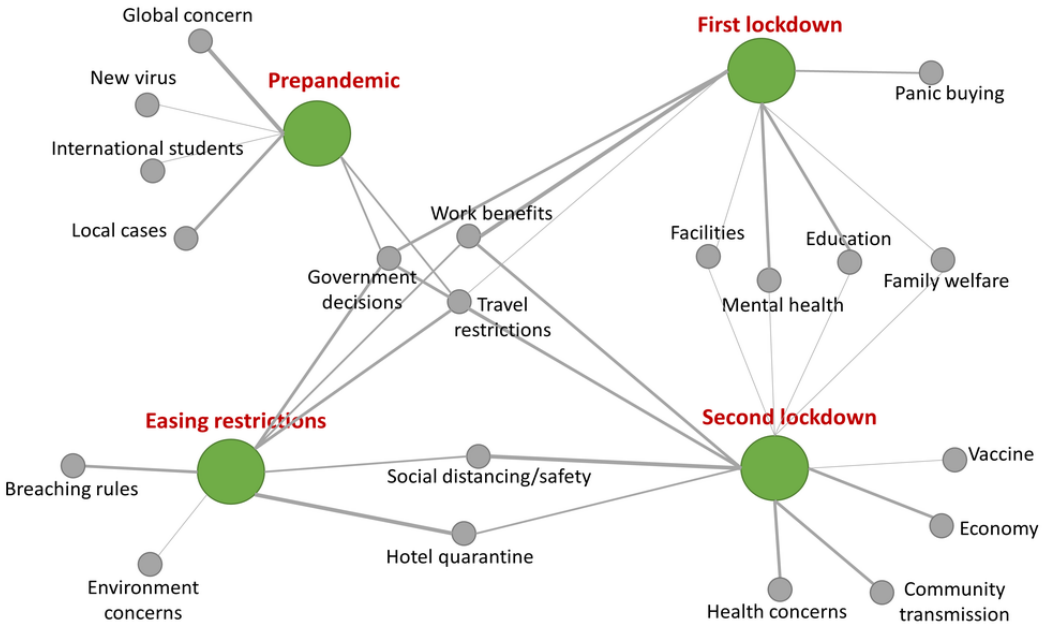
Topics and Topic Associations

Social media contains a high volume of expressions by individuals and groups alike. However, most of these expressions can be aggregated into a finite number of themes and topics [72]. We conducted a topic analysis and a topic association analysis to explore the distinct and common themes across the four stages of the pandemic. Table 2 presents the topics discussed during the four stages of the pandemic, from January to September 2020. The most prominent topics in each stage are presented in Table 2, where the quantitative measure indicates the weight of each topic based on the volume. The topics were also associated across stages as depicted in Figure 2. Key observations are that *government decisions* and *travel restrictions* were discussed during all four stages, and both lockdown stages expressed similar topics oriented toward lifestyle factors. These common topics included *education*, *mental health*, *family welfare*, and *facilities*. *Panic buying* was only mentioned during the first lockdown and was entirely absent during the second lockdown. During the easing restrictions stage, people mostly talked about *social distancing and safety* as well as *hotel quarantine*, which were common topics during the second lockdown as well. Apart from the common topics, the second lockdown mainly focused on *health care*, *community transmission*, *economy*, and *vaccine*. These denote that the conversations were more informed and regulated by the second lockdown. It should be noted that only conversations with prominent topics were included in the analysis and that general conversations have been omitted from the topic analysis. The outcomes of this analysis also aid in comprehending the emotions by providing contextual information related to each stage.

Table 2. Topics of discussion across the four stages of the pandemic.

Pandemic stage and main themes	Volume of conversations, n (%)
Prepandemic (n=15,302)	
Global concern	6021 (39.35)
Local cases	3522 (23.02)
Travel restrictions	2518 (16.46)
Government decisions	2127 (13.90)
International students	588 (3.84)
New virus	526 (3.44)
First lockdown (n=23,201)	
Work benefits	6116 (26.36)
Education	5832 (25.14)
Government decisions	3765 (16.23)
Travel restrictions	3069 (13.23)
Mental health	2568 (11.07)
Family welfare	1176 (5.07)
Facilities	413 (1.78)
Panic buying	262 (1.13)
Easing restrictions (n=11,801)	
Government decisions	2811 (23.82)
Travel restrictions	2700 (22.88)
Hotel quarantine	2294 (19.44)
Breaching rules	1554 (13.17)
Work benefits	1184 (10.03)
Social distancing and safety	925 (7.84)
Environment concerns	333 (2.82)
Second lockdown (n=16,198)	
Social distancing and safety	2910 (17.96)
Health care	2112 (13.04)
Work benefits	2046 (12.63)
Government decisions	1902 (11.74)
Community transmission	1793 (11.07)
Facilities	897 (5.54)
Economy	886 (5.47)
Hotel quarantine	867 (5.35)
Education	841 (5.19)
Family welfare	829 (5.12)
Mental health	489 (3.02)
Vaccine	381 (2.35)
Travel restrictions	245 (1.51)

Figure 2. Topic associations between the four stages of the pandemic.



Emotion Classification and Intensities

The AI approach used for the study of emotions expressed during the pandemic began with the classification of emotions, based on the widely cited psychological emotion model proposed by Plutchik [50]. Classification of emotions was followed by the quantification of the intensity of each of the eight emotions as they were expressed during the four stages.

The outcomes of the emotion analysis suggested that both *sadness* and *fear* emotions were the most intense during the prepandemic stages. This is indicative of the central emotional response during a pandemic being *fear* [2]. In addition, it is said that given the instinctive defensive systems for combating ecological threats, humans experience negative emotions resulting from a threat that can be contagious. The high intensity of *sadness* and *fear* identified during the announcement stage of the pandemic exhibits this emotional pattern. This behavior was also confirmed by the emotion transition model in this stage, where higher probabilities were observed for these emotions. Furthermore, *fear* is also associated with *panic*, where people act blindly and excessively out of self-preservation, possibly endangering the survival of others [2]; this was experienced via incidents like *panic buying*, where people unreasonably stocked up on essential items for impending self-isolation. The surfacing of fear and sadness emotions represents increased anxiety, which impacts people's mental health [73].

The temporal emotion analysis showed that, out of the negative emotions, *anger* and *disgust* were more strongly expressed toward the latter part of the pandemic due to the security breaches and restrictions. Terms such as "covidiot," "quarantine issues," "restrictions," and "another lockdown" were mostly mentioned in the Twitter conversations related to these emotions. This indicates that starting from the negativity of fear and sadness, public emotions have transferred to more anger and disgust at a later stage.

Positive emotion expressions were also captured by the analysis. It was noted that *joy* was more strongly expressed during the easing restriction stage and least during the second lockdown stage. This behavior aligned with real-world behaviors, as it was expected that people would express disappointment for having to experience another lockdown. The emotions *trust* and *anticipation* demonstrated higher intensities during the first lockdown and easing restriction stages; however, they have subsided over time (Figure 3).

We conducted a comparative study based on the two lockdown stages to identify the differences in emotion intensity (Table 3). The differences were compared using the Pearson chi-square test [74,75], and the confidence intervals were calculated based on the recommended method given by Altman et al [76]. A significance level of $P<.05$ was used to determine if the differences in the intensities were significant.

The comparison of negative emotions shows that, except for *sadness*, all other negative emotion intensities were significantly higher in the second lockdown ($P<.05$). This evidently demonstrates the increased disappointment and negativity of having to experience a second lockdown. The emotion *sadness* was persistent in both lockdowns.

The positive emotion comparison shows that, apart from *anticipation*, all other positive emotions were strongly expressed in the first lockdown ($P<.05$). This suggests that people had elicited more positive thoughts during the first lockdown compared to the second. The increase in the intensity of negative emotions and the reduction of positive emotions with further lockdowns confirm the increased levels of distress and deterioration of mental health. Based on this, relevant strategies should be initiated in order to eliminate the spread of negativity among people in further lockdowns.

Figure 3. Emotion intensity fluctuations over time.

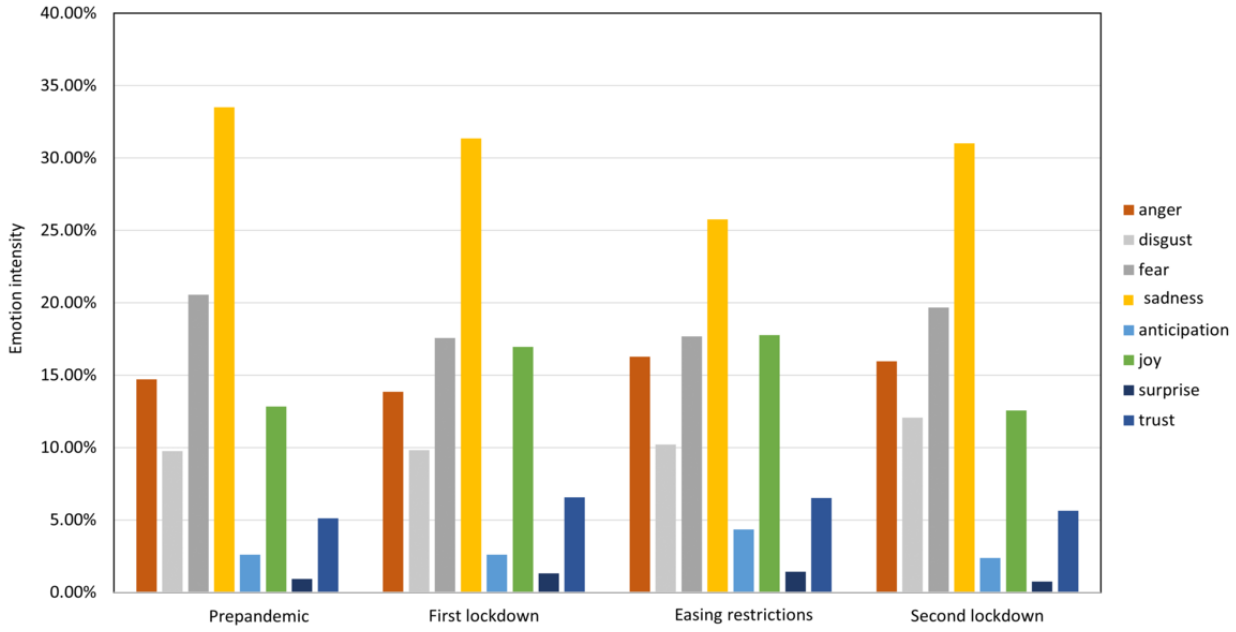


Table 3. Comparison of emotion intensities between the first and second lockdowns.

Emotion	Normalized emotion intensity during the first lockdown (n=36,317), %	Normalized emotion intensity during the second lockdown (n=10,604), %	Difference, % (95% CI)	*P* value
Negative emotions				
Anger	13.84	15.97	2.13 (1.3567 to 2.9218)	<.001
Sadness	31.36	31.01	0.35 (–0.6563 to 1.3461)	.49
Disgust	9.81	12.06	2.25 (1.5691 to 2.9518)	<.001
Fear	17.58	19.66	2.08 (1.2365 to 2.9399)	<.001
Positive emotions				
Joy	16.94	12.56	4.38 (3.6308 to 5.1097)	<.001
Surprise	1.32	0.74	0.58 (0.3662 to 0.7708)	<.001
Trust	6.56	5.63	0.93 (0.4108 to 1.4260)	<.001
Anticipation	2.59	2.37	0.22 (–0.1257 to 0.5406)	.21

Emotion Transitions

Mapping and quantifying the transition of emotions over conversations can be useful to predict the likelihood of emotion state changes as well as to observe temporal change in emotions and intensities over time. In a social media setting, this represents collective emotions in a community over time. The emotion state transitions were modeled separately for each stage, which enabled temporal analysis of emotions from the prepandemic stage to the second lockdown stage. For generating emotion transitions, the emotion space was defined as a set of discrete emotions—anger, fear, sadness, disgust, joy, anticipation, trust, and surprise—from the emotion extraction described earlier. Based on the temporal emotion sequence extracted from the emotion extraction module, a matrix was formed with the frequency of transitions of emotions from one state to another, and the matrix was used to create the emotion state transition diagrams shown in Figure 4; the shift and emergence of emotions are demonstrated via the arrows connecting the emotion states. The probabilities denote the likelihood of changed emotion states (Table 4).

Transitions at each stage demonstrate the temporal analysis of emotions during the pandemic. This emotion flow is useful when determining how people's emotions have changed over time. Health care practitioners and the government can use this information and analysis to understand the mental health status of the population.

Regarding the outcomes, it was observed that most of the emotion transitions involved the *sadness* state in the prepandemic stage. *Sadness to fear* was also prominent relative to other emotion changes. By the first lockdown, there was a shift toward positive emotions, as the emotion states *anticipation* and *joy* were seen to be prominent. This pattern deviated slightly

in the third stage, where positive emotions had amplified and negative emotions had reduced. The emotion propagation made a drastic change in the second lockdown, where positive emotion behaviors had decreased and negative emotions had been amplified, eliciting more tendencies toward *sadness*, *fear*, and *disgust*. This propagation of emotions denoted the emotion flow among people as they voiced their thoughts on social media.

The emergence and shift of emotions over different points in time were seen to be associated with the different stages across different time points in the COVID-19 pandemic. It is apparent that people experienced diverse emotions and had expressed them on social media reflecting their individual interests, opinions, and priorities. Further analysis was conducted using profiling techniques to investigate the intensity (ie, valence) of these emotions and the variation across different stages, which provided insights into who these people were and their priorities and opinions during the pandemic.

Figure 4. Emotion transitions of the four stages of the pandemic. Each arrow represents a transferring of one emotional state to another. The numbers on the arrows represent the likelihoods (ie, probabilities) of changed emotion states.

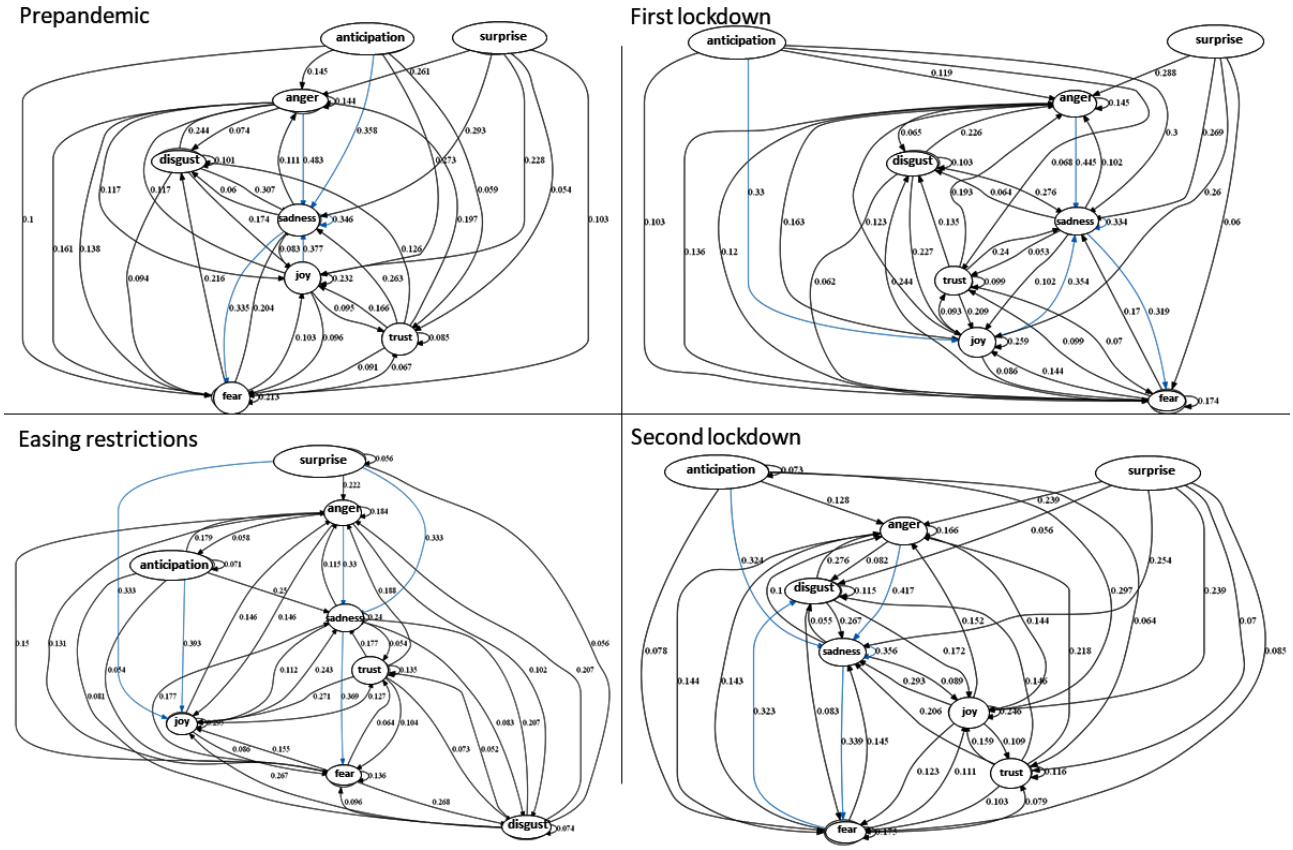


Table 4. Top five emotion transitions and probabilities for each pandemic stage.

Emotion transition	Probability
Prepandemic	
Anger → sadness	0.483
Joy → sadness	0.378
Anticipation → sadness	0.358
Sadness → sadness	0.346
Sadness → fear	0.335
First lockdown	
Anger → sadness	0.445
Joy → sadness	0.354
Sadness → sadness	0.334
Anticipation → joy	0.329
Sadness → anticipation	0.288
Easing restrictions	
Anticipation → joy	0.393
Surprise → joy	0.389
Sadness → fear	0.362
Anger → sadness	0.340
Joy → joy	0.295
Second lockdown	
Anger → sadness	0.417
Sadness → sadness	0.357
Sadness → fear	0.341
Anticipation → sadness	0.324
Fear → disgust	0.322

Group Profiles Based on Topics and Emotions

The GSOM algorithm was used to generate profiles of individuals based on topics and emotions. This enabled the identification of different clusters of citizens that existed within the community in terms of their concerns and emotions. Figures 5 and 6 illustrate the profiles identified for the period of the first and second lockdowns, respectively. Seven profiles were identified for the first lockdown based on the topics and emotions expressed by individuals, and they were labeled based on the most prominent topics that have been discussed. The profiles were (1) children's education, (2) family-oriented discussions, (3) work concerns, (4) panic buying, (5) lifestyle-oriented discussions, (6) travelers and traveling, and (7) higher education–focused discussions.

These profiles demonstrated a distinct focus in their conversations, which resulted in grouping those individuals together. Among these profiles, the *panic buying* cluster demonstrated higher levels of *fear*, whereas *sadness* was prominent in all clusters. In addition, the *family* cluster demonstrated more *joy*, which can be aligned with spending more time with family.

The second lockdown profiles shifted slightly from the profiles identified in the first lockdown. The identified behavior profiles were (1) government decisions, (2) health and mental health concerns, (3) community transmission, (4) safety measures, and (5) work concerns.

Compared to the first lockdown, the profiles of the second lockdown exhibited more negative emotions, where increased levels of negative emotions can be seen in health and mental health concerns. It is noteworthy to observe that the profiles formed in the second lockdown were more focused toward broader aspects of the pandemic when compared to the more lifestyle-oriented profiles in the first lockdown. This difference indicates that as the pandemic progressed, people's concerns shifted more toward broader aspects than concerns at the early stages, which may have been normalized with time.

Figure 5. Behavior profiles in the first lockdown. The mean strength of each concern is denoted as a blue line. GSOM: growing self-organizing map.

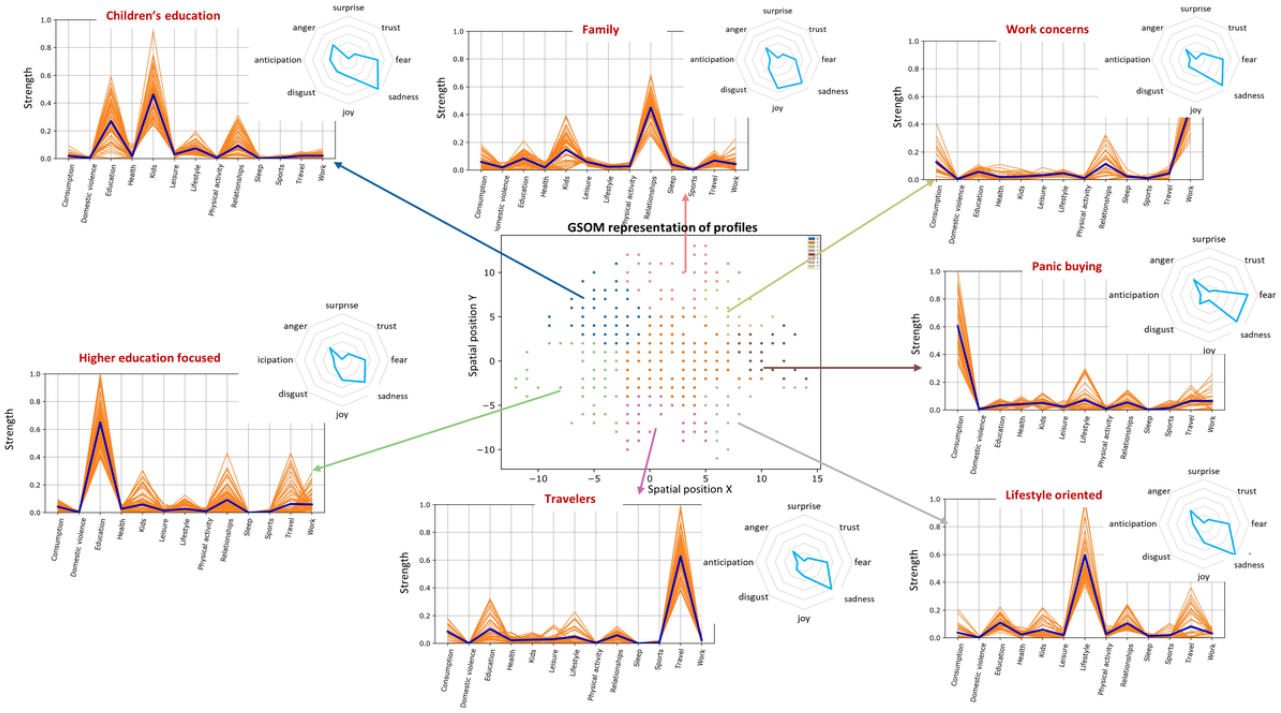
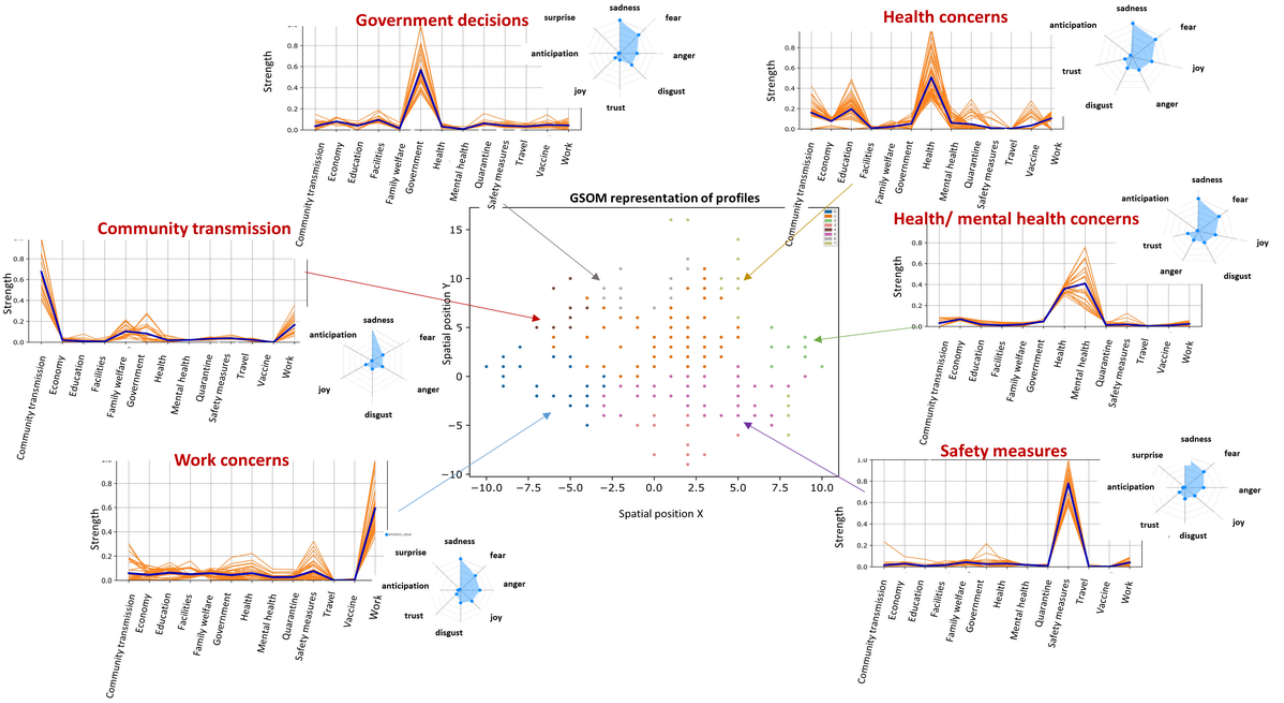


Figure 6. Behavior profiles in the second lockdown. The mean strength of each concern is denoted as a blue line. GSOM: growing self-organizing map.



Discussion

Principal Findings and Implications

At a high level, the themes and topics that were seen at the individual level in the first stages of the pandemic had cascaded into broader and informed topics during the second lockdown. Individuals frequently discussed matters related to family, panic buying, and facilities during the first stage and transitioned into generic topics, such as the vaccine and economy, in the latter stages. Besides, several common concerns persisted over the course of time. Identification of such behaviors related to topics and topic associations from social media conversations can inform public health authorities regarding necessities and the issues that disrupt people's quality of life during this global pandemic. However, it is crucial to investigate the underlying emotions in these conversations to gauge the status of people's mental health.

Limitations and Future Work

One of the limitations of this study is the use of a single data source. As Twitter users are not representative of the whole of Australia, we believe that this could pose a limitation when modeling the emotions and concerns of the Australian population. However, given the importance of Twitter as a social media platform and the substantial amount of data retrieved (73,000 conversations), we were able to capture a broad cross-section of the community. As future work, the study could be extended by using multiple data sources, such as online forums and other social media channels. This will create a holistic view of the community. Moreover, the investigation could be conducted using data from another country, as this would provide different insights into how people have reacted toward different restriction measures and pandemic stages.

Comparison With Prior Work

Our findings regarding emotions are consistent with and align with public health studies related to the COVID-19 pandemic. The emergence of the *sadness* and *fear* emotions during the initial stages resemble the increased anxiety and depression faced by Australians in the first few months of the pandemic [31,32]. Past research indicates that public health emergencies often trigger negative emotions, as people tend to develop aversion, anxiety, and fear [77,78]. The emotion *fear* is often associated with such emergencies and pandemics, as it is one of the core emotions linked with survival. It has been reported that the human mind's instinctive defensive systems for combating ecological threats are often wired with negative emotions resulting from a threat that can be contagious [78]. Moreover, fear is also associated with panic behaviors, where humans act egoistically, endangering others' lives as well. These behaviors were captured via the emotion and topic analysis of this study, demonstrating that social media reflected people's underlying emotions and mental health concerns.

Studies investigating social media sentiments and topics during the pandemic align with our study [20,79,80]. In addition, our study provides an extension to the current body of COVID-19 social media research by modeling the emotion state transitions over time to represent the shift of emotions over time. The profiling of citizens based on emotions and concerns also enables the identification of clusters in the community during different stages of the pandemic. Recent research related to identifying the role of social media in the COVID-19 pandemic suggested further exploration of how social media can be utilized to inform health care practices [23]. We believe that our study contributes to providing outcomes that can enhance health care practices related to community understanding and uplifting of mental health care during a similar crisis.

Conclusions

As the entire world is still grappling with the direct or indirect effects of the COVID-19 pandemic, it is pertinent that we explore and analyze human emotions and related expressions in order to aid postpandemic recovery, as well as to prepare for future crises. Given the current social isolation settings, people are experiencing increased mental health issues and emotion changes that are affecting their quality of life [81,82]. During times of social isolation, the abundance of data available in digital spaces provides a reflection of people's behaviors, which is otherwise challenging to assess via clinical trials or interviews. This enables emotion analysis using digital spaces to act as an adjunct to real-world behavior analysis [83].

This study focused on capturing human emotions and concerns, which were reflected through social media conversations as the COVID-19 pandemic progressed. Our work is targeted at detecting and helping to understand the underlying mental health– and well-being–related issues via emotions expressed and topics discussed on social media by different segments of society. Although prevention of spread with social distancing and masking, as well as attempts at the elimination of COVID-19 with vaccine programs, have been widely discussed, less attention has been paid to the less visible issue of emotional distress and mental health. Governments and health care organizations are now realizing the deep and long-term effects of these issues; funding is being allocated to address these issues and programs for understanding the implications are now being established [81]. The work described here is focused on this less discussed area. We proposed an AI-based approach and designed and developed a supporting technical framework to learn about the emotion and content representations from unstructured, voluminous social media data, which are otherwise challenging to assess manually. The outcomes from this study aligned with theories of social sharing, emotional responses during pandemics, and collective emotion theories, and they represent insights that can be used to improve the mental health of citizens.

The outcomes of the emotion analysis demonstrated a high intensity of fear and sadness emotions during the announcement stage of the COVID-19 pandemic, and the emotion transition model also demonstrated higher probabilities of expressing fear and sadness during the prepandemic stage among Australian citizens. Noteworthy insights were produced as the outcomes of the temporal emotion analysis. The emotion transition models across the pandemic demonstrated how the emotions had changed over time. It was noted that, although fear and sadness emotions were more prominent over the first few months, they had eventually transitioned into anger and disgust as people expressed their dissatisfaction and frustration with the continuing pandemic and restrictions. The temporal emotion analysis during the pandemic using emotion transition models was one of the main contributions of this study, as it demonstrates how people's emotions change over time. On the other hand, positive emotions such as trust and anticipation were observed in the first lockdown and easing restrictions stages; however, they had diminished over time as the pandemic progressed. This *emergence and shift* of different emotions is a strong indicator of people's varying emotions, and it signals the underlying mental health of the population. The decreased intensity of sadness and fear in the latter stage showed the normalization of emotions as people tended to realize the nature of the pandemic. However, the increased emotion transitions toward anger and disgust as well as broader concerns represented the dissatisfaction toward the management of the pandemic. These emotion behaviors presented insights into people's emotions, which were challenging to quantify and measure using

traditional surveys or clinical trials, given the social isolation settings. Social media has been effective in helping us understand human emotion expressions during times of crisis, and it can be utilized to improve health care practices in similar crises.

On the other hand, emotions expressed on social media can be contagious, as negative content attracts more attention and propagation on social media [84,85]. The increased intensity of negative emotions and the persistence of negativity over time that was captured across different stages in this study aids this phenomenon. Negative emotions were observed in both first and second lockdowns and were significantly higher during the second lockdown. Therefore, it is crucial to observe this negative content to avoid further escalation. These swirls of emotions generated at one point in time have the ability to propagate through time using social media as the medium, as social media holds the potential for cascading and going viral to transmit emotional content more rapidly and broadly than other media [86,87]. These behaviors can also be aligned with the social sharing of emotion, which states that collective emotional events are anticipated to trigger a social process of emotion propagation [12]. It has been stated that emotion strength, presence, and valence can affect and influence citizen engagement through social media; therefore, it is imperative to moderate and monitor such emotional behaviors on digital platforms [88]. As potential implications, the proposed emotion behavior analysis in the study can be used to moderate online negativity and even detect distress among people during the pandemic to uplift their mental health as well as in postpandemic times as measures of precautions.

Another key observation of this study was the profiling of people based on similar emotions and thematic expressions on social media. It was observed that while certain groups of people were more enthused about family matters, some groups expressed more concerns related to employment issues. The deviation of profiles from the first lockdown to the second lockdown also shows how the identified profiles took different shapes depending on the concerns at that point in time. Such behavior profiling enables the understanding of different groups of people with different concerns and emotions. Such groups elicit significant insights and enable us to look at the community from different perspectives. Based on these behavior changes, governance and health care policy-making practices could be tailored to consider different groups of concerns emerging from different sectors within the community. Insights can also inform future planning to uplift citizens' mental health during disastrous events similar to the current pandemic.

Apart from the aforementioned implications, one of the key outcomes of this study was to showcase the applicability and capability of using social media conversations to identify, analyze, and understand human behaviors at scale during a global crisis. Digital avatars created in social media have become a derivative of reality. This study showcases that in the age of social media, it is helpful, or even essential, to study human behaviors and emotions using their digital representations.

Conflicts of Interest

None declared.

References

1. Johns Hopkins Coronavirus Resource Center. URL: https://coronavirus.jhu.edu/ [accessed 2020-12-16]
2. Bavel JJV, Baicker K, Boggio PS, Capraro V, Cichocka A, Cikara M, et al. Using social and behavioural science to support COVID-19 pandemic response. Nat Hum Behav 2020 May;4(5):460-471. [doi: 10.1038/s41562-020-0884-z] [Medline: 32355299]
3. Gao J, Zheng P, Jia Y, Chen H, Mao Y, Chen S, et al. Mental health problems and social media exposure during COVID-19 outbreak. PLoS One 2020;15(4):e0231924 [FREE Full text] [doi: 10.1371/journal.pone.0231924] [Medline: 32298385]
4. Valdez D, Ten Thij M, Bathina K, Rutter L, Bollen J. Social media insights into US mental health during the COVID-19 pandemic: Longitudinal analysis of Twitter data. J Med Internet Res 2020 Dec 14;22(12):e21418 [FREE Full text] [doi: 10.2196/21418] [Medline: 33284783]
5. Goel A, Gupta L. Social media in the times of COVID-19. J Clin Rheumatol 2020 Jun 11;26(6):220-223. [doi: 10.1097/rhu.0000000000001508]
6. Alexander DE. Social media in disaster risk reduction and crisis management. Sci Eng Ethics 2014 Sep;20(3):717-733. [doi: 10.1007/s11948-013-9502-z] [Medline: 24306994]
7. Gao H, Barbier G, Goolsby R. Harnessing the crowdsourcing power of social media for disaster relief. IEEE Intell Syst 2011 May;26(3):10-14. [doi: 10.1109/mis.2011.52]
8. Palen L, Hughes A. Social media in disaster communication. In: Rodríguez H, Donner W, Trainor J, editors. Handbook of Disaster Research. Cham, Switzerland: Springer; 2018:497-518.
9. Abd-Alrazaq A, Alhuwail D, Househ M, Hamdi M, Shah Z. Top concerns of tweeters during the COVID-19 pandemic: Infoveillance study. J Med Internet Res 2020 Apr 21;22(4):e19016 [FREE Full text] [doi: 10.2196/19016] [Medline: 32287039]
10. Panayiotou G, Panteli M, Leonidou C. Coping with the invisible enemy: The role of emotion regulation and awareness in quality of life during the COVID-19 pandemic. J Contextual Behav Sci 2021 Jan;19:17-27. [doi: 10.1016/j.jcbs.2020.11.002]

11. Restubog SLD, Ocampo ACG, Wang L. Taking control amidst the chaos: Emotion regulation during the COVID-19 pandemic. J Vocat Behav 2020 Jun;119:103440 [FREE Full text] [doi: 10.1016/j.jvb.2020.103440] [Medline: 32390659]
12. Rimé B. Emotion elicits the social sharing of emotion: Theory and empirical review. Emot Rev 2009;1(1):60-85. [doi: 10.1177/1754073908097189]
13. Jin Y, Liu BF, Anagondahalli D, Austin L. Scale development for measuring publics' emotions in organizational crises. Public Relat Rev 2014 Sep;40(3):509-518. [doi: 10.1016/j.pubrev.2014.04.007]
14. Rimé B, Páez D, Basabe N, Martínez F. Social sharing of emotion, post-traumatic growth, and emotional climate: Follow-up of Spanish citizen's response to the collective trauma of March 11th terrorist attacks in Madrid. Eur J Soc Psychol 2009 Sep 09;40(6):1029-1045. [doi: 10.1002/ejsp.700]
15. Jin Y, Fraustino JD, Liu BF. The scared, the outraged, and the anxious: How crisis emotions, involvement, and demographics predict publics' conative coping. Int J Strateg Commun 2016 Apr 19;10(4):289-308. [doi: 10.1080/1553118x.2016.1160401]
16. Adikari A, Gamage G, de Silva D, Mills N, Wong SJ, Alahakoon D. A self structuring artificial intelligence framework for deep emotions modeling and analysis on the social web. Future Gener Comput Syst 2021 Mar;116:302-315. [doi: 10.1016/j.future.2020.10.028]
17. Adikari A, Alahakoon D. Understanding citizens' emotional pulse in a smart city using artificial intelligence. IEEE Trans Industr Inform 2021 Apr;17(4):2743-2751. [doi: 10.1109/tii.2020.3009277]
18. Adikari A, de Silva D, Ranasinghe WKB, Bandaragoda T, Alahakoon O, Persad R, et al. Can online support groups address psychological morbidity of cancer patients? An artificial intelligence based investigation of prostate cancer trajectories. PLoS One 2020;15(3):e0229361 [FREE Full text] [doi: 10.1371/journal.pone.0229361] [Medline: 32130256]
19. Xue J, Chen J, Hu R, Chen C, Zheng C, Su Y, et al. Twitter discussions and emotions about the COVID-19 pandemic: Machine learning approach. J Med Internet Res 2020 Nov 25;22(11):e20550 [FREE Full text] [doi: 10.2196/20550] [Medline: 33119535]
20. Chandrasekaran R, Mehta V, Valkunde T, Moustakas E. Topics, trends, and sentiments of tweets about the COVID-19 pandemic: Temporal infoveillance study. J Med Internet Res 2020 Oct 23;22(10):e22624 [FREE Full text] [doi: 10.2196/22624] [Medline: 33006937]
21. Hung M, Lauren E, Hon ES, Birmingham WC, Xu J, Su S, et al. Social network analysis of COVID-19 sentiments: Application of artificial intelligence. J Med Internet Res 2020 Aug 18;22(8):e22590 [FREE Full text] [doi: 10.2196/22590] [Medline: 32750001]
22. Al-Rawi A, Siddiqi M, Morgan R, Vandan N, Smith J, Wenham C. COVID-19 and the gendered use of emojis on Twitter: Infodemiology study. J Med Internet Res 2020 Nov 05;22(11):e21646 [FREE Full text] [doi: 10.2196/21646] [Medline: 33052871]
23. Al-Dmour H, Masa'deh R, Salman A, Abuhashesh M, Al-Dmour R. Influence of social media platforms on public health protection against the COVID-19 pandemic via the mediating effects of public health awareness and behavioral changes: Integrated model. J Med Internet Res 2020 Aug 19;22(8):e19996 [FREE Full text] [doi: 10.2196/19996] [Medline: 32750004]
24. Choudhury M, De S. Mental health discourse on reddit: Self-disclosure, social support, and anonymity. In: Proceedings of the 8th International AAAI Conference on Weblogs and Social Media. 2014 Presented at: 8th International AAAI Conference on Weblogs and Social Media; June 1-4, 2014; Ann Arbor, MI p. 71-80 URL: https://ojs.aaai.org/index.php/ICWSM/article/view/14526/14375
25. De Choudhury M, Counts S, Horvitz E. Predicting postpartum changes in emotion and behavior via social media. In: Proceedings of the 31st SIGCHI Conference on Human Factors in Computing Systems. New York, NY: Association for Computing Machinery; 2013 Presented at: 31st SIGCHI Conference on Human Factors in Computing Systems; April 27-May 2, 2013; Paris, France p. 3267-3276. [doi: 10.1145/2470654.2466447]
26. De CM, Counts S, Horvitz E. Social media as a measurement tool of depression in populations. In: Proceedings of the 5th Annual ACM Web Science Conference. New York, NY: Association for Computing Machinery; 2013 Presented at: 5th Annual ACM Web Science Conference; May 2-4, 2013; Paris, France p. 47-56. [doi: 10.1145/2464464.2464480]
27. Adikari A, Gamage G, de Silva D, Mills N, Wong SJ, Alahakoon D. A self structuring artificial intelligence framework for deep emotions modeling and analysis on the social web. Future Gener Comput Syst 2021 Mar;116:302-315. [doi: 10.1016/j.future.2020.10.028]
28. Cha M, Haddadi H, Benevenuto F, Gummadi K. Measuring user influence in Twitter: The million follower fallacy. In: Proceedings of the 4th International AAAI Conference on Weblogs and Social Media. 2010 Presented at: 4th International AAAI Conference on Weblogs and Social Media; May 23-26, 2010; Washington, DC p. 10-17 URL: https://aaai.org/ocs/index.php/ICWSM/ICWSM10/paper/view/1538/1826
29. Yosufzai R. Australia among top 10 countries ranked for pandemic response, NZ tops list. SBS News. 2021 Jan 28. URL: https://www.sbs.com.au/news/australia-among-top-10-countries-ranked-for-pandemic-response-nz-tops-list [accessed 2020-12-16]
30. Hong J, Chang R, Varley K. The Covid resilience ranking: These are the best and worst places to be during the Covid pandemic. Bloomberg. 2020. URL: https://www.bloomberg.com/graphics/covid-resilience-ranking/ [accessed 2020-12-16]

31. Fisher J, Tran T, Hammarberg K, Sastry J, Nguyen H, Rowe H, et al. Mental health of people in Australia in the first month of COVID-19 restrictions: A national survey. Med J Aust 2020 Nov;213(10):458-464. [doi: 10.5694/mja2.50831] [Medline: 33107063]
32. Pachana N, Beattie E, Byrne G, Brodaty H. COVID-19 and psychogeriatrics: The view from Australia. Int Psychogeriatr 2020 May 12;32(10):1135-1141. [doi: 10.1017/s1041610220000885]
33. Rice ES, Haynes E, Royce P, Thompson SC. Social media and digital technology use among Indigenous young people in Australia: A literature review. Int J Equity Health 2016 May 25;15:81 [FREE Full text] [doi: 10.1186/s12939-016-0366-0] [Medline: 27225519]
34. Balakrishnan V, Teoh KK, Pourshafie T, Liew TK. Social media and their use in learning: A comparative analysis between Australia and Malaysia from the learners' perspectives. Australas J Educ Technol 2016 Nov 11;33:81-97. [doi: 10.14742/ajet.2469]
35. Ferguson C, Inglis SC, Newton PJ, Cripps PJ, MacDonald PS, Davidson PM. Collegian 2014;21(2):89-93. [doi: 10.1016/j.colegn.2014.03.002] [Medline: 25109206]
36. Anikeeva O, Steenkamp M, Arbon P. The future of social media use during emergencies in Australia: Insights from the 2014 Australian and New Zealand Disaster and Emergency Management conference social media workshop. Aust J Emerg Manag 2015 Jan;30(1):22-26 [FREE Full text]
37. Flew T, Bruns A, Burgess J, Crawford K, Shaw F. Social media and its impact on crisis communication: Case studies of Twitter use in emergency management in Australia and New Zealand. In: Proceedings of the Communication and Social Transformation, ICA Regional Conference. 2013 Presented at: Communication and Social Transformation, ICA Regional Conference; November 8-10, 2013; Shanghai, China URL: https://eprints.qut.edu.au/63707/15/ICA_Shanghai_paper_Social_Media_and_Crisis_Comms.pdf
38. De Silva D, Ranasinghe W, Bandaragoda T, Adikari A, Mills N, Iddamalgoda L, et al. Machine learning to support social media empowered patients in cancer care and cancer treatment decisions. PLoS One 2018;13(10):e0205855 [FREE Full text] [doi: 10.1371/journal.pone.0205855] [Medline: 30335805]
39. Esuli A, Sebastiani F. SENTIWORDNET: A publicly available lexical resource for opinion mining. In: Proceedings of the 5th International Conference on Language Resources and Evaluation.: European Language Resources Association (ELRA); 2006 Presented at: 5th International Conference on Language Resources and Evaluation; May 22-28, 2006; Genoa, Italy p. 417-422 URL: http://www.lrec-conf.org/proceedings/lrec2006/pdf/384_pdf.pdf
40. Baccianella S, Esuli A, Sebastiani F. SENTIWORDNET 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In: Proceedings of the 7th International Conference on Language Resources and Evaluation. 2010 Presented at: 7th International Conference on Language Resources and Evaluation; May 17-23, 2010; Valletta, Malta p. 2200-2204 URL: http://www.lrec-conf.org/proceedings/lrec2010/pdf/769_Paper.pdf
41. Strapparava C, Valitutti A. WordNet-Affect: An affective extension of WordNet. In: Proceedings of the 4th International Conference on Language Resources and Evaluation. 2010 Presented at: 4th International Conference on Language Resources and Evaluation; May 26-28, 2004; Lisbon, Portugal p. 1083-1086 URL: http://www.lrec-conf.org/proceedings/lrec2004/pdf/369.pdf
42. Franzoni V, Biondi G, Milani A. A web-based system for emotion vector extraction. In: Proceedings of the 17th International Conference on Computational Science and its Applications. Cham, Switzerland: Springer International Publishing; 2017 Presented at: 17th International Conference on Computational Science and its Applications; July 3-6, 2017; Trieste, Italy p. 653-668. [doi: 10.1007/978-3-319-62392-4]
43. Araque O, Corcuera-Platas I, Sánchez-Rada JF, Iglesias CA. Enhancing deep learning sentiment analysis with ensemble techniques in social applications. Expert Syst Appl 2017 Jul;77:236-246. [doi: 10.1016/j.eswa.2017.02.002]
44. Basiri ME, Nemati S, Abdar M, Cambria E, Acharya UR. ABCDM: An Attention-based Bidirectional CNN-RNN Deep Model for sentiment analysis. Future Gener Comput Syst 2021 Feb;115:279-294. [doi: 10.1016/j.future.2020.08.005]
45. Cambria E, Schuller B, Xia Y, Havasi C. New avenues in opinion mining and sentiment analysis. IEEE Intell Syst 2013 Mar;28(2):15-21. [doi: 10.1109/MIS.2013.30]
46. Giatsoglou M, Vozalis MG, Diamantaras K, Vakali A, Sarigiannidis G, Chatzisavvas KC. Sentiment analysis leveraging emotions and word embeddings. Expert Syst Appl 2017 Mar;69:214-224. [doi: 10.1016/j.eswa.2016.10.043]
47. Franzoni V, Vallverdù J, Milani A. Errors, biases and overconfidence in artificial emotional modeling. In: Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence. New York, NY: Association for Computing Machinery; 2019 Presented at: IEEE/WIC/ACM International Conference on Web Intelligence; October 14-17, 2019; Thessaloniki, Greece p. 86-90. [doi: 10.1145/3358695.3361749]
48. Ekman P. An argument for basic emotions. Cogn Emot 2008 Jan 07;6(3-4):169-200. [doi: 10.1080/02699939208411068]
49. Mohammad SM. Sentiment analysis: Detecting valence, emotions, and other affectual states from text. In: Meiselman HL, editor. Emotion Measurement. Cambridge, UK: Woodhead Publishing; 2016:201-237.
50. Plutchik R. A psychoevolutionary theory of emotions. Soc Sci Inf 2017 Feb 06;21(4-5):529-553. [doi: 10.1177/053901882021004003]
51. Binali H, Wu C, Potdar V. Computational approaches for emotion detection in text. In: Proceedings of the 4th IEEE International Conference on Digital Ecosystems and Technologies. 2010 Presented at: 4th IEEE International Conference

on Digital Ecosystems and Technologies; April 12-15, 2010; Dubai, United Arab Emirates p. 172-177. [doi: 10.1109/dest.2010.5610650]
52. Mashal S, Asnani K. Emotion intensity detection for social media data. In: Proceedings of the International Conference on Computing Methodologies and Communication. 2017 Presented at: International Conference on Computing Methodologies and Communication; July 18-19, 2017; Erode, India p. 155-158. [doi: 10.1109/iccmc.2017.8282664]
53. Rao Y, Li Q, Wenyin L, Wu Q, Quan X. Affective topic model for social emotion detection. Neural Netw 2014 Oct;58:29-37. [doi: 10.1016/j.neunet.2014.05.007] [Medline: 24913903]
54. Herzig J, Shmueli-Scheuer M, Konopnicki D. Emotion detection from text via ensemble classification using word embeddings. In: Proceedings of the ACM SIGIR International Conference on Theory of Information Retrieval. New York, NY: Association for Computing Machinery; 2017 Presented at: ACM SIGIR International Conference on Theory of Information Retrieval; October 1-4, 2017; Amsterdam, the Netherlands p. 269-272. [doi: 10.1145/3121050.3121093]
55. Zhang L, Wang S, Liu B. Deep learning for sentiment analysis: A survey. WIREs Data Min Knowl Discov 2018 Mar 30;8(4):1. [doi: 10.1002/widm.1253]
56. Felbo B, Mislove A, Søgaard A, Rahwan I, Lehmann S. Using millions of emoji occurrences to learn any-domain representations for detecting sentiment, emotion and sarcasm. In: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. 2017 Presented at: 2017 Conference on Empirical Methods in Natural Language Processing; September 7-11, 2017; Copenhagen, Denmark p. 1616-1626 URL: https://www.aclweb.org/anthology/D17-1169.pdf [doi: 10.18653/v1/d17-1169]
57. Majumder N, Poria S, Hazarika D, Mihalcea R, Gelbukh A, Cambria E. DialogueRNN: An attentive RNN for emotion detection in conversations. In: Proceedings of the 33rd AAAI Conference on Artificial Intelligence. 2019 Jul Presented at: 33rd AAAI Conference on Artificial Intelligence; January 27-February 1, 2019; Honolulu, HI p. 6818-6825. [doi: 10.1609/aaai.v33i01.33016818]
58. Rathnayaka P, Abeysinghe S, Samarajeewa C, Manchanayake I, Walpola M, Nawaratne R, et al. Gated recurrent neural network approach for multilabel emotion detection in microblogs. ArXiv. Preprint posted online on July 17, 2019 [FREE Full text]
59. Wood ID, Ruder S. Emoji as emotion tags for tweets. In: Proceedings of the 10th International Conference on Language Resources and Evaluation, Emotion and Sentiment Analysis Workshop. 2016 Presented at: 10th International Conference on Language Resources and Evaluation, Emotion and Sentiment Analysis Workshop; May 23-28, 2016; Portorož, Slovenia p. 76-79 URL: http://www.lrec-conf.org/proceedings/lrec2016/workshops/LREC2016Workshop-ESA_Proceedings.pdf
60. Akhtar MS, Ekbal A, Cambria E. How intense are you? Predicting intensities of emotions and sentiments using stacked ensemble [application notes]. IEEE Comput Intell Mag 2020 Feb;15(1):64-75. [doi: 10.1109/mci.2019.2954667]
61. Cambria E. Affective computing and sentiment analysis. IEEE Intell Syst 2016 Mar;31(2):102-107. [doi: 10.1109/mis.2016.31]
62. O'Sullivan D, Rahamathulla M, Pawar M. The impact and implications of COVID-19: An Australian perspective. Int J Community Soc Dev 2020 Jul 01;2(2):134-151. [doi: 10.1177/2516602620937922]
63. Mikolov T, Chen K, Corrado G, Dean J. Efficient estimation of word representations in vector space. In: Proceedings of the 1st International Conference on Learning Representations. 2013 Presented at: 1st International Conference on Learning Representations; May 2-4, 2013; Scottsdale, AZ URL: https://arxiv.org/pdf/1301.3781
64. Picard RW. Affective Computing. Cambridge, MA: The MIT Press; 1997.
65. Han J, Xie L, Liu X, Wang Z. Emotional state transfer model based on FSM. In: Proceedings of the International Conference on Multisensor Fusion and Information Integration for Intelligent Systems. 2014 Presented at: International Conference on Multisensor Fusion and Information Integration for Intelligent Systems; September 28-30, 2014; Beijing, China p. 1-5. [doi: 10.1109/mfi.2014.6997732]
66. Meng Q, Wu W. Artificial emotional model based on finite state machine. J Cent South Univ Technol 2008 Oct 1;15(5):694-699. [doi: 10.1007/s11771-008-0129-1]
67. Blei DM, Ng AY, Jordan MI. Latent Dirichlet allocation. J Mach Learn Res 2003;3:993-1022 [FREE Full text]
68. Rose S, Engel D, Cramer N, Cowley W. Automatic keyword extraction from individual documents. In: Berry MW, Kogan J, editors. Text Mining: Applications and Theory. Chichester, UK: John Wiley & Sons; 2010:3-20.
69. Alahakoon D, Halgamuge SK, Srinivasan B. Dynamic self-organizing maps with controlled growth for knowledge discovery. IEEE Trans Neural Netw 2000;11(3):601-614. [doi: 10.1109/72.846732] [Medline: 18249788]
70. Nawaratne R, Alahakoon D, De Silva D, Kumara H, Yu X. Hierarchical two-stream growing self-organizing maps with transience for human activity recognition. IEEE Trans Industr Inform 2020 Dec;16(12):7756-7764. [doi: 10.1109/tii.2019.2957454]
71. Nawaratne R, Alahakoon D, De SD, Yu X. HT-GSOM: Dynamic self-organizing map with transience for human activity recognition. In: Proceedings of the IEEE 17th International Conference on Industrial Informatics. 2019 Presented at: IEEE 17th International Conference on Industrial Informatics; July 22-25, 2019; Helsinki-Espoo, Finland p. 270. [doi: 10.1109/indin41052.2019.8972260]

72. Rafea A, Gaballah N. Topic extraction in social media. In: Proceedings of the 2013 International Conference on Collaboration Technologies and Systems. 2013 Presented at: 2013 International Conference on Collaboration Technologies and Systems; May 20-24, 2013; San Diego, CA. [doi: 10.1109/cts.2013.6567212]
73. Podlogar MC, Rogers ML, Stanley IH, Hom MA, Chiurliza B, Joiner TE. Anxiety, depression, and the suicidal spectrum: A latent class analysis of overlapping and distinctive features. Cogn Emot 2018 Nov;32(7):1464-1477. [doi: 10.1080/02699931.2017.1303452] [Medline: 28317414]
74. Campbell I. Chi-squared and Fisher-Irwin tests of two-by-two tables with small sample recommendations. Stat Med 2007 Aug 30;26(19):3661-3675. [doi: 10.1002/sim.2832] [Medline: 17315184]
75. Richardson JTE. The analysis of 2 × 2 contingency tables--Yet again. Stat Med 2011 Apr 15;30(8):890; author reply 891-890; author reply 892. [doi: 10.1002/sim.4116] [Medline: 21432882]
76. Altman D, Machin D, Bryant T, Gardner M, editors. Statistics with Confidence: Confidence Intervals and Statistical Guidelines. 2nd edition. London, UK: BMJ Books; 2000.
77. Li S, Wang Y, Xue J, Zhao N, Zhu T. The impact of COVID-19 epidemic declaration on psychological consequences: A study on active Weibo users. Int J Environ Res Public Health 2020 Mar 19;17(6):2032 [FREE Full text] [doi: 10.3390/ijerph17062032] [Medline: 32204411]
78. Pakpour AH, Griffiths MD. The fear of COVID-19 and its role in preventive behaviors. J Concurr Disord 2020;2(1):58-63 [FREE Full text]
79. Boon-Itt S, Skunkan Y. Public perception of the COVID-19 pandemic on Twitter: Sentiment analysis and topic modeling study. JMIR Public Health Surveill 2020 Nov 11;6(4):e21978 [FREE Full text] [doi: 10.2196/21978] [Medline: 33108310]
80. Rajput N, Grover B, Rathi V. Word frequency and sentiment analysis of Twitter messages during Coronavirus pandemic. ArXiv. Preprint posted online on April 8, 2020 [FREE Full text]
81. Pfefferbaum B, North CS. Mental health and the Covid-19 pandemic. N Engl J Med 2020 Aug 06;383(6):510-512. [doi: 10.1056/NEJMp2008017] [Medline: 32283003]
82. Rajkumar RP. COVID-19 and mental health: A review of the existing literature. Asian J Psychiatr 2020 Aug;52:102066 [FREE Full text] [doi: 10.1016/j.ajp.2020.102066] [Medline: 32302935]
83. Meshi D, Tamir DI, Heekeren HR. The emerging neuroscience of social media. Trends Cogn Sci 2015 Dec;19(12):771-782. [doi: 10.1016/j.tics.2015.09.004] [Medline: 26578288]
84. Peng W, Adikari A, Alahakoon D, Gero J. Discovering the influence of sarcasm in social media responses. WIREs Data Min Knowl Discov 2019 Jul 29;9(6):1. [doi: 10.1002/widm.1331]
85. Naveed N, Gottron T, Kunegis J, Alhadi A. Bad news travel fast: A content-based analysis of interestingness on Twitter. In: Proceedings of the 3rd International Web Science Conference. New York, NY: Association for Computing Machinery; 2011 Presented at: 3rd International Web Science Conference; June 15-17, 2011; Koblenz, Germany. [doi: 10.1145/2527031.2527052]
86. Berger J, Milkman KL. Social Transmission, Emotion, and the Virality of Online Content. Marketing Science Institute Working Paper Series 2010 Report No. 10-114. New York, NY: Marketing Science Institute; 2010. URL: https://www.msi.org/wp-content/uploads/2020/06/MSI_Report_10-114.pdf [accessed 2021-04-22]
87. Hansen L, Arvidsson A, Nielsen F, Colleoni E, Etter M. Good friends, bad news - Affect and virality in Twitter. In: Proceedings of the 6th International Conference on Future Information Technology. 2011 Presented at: 6th International Conference on Future Information Technology; June 28-30, 2011; Crete, Greece p. 34-43. [doi: 10.1007/978-3-642-22309-9_5]
88. Chen Q, Min C, Zhang W, Wang G, Ma X, Evans R. Unpacking the black box: How to promote citizen engagement through government social media during the COVID-19 crisis. Comput Human Behav 2020 Sep;110:106380 [FREE Full text] [doi: 10.1016/j.chb.2020.106380] [Medline: 32292239]

Abbreviations

AI: artificial intelligence
GSOM: growing self-organizing map
NLU: natural language understanding

Edited by C Basch; submitted 21.01.21; peer-reviewed by E Cambria, V Franzoni; comments to author 15.02.21; revised version received 03.03.21; accepted 01.04.21; published 30.04.21.

Please cite as:
Adikari A, Nawaratne R, De Silva D, Ranasinghe S, Alahakoon O, Alahakoon D
Emotions of COVID-19: Content Analysis of Self-Reported Information Using Artificial Intelligence
J Med Internet Res 2021;23(4):e27341
URL: https://www.jmir.org/2021/4/e27341
doi:10.2196/27341
PMID:33819167

Original Paper

Using Speech Data From Interactions With a Voice Assistant to Predict the Risk of Future Accidents for Older Drivers: Prospective Cohort Study

Yasunori Yamada[1], PhD; Kaoru Shinkawa[1], ME; Masatomo Kobayashi[1], PhD; Hironobu Takagi[1], PhD; Miyuki Nemoto[2], PhD; Kiyotaka Nemoto[3], MD, PhD; Tetsuaki Arai[3], MD, PhD

[1]IBM Research, Tokyo, Japan

[2]Department of Psychiatry, University of Tsukuba Hospital, Ibaraki, Japan

[3]Department of Psychiatry, Faculty of Medicine, University of Tsukuba, Ibaraki, Japan

Corresponding Author:
Yasunori Yamada, PhD
IBM Research
Nihonbashi, Hakozaki-cho, Chuo-ku
Tokyo, 103-8510
Japan
Phone: 81 80 6706 9381
Email: ysnr@jp.ibm.com

Abstract

Background: With the rapid growth of the older adult population worldwide, car accidents involving this population group have become an increasingly serious problem. Cognitive impairment, which is assessed using neuropsychological tests, has been reported as a risk factor for being involved in car accidents; however, it remains unclear whether this risk can be predicted using daily behavior data.

Objective: The objective of this study was to investigate whether speech data that can be collected in everyday life can be used to predict the risk of an older driver being involved in a car accident.

Methods: At baseline, we collected (1) speech data during interactions with a voice assistant and (2) cognitive assessment data—neuropsychological tests (Mini-Mental State Examination, revised Wechsler immediate and delayed logical memory, Frontal Assessment Battery, trail making test-parts A and B, and Clock Drawing Test), Geriatric Depression Scale, magnetic resonance imaging, and demographics (age, sex, education)—from older adults. Approximately one-and-a-half years later, we followed up to collect information about their driving experiences (with respect to car accidents) using a questionnaire. We investigated the association between speech data and future accident risk using statistical analysis and machine learning models.

Results: We found that older drivers (n=60) with accident or near-accident experiences had statistically discernible differences in speech features that suggest cognitive impairment such as reduced speech rate (P=.048) and increased response time (P=.040). Moreover, the model that used speech features could predict future accident or near-accident experiences with 81.7% accuracy, which was 6.7% higher than that using cognitive assessment data, and could achieve up to 88.3% accuracy when the model used both types of data.

Conclusions: Our study provides the first empirical results that suggest analysis of speech data recorded during interactions with voice assistants could help predict future accident risk for older drivers by capturing subtle impairments in cognitive function.

(J Med Internet Res 2021;23(4):e27667) doi:10.2196/27667

KEYWORDS

cognitive impairment; smart speaker; speech analysis; accident; prevention; older adults; prediction; risk; assistant

Introduction

As the world's older adult population increases, car accidents involving older adults have become an increasingly serious social problem. While it has been reported that older drivers have an increased risk of car accident involvement per unit distance travelled [1-4], they also showed a substantially higher rate of serious injury than that of middle-age car drivers [5,6].

Even in normal aging, there is a decline in many cognitive abilities related to driving, and this cognitive decline is known to be one of the risk factors for older adults being involved in car accidents [7-9]. Associating cognitive assessment scores with either self-reported car accidents, crash records, or on-road driving measures has been investigated to identify predictors of driving safety (in previous empirical studies [7]). In particular, cognitive abilities such as visual attention, short-term memory, and executive functions (evaluated with neuropsychological tests) were consistently shown to have associations with driving safety [7,10-12]. In this respect, if cognitive impairments relevant to driving safety in older adults can be inferred solely from behavior data in everyday situations in a passive way, this would be beneficial for accident prevention.

Speech in daily life can be used a potential data sources for determining cognitive impairments related to driving safety. Speech involves multiple interacting cognitive abilities including attention, memory, and executive functions [13,14]. Many empirical studies have used speech data to identify cognitive impairments resulting from aging and diseases such as Alzheimer disease [15-18] and characterized speech changes related to cognitive impairments by extracting linguistic and paralinguistic features from speech data [19-25]. For example, difficulties with word finding and word retrieving have been quantified by tallying pronoun frequency and pause durations [19,20,24,26-28]. A reduction in speech expressiveness has also been quantified by measuring lexical diversity and speech rate [19,23,29-31]. Using a combination of these features, previous studies [19-25,29] have succeeded in differentiating individuals with cognitive impairments from healthy controls. Although no study has investigated the relationship between speech data and driving safety, it is reasonable to explore the possibility that speech data could be used for inferring ability to drive safely from changes in cognitive functioning in older drivers.

At the same time, there is growing interest in using speech data that can be collected in everyday situations for applications in health care owing to the popularity of voice-based interaction systems such as voice assistants in smart speakers and smartphones [32-34]. One approach is to provide various types of voice-based tests via a smart-speaker platform. For example, previous studies [35,36] have used mobile apps for collecting speech responses to neuropsychological tasks such as verbal fluency and picture description tasks; they showed accurate classification rates in detecting patients with Alzheimer disease [35] and dementia [36]. Another approach is to analyze health-related insights from speech data collected during daily voice-based interactions. For example, vocal characteristics in speech data during typical tasks on smart speakers appeared to be associated with neuropsychological test scores [37], while linguistic features extracted from phone conversation data were significant indicators for differentiating patients with Alzheimer disease from older adults with normal cognition [38]. This approach, focusing on speech data that can be collected in everyday situations, would increase opportunities for frequent assessment by facilitating passive and unobtrusive monitoring.

In this study, we aimed to investigate the relationship between speech data and future driving experiences related to car accidents in healthy older adults by collecting speech data during interactions with a voice assistant with simulated tasks on smart speakers and smartphones. We hypothesized that these speech data could be used for predicting accident risk for older drivers.

Methods

Participants

We recruited healthy older adults aged 60 years or older through recruiting agencies and advertisements in the local community in Ibaraki, Japan. All examinations were conducted in Japanese. Older adults met the inclusion criteria if they were in good physical and mental health and had no serious diseases, disabilities, mental illness (eg, major depression, bipolar disorder, and schizophrenia), or neurodegenerative diseases (eg, Parkinson disease and dementia). This study was conducted with the approval of the University of Tsukuba Hospital Ethics Committee (H29-065). All participants provided written consent after the procedures of the study had been fully explained.

A total of 71 older individuals participated in the cognitive assessments and speech data collection (women: 38/71, 53.5%; age: range 61-80 years, mean 71.1, SD 4.9). Of the original 71 participants, 60 consented to the follow-up study about their driving experiences (women: 33/60, 55.0%; age: range 61-80 years, mean 70.8, SD 5.1; Table 1). They were contacted again approximately one-and-a-half years after the speech data collection (mean 17.3 months, SD 2.7) and answered a questionnaire on their driving experiences within the past year. The questionnaire included free-form questions about accidents and near accidents; *near accidents* were described as infractions and any other incidents while driving that they deemed to be dangerous regardless of severity and culpability.

Table 1. Demographic and assessment data for study participants.

Variable	Total (N=60)	Individuals without accident or near-accident experiences (n=34)	Individuals with accident or near-accident experiences (n=26)	*P* value
Age (years), mean (SD)	70.8 (5.1)	70.5 (4.9)	71.3 (5.3)	.45
Education (years), mean (SD)	13.7 (2.2)	13.7 (2.2)	13.6 (2.1)	.93
Sex, n (%)				.53
Men	27 (45)	17 (50)	10 (38)	
Women	33 (55)	17 (50)	16 (62)	
Mini-Mental State Examination[a], mean (SD)	27.6 (1.8)	27.4 (1.8)	27.9 (1.8)	.28
LM IA[b], mean (SD)	9.6 (3.8)	9.1 (3.7)	10.2 (4.0)	.43
LM IIA[c], mean (SD)	7.5 (3.6)	7.3 (3.7)	7.6 (3.6)	.74
Frontal Assessment Battery[d], mean (SD)	13.7 (2.7)	13.4 (2.7)	14.2 (2.7)	.45
Trail making test-part A (seconds), mean (SD)	33.2 (9.8)	33.6 (9.7)	32.6 (10.1)	.72
Trail making test-part B (seconds), mean (SD)	89.5 (49.7)	95.7 (60.9)	81.3 (28.2)	.71
Clock Drawing Test[e], mean (SD)	6.7 (0.8)	6.7 (0.7)	6.7 (1.0)	.36
Geriatric Depression Scale[f], mean (SD)	2.9 (2.4)	2.8 (2.4)	3.1 (2.4)	.62
Severity scores for atrophy in medial temporal structures, mean (SD)	0.9 (0.6)	0.8 (0.4)	0.9 (0.7)	.86

[a]The total possible score ranges from 0 to 30.

[b]LM IA: immediate recall of the logical memory-story A of the Wechsler memory scale-revised for episodic memory; the total possible score ranges from 0 to 25.

[c]LM IIA: delayed recall of the logical memory-story A of the Wechsler memory scale-revised for episodic memory; the total possible score ranges from 0 to 25.

[d]The total possible score ranges from 0 to 18.

[e]The total possible score ranges from 0 to 7.

[f]The total possible score ranges from 0 to 15.

Cognitive Assessments

Cognitive assessments and examinations were those typically used for the diagnosis of dementia and comprised 12 variables: age, sex, education, 7 neuropsychological test scores (Mini-Mental State Examination for global cognition; immediate and delayed recall of the logical memory-story A of the Wechsler memory scale-revised for episodic memory; the Frontal Assessment Battery for executive function; the trail making test-part A and B for executive function and attention; and the clock drawing test for visuospatial function), and 2 clinical scores (Geriatric Depression Scale and the severity of medial temporal lobe atrophy). The severity of medial temporal lobe atrophy was evaluated using structural magnetic resonance imaging (MRI) scans—1.5 T, T1-weighted images and a 3D gradient-echo sequence—with the following parameters: sagittal orientation with 1.2-mm thick sections; time repetition/time echo: 2400/3.52 milliseconds; flip angle: 8°; field of view: 192×192. We expressed the severity of medial temporal lobe atrophy as a *Z* score relative to cognitively healthy adults by using a standalone, voxel-based specific regional analysis system for Alzheimer disease [39]. Two psychiatrists (KN and TA) reviewed the results of the cognitive assessments and confirmed that participants did not meet the criteria for dementia based on those of the National Institute on Aging and Alzheimer's Association and Alzheimer disease Neuroimaging Initiative 2 [40].

Speech Data Collection

We simulated conversations with a voice assistant on modern smart speakers and smartphones and collected the speech data while performing 3 typical task scenarios: information retrieval (asking for tomorrow's weather), shopping online (booking a movie ticket), and personal schedule management (creating a calendar event). The tasks began with a simple scenario and then advanced to the more complicated ones. Each task started with an initiating question from the system ("what can I help you with?"), with follow-up questions that asked for detailed information related to the task. The follow-up questions were presented in a fixed order. The questions consisted of four categories—open-ended, to which participants responded with a free-form sentence (Multimedia Appendix 1: Table S1); multiple choice, to which participants responded by choosing one of the options stated in the question; prepared input, to which participants responded with information (eg, passcode) specified by the experimenter; and confirmation, to which participants responded by accepting or rejecting a statement made by the system. The system presented at least 22 questions in total to each participant for the 3 tasks.

To simulate conversations, we took a Wizard-of-Oz [41] approach, in which the participants were told that they were talking with a computer system, though in fact the interaction was mediated by an experimenter (ie, the wizard). We chose this approach so that we could avoid uncertain factors such as errors in automatic speech recognition. During the tasks, the experimenter made the system present a question. After the participant responded, the experimenter prompted the system to move onto the next question if the response contained the necessary information corresponding to the question; otherwise, they would repeat the same question. Each open-ended, multiple choice, and prepared input question presented by the system was scripted in advance and the same for all participants. For confirmation questions, we prepared several variations for each question and the experimenter chose one, depending on the participant's previous response. For example, the experimenter chose "you are purchasing one ticket, is it OK?" or "you are purchasing two tickets, is it OK?" to have the participant confirm the number of tickets to book.

The interface for speech data collection was implemented as a tablet-based app on an Apple iPad Air 2. In the experiment, participants sat down in front of the tablet and talked with the system (Figure 1a). During the tasks, the tablet showed a screen indicating whether it was speaking (Figure 1b) or listening (Figure 1c). The experimenter sat behind the participant and operated the system by using a separate interface hidden from the participants. Speech data were recorded in raw format with a sampling rate of 44.1 kHz through the embedded microphone in the tablet. Each experimental session took approximately 30 minutes per participant, including instructions and wrap-up. Additional details about our apparatus and procedure have been previously published [37].

Figure 1. Overview of experimental setup: (a) setup for collecting speech data, (b) screen showing participant's turn, and (c) screen showing the tablet's turn.

Data Analysis

From each participant's speech data, we automatically extracted 84 paralinguistic speech features used in previous studies on inferring cognitive impairments and detecting early signs of Alzheimer disease [19,20,23,27-29,31,42,43]. They consisted of 56 acoustic features and 28 prosodic features.

The acoustic features consisted of features related to mel-frequency cepstral coefficients (MFCCs), jitter, and shimmer. We used the mean and first-order derivatives of the first 12 MFCCs, which represent the short-term power spectrum of the speech signal. Jitter and shimmer features measure cycle-to-cycle variations of fundamental frequency and amplitude [44]. Prosodic features included speech rate, pitch variability, phonation time, number of phonemes needed for completing tasks, response time, total pause duration, and proportion of long pauses (pauses >0.8 seconds). Both acoustic and prosodic features were extracted from each task's speech data separately. We used Python (version 3.8) audio-processing libraries (librosa, version 0.8.0 [45]; Signal_Analysis, version 0.1.26 [46]).

Statistical analyses were performed using Statistics and Machine Learning Toolbox (version 11.1) for MATLAB (version R2017a, The MathWorks Inc) environment. To assess the differences in each variable between participants with and without accident or near-accident experiences, we used 2-sided Mann-Whitney tests for continuous data and chi-square tests for categorical data. We did not correct for multiple comparisons, and *P* values <.05 were considered significant.

The prediction models for differentiating individuals with and without accident or near-accident experiences were built using multiple types of binary classifiers with automatic sequential forward selection of features. Model performance was evaluated with both leave-one-subject-out cross validation and 100 iterations of 10-fold cross-validation methods. The classifiers included *k*-nearest neighbors [47], random forest [48] and support vector machine [49]. The parameters that we studied were as follows: the number of neighbors for the *k*-nearest neighbors; the number and the maximum depth of trees for random forest; kernel functions, penalty parameter, and the parameter associated with the width of the radial basis function kernel for the support vector machine. We performed an exhaustive grid search to determine these parameters. The algorithms were implemented using the Python scikit-learn package (version 0.23.2).

Results

For speech data collection (the 30-minute sessions), we obtained an average of 23.8 responses within 100.2 seconds (SD 28.6) from each participant. The average response duration of each task scenario ranged from 17.4 to 59.6 seconds (mean 33.41, SD 9.5). The average duration of a single response for each participant ranged from 1.1 to 7.6 seconds (mean 4.2, SD 1.1). At follow-up, 26 of the 60 participants (43.3%) reported car accident or near-accident experiences within the previous year.

Of those, 23 participants reported a near-accident experience, 2 reported accidents, and 1 reported both. The near-accidents consisted of near-misses with a car or pedestrian resulting in a sense of fear and anxiety (eg, from failure to notice a crossing pedestrian), errors in operation (eg, stepping on the accelerator instead of the brake), and unintentional violations (eg, entering the opposite lane).

In comparisons between individuals with and without accident or near-accident experiences, there were no significant differences in any cognitive assessment variables (age: P=.45; education year: P=.93; sex: P=.53; Mini-Mental State Examination: P=.28; immediate and delayed recall of the logical memory-story A of the Wechsler memory scale-revised: P=.43, P=.74; the Frontal Assessment Battery: P=.45; the trail making test-part A and B: P=.72, P=.71; the clock drawing test: P=.36; Geriatric Depression Scale: P=.62; severity scores for atrophy in medial temporal structures: P=.86; Table 1); however, we found 10 speech features with significant differences—$\Delta MFCC_1$: P=.005, $\Delta MFCC_4$: P=.043, $\Delta MFCC_5$: P=.011, $\Delta MFCC_7$: P=.035, $\Delta MFCC_{12}$: P=.023; jitter: P=.034; response time: P=.040; proportion of long pauses: P=.044; speech rate: P=.048; and number of phonemes needed for completing tasks: P=.049 (Figure 2; Multimedia Appendix 1: Table S2). Those with accident or near-accident experiences showed decreased speech rate and jitter as well as increased response time and long pauses. These speech features were reported in previous studies as significant indicators of changes in cognitive function, and the trends in their changes were consistent with those observed in individuals with cognitive impairments and patients with Alzheimer disease and mild cognitive impairment (for speech rate [23,27,31]; for jitter [42,43]; for response time [20,27]; for proportion of long pause [27,28]).

To visualize whether the variance seen among a variable set is capable of discriminating between individuals with and without potential future accident or near-accident experiences, we performed principal component analysis on 2 variable sets: the 12 cognitive assessment variables and 10 speech features (Figure 3). The cognitive assessment variable set had little capability to differentiate the groups; there was considerable overlap and no clear separation. In contrast, the speech variable set enabled some separation of the groups.

Input variables for the classification models were either or both the 12 cognitive assessment variables and 10 speech features. When model performance was evaluated with leave-one-subject-out cross-validation, with only the cognitive assessment variables, we obtained 75.0% accuracy (65.4% sensitivity, 82.4% specificity, and 69.4% F1 score; Figure 4a), with only the speech features, the model accuracy increased to 81.7% accuracy (65.4% sensitivity, 94.1% specificity, and 75.6% F1 score; Figure 4b), and with speech features and cognitive assessment variables combined, performance improved further (88.3% accuracy, 88.5% sensitivity, 88.2% specificity, and 86.8% F1 score; Figure 4c). When we evaluated the model using 10-fold cross validation, the results showed similar trends (Multimedia Appendix 1: Table S3): the model using the cognitive assessment variables achieved 75.5% accuracy (95% CI 75.1-75.9), the model using speech features achieved 80.1% accuracy (95% CI 79.7-80.5), and the model using both types of features achieved 85.5% accuracy (95% CI 85.1-85.9).

Figure 2. Box plots (line and diamond represent median and mean, respectively) for speech features with significant differences between individuals with and without accident or near-accident experiences—jitter: P=.034; response time: P=.040; speech rate: P=.048.

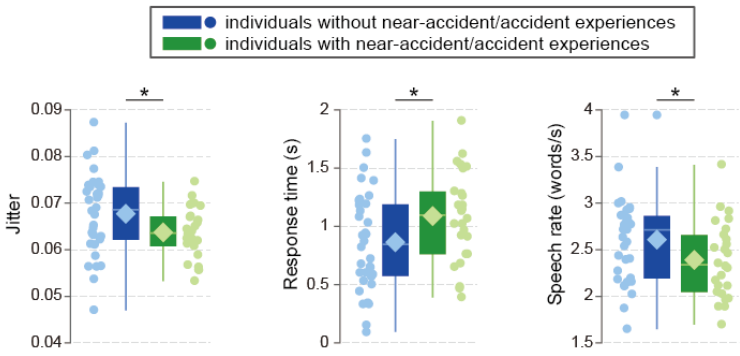


Figure 3. Principal component analysis plots using (a) cognitive assessment variables and (b) speech features, with confidence interval ellipsoid set to 0.95. PC: principal component.

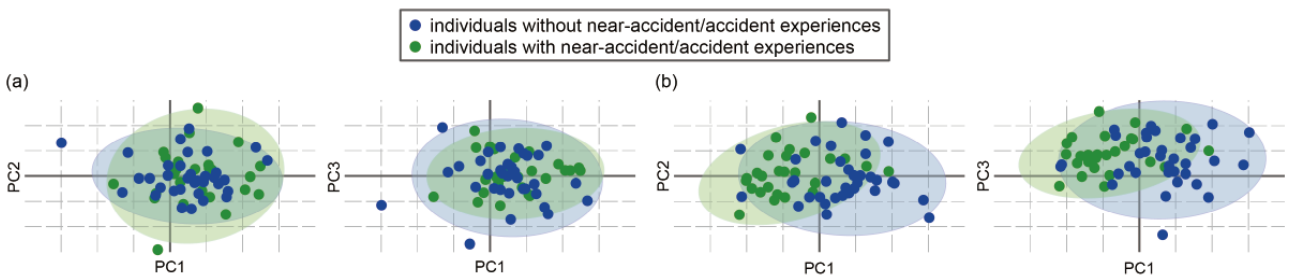
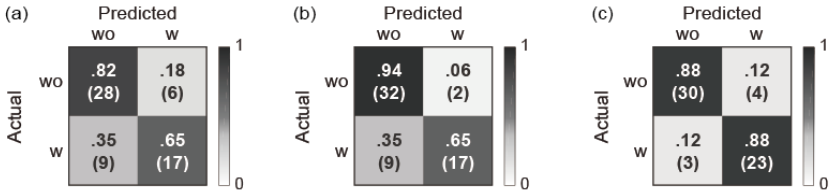


Figure 4. Confusion matrixes for predicting future accident risks of older drivers obtained using leave-one-subject-out cross-validation for models with (a) cognitive assessment variables, (b) speech features, and (c) cognitive assessment variables and speech features combined. The number in parentheses indicates the number of participants. wo: without; w: with.



Discussion

Principal Results

In light of the increasing demand for preventing car accidents involving older adults, we investigated the possibility that future accident risk related to cognitive impairments could be automatically predicted with passive unobtrusive monitoring. To this end, we focused on speech data because many previous studies have succeeded in quantifying and detecting cognitive impairments from speech data [19,20,23,27-29,31,42,43], speech data are becoming more accessible, and voice-based interaction systems such as voice assistants are becoming more popular [32-34].

The statistical analysis showed that the speech data collected during typical tasks on smart speakers and smartphones had statistically discernible speech features between older drivers with and without accident or near-accident experiences. These speech features indicated that older drivers with these experiences tended to show decreased speech rate and jitter as well as increased response time and long pauses. These changes in speech features were reported as statistically significant signatures for cognitive impairments by previous studies on patients with Alzheimer disease and mild cognitive impairment [19,20,23,27-29,31,42,43]. The results suggest that speech features could capture subtle impairments of cognitive function in older drivers. On the other hand, we found no differences in any cognitive assessment variables, but this could be explained by the criteria for driving risks that differed from those in previous studies [7,10,12,50-53]. While previous studies compared older drivers with and without car-accident experiences regardless of having near-car-accident experiences and reported significant differences in cognitive assessment scores between them [7,10,12,50-53], we focused on both accident and near-accident experiences, and the majority of the high-risk group in our study were individuals with near-accident experiences but without actual car accidents. Speech data and cognitive assessment results suggest that eliciting discernible changes relevant to future near-accident experiences may require cognitive assessment for subtle impairments, such as, test batteries used for screening preclinical Alzheimer disease [54,55]. Even so, if speech data during interactions with voice assistants can be used for predicting future accident risk, it would greatly increase the accessibility of early screening with a relatively low burden.

The classification model using speech features achieved 81.7% accuracy, which is 6.7% higher than that using cognitive assessment data, and models achieved up to 88.3% accuracy with both combined. Dimensional reduction and visualization using principal component analysis, an unsupervised method, showed that the feature space with speech data was better able to separate those with and without accident or near-accident experiences than the feature space with cognitive assessment variables. These results and those of the statistical analysis indicate that speech data during typical tasks with voice assistants could have comparable (or possibly more) information for predicting future accident risks of older drivers compared with the standard cognitive assessments.

Our results show paralinguistic speech characteristics were useful for predicting future accident risks of older drivers. Previous user-interface studies reported that voice input was effective and was preferable as an input modality for older adults [56-58], while other studies reported that the performance of automatic speech recognition tended to be worse in older adults than in other age groups [59,60]. From this perspective, our results suggest that models for predicting future accident risks of older drivers can be made robust against errors of automatic speech recognition by exploiting paralinguistic features.

Our results highlight the possibility that cognitive impairments related to future car accident risks could be detected using speech data collected in everyday life. Assistive and automated driving systems are promising technologies that may help older adults with cognitive challenges to safely continue driving [61]. Recent studies suggested the importance of individual differences in cognitive abilities for assistive and automated driving technologies for older adults [62,63] because literature has suggested that cognitive abilities affect both performance with automated technology and perceptions of automation (ie, trust) [64,65]. Hence, our approach to detect cognitive impairments associated with driving risks might provide useful information for the personalization of assistive and automated driving systems based on the cognitive abilities of older adults.

Limitations

Our work had several limitations. First, we collected speech data in a lab setting. The controlled setting might affect the way people interact with a voice assistant. In future work, data collection in free-living situations using voice assistants would be needed along with additional interaction scenarios. Second, the sample size was limited. In spite of this limitation, our statistical analysis of speech features showed consistent trends indicating subtle cognitive impairments in older adults with future accident or near-accident experiences, and the prediction performance (to predict independent future accidents) using speech features was as high as 88%, even when the classifier was trained on a subsample. From these perspectives, we believe

that our results can be confirmed by future studies. Third, our definition of future car accident risks was based on self-reports of accident and near-accident experiences. In future work, we need to consider obtaining more objective measures for accident risks by combining self-reports with on-road driving assessments, informant reports, or drive recorder videos.

Conclusion

Given the increasing demand for car accident prevention involving older adults, we explored the possibility of predicting future accident risks associated with cognitive impairments by using behavioral data that can be collected in everyday life. To this end, we focused on speech data collected during interactions with voice assistants in smart speakers and smartphones and investigated the associations with future accident risks by following up with older drivers. We found that (1) older drivers with accident or near-accident experiences had statistically discernible changes in speech features, implying cognitive impairments, and (2) the machine learning model using speech features could predict future accident or near-accident experiences with up to 88.3% accuracy. Although further studies with speech data collected in everyday life and objective data for near-accidents are needed, our study provides the first empirical results suggesting that speech data during interactions with voice assistants in smart speakers and smartphones could help predict future accident risks of older drivers by capturing subtle impairments in cognitive function. We believe that our results can be used in future efforts toward preventing driving accidents of older adults through continuous passive unobtrusive monitoring.

Acknowledgments

This work was supported by Japan Society for the Promotion of Science KAKENHI (grants 19H01084 and 18K18164).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary tables.
[PDF File (Adobe PDF File), 194 KB - jmir_v23i4e27667_app1.pdf]

References

1. Cicchino JB, McCartt AT. Trends in older driver crash involvement rates and survivability in the United States: an update. Accid Anal Prev 2014 Nov;72:44-54. [doi: 10.1016/j.aap.2014.06.011] [Medline: 25003969]
2. Cicchino JB, McCartt AT. Critical older driver errors in a national sample of serious U.S. crashes. Accid Anal Prev 2015 Jul;80:211-219. [doi: 10.1016/j.aap.2015.04.015] [Medline: 25916662]
3. Ma L, Yan X. Examining the nonparametric effect of drivers' age in rear-end accidents through an additive logistic regression model. Accid Anal Prev 2014 Jun;67:129-136. [doi: 10.1016/j.aap.2014.02.021] [Medline: 24642249]
4. Lombardi DA, Horrey WJ, Courtney TK. Age-related differences in fatal intersection crashes in the United States. Accid Anal Prev 2017 Feb;99(Pt A):20-29 [FREE Full text] [doi: 10.1016/j.aap.2016.10.030] [Medline: 27855312]
5. Ryan GA, Legge M, Rosman D. Age related changes in drivers' crash risk and crash type. Accid Anal Prev 1998 May;30(3):379-387. [doi: 10.1016/s0001-4575(97)00098-5] [Medline: 9663297]
6. Meuleners LB, Harding A, Lee AH, Legge M. Fragility and crash over-representation among older drivers in Western Australia. Accid Anal Prev 2006 Sep;38(5):1006-1010. [doi: 10.1016/j.aap.2006.04.005] [Medline: 16713982]
7. Anstey KJ, Wood J, Lord S, Walker JG. Cognitive, sensory and physical factors enabling driving safety in older adults. Clin Psychol Rev 2005 Jan;25(1):45-65. [doi: 10.1016/j.cpr.2004.07.008] [Medline: 15596080]
8. Bryan J, Luszcz MA. Measurement of executive function: considerations for detecting adult age differences. J Clin Exp Neuropsychol 2000 Feb;22(1):40-55. [doi: 10.1076/1380-3395(200002)22:1;1-8;FT040] [Medline: 10649544]
9. McGwin, Jr G, Brown DB. Characteristics of traffic crashes among young, middle-aged, and older drivers. Accid Anal Prev 1999 May;31(3):181-198. [doi: 10.1016/s0001-4575(98)00061-x] [Medline: 10196595]
10. Odenheimer GL, Beaudet M, Jette AM, Albert MS, Grande L, Minaker KL. Performance-based driving evaluation of the elderly driver: safety, reliability, and validity. J Gerontol 1994 Jul;49(4):M153-M519. [doi: 10.1093/geronj/49.4.m153] [Medline: 8014389]
11. Sims RV, McGwin G, Allman RM, Ball K, Owsley C. Exploratory study of incident vehicle crashes among older drivers. J Gerontol A Biol Sci Med Sci 2000 Jan;55(1):M22-M27. [doi: 10.1093/gerona/55.1.m22] [Medline: 10719769]
12. Daigneault G, Joly P, Frigon J. Executive functions in the evaluation of accident risk of older drivers. J Clin Exp Neuropsychol 2002 Apr;24(2):221-238. [doi: 10.1076/jcen.24.2.221.993] [Medline: 11992205]
13. Gibson MT, Mangat D, Gagliano G, Wilson M, Fletcher J, Bulman J, et al. Evaluation of the efficacy of a redox agent in the treatment of chronic periodontitis. J Clin Periodontol 1994 Nov;21(10):690-700. [doi: 10.1111/j.1600-051x.1994.tb00788.x] [Medline: 7852614]

14. Kirshner HS. Primary progressive aphasia and Alzheimer's disease: brief history, recent evidence. Curr Neurol Neurosci Rep 2012 Dec;12(6):709-714. [doi: 10.1007/s11910-012-0307-2] [Medline: 22932755]
15. Taler V, Phillips NA. Language performance in Alzheimer's disease and mild cognitive impairment: a comparative review. J Clin Exp Neuropsychol 2008 Jul;30(5):501-556. [doi: 10.1080/13803390701550128] [Medline: 18569251]
16. Garrard P, Lambon Ralph MA, Watson PC, Powis J, Patterson K, Hodges JR. Longitudinal profiles of semantic impairment for living and nonliving concepts in dementia of Alzheimer's type. J Cogn Neurosci 2001 Oct 01;13(7):892-909. [doi: 10.1162/089892901753165818] [Medline: 11595093]
17. Oulhaj A, Wilcock GK, Smith AD, de Jager CA. Predicting the time of conversion to MCI in the elderly: role of verbal expression and learning. Neurology 2009 Nov 03;73(18):1436-1442 [FREE Full text] [doi: 10.1212/WNL.0b013e3181c0665f] [Medline: 19794124]
18. Forbes-McKay KE, Venneri A. Detecting subtle spontaneous language decline in early Alzheimer's disease with a picture description task. Neurol Sci 2005 Oct;26(4):243-254. [doi: 10.1007/s10072-005-0467-9] [Medline: 16193251]
19. Fraser KC, Meltzer JA, Rudzicz F. Linguistic features identify Alzheimer's disease in narrative speech. J Alzheimers Dis 2016;49(2):407-422. [doi: 10.3233/JAD-150520] [Medline: 26484921]
20. König A, Satt A, Sorin A, Hoory R, Toledo-Ronen O, Derreumaux A, et al. Automatic speech analysis for the assessment of patients with predementia and Alzheimer's disease. Alzheimers Dement (Amst) 2015 Mar;1(1):112-124 [FREE Full text] [doi: 10.1016/j.dadm.2014.11.012] [Medline: 27239498]
21. Roark B, Mitchell M, Hosom J, Hollingshead K, Kaye J. Spoken language derived measures for detecting mild cognitive impairment. IEEE Trans Audio Speech Lang Process 2011 Sep 01;19(7):2081-2090 [FREE Full text] [doi: 10.1109/TASL.2011.2112351] [Medline: 22199464]
22. Sajjadi SA, Patterson K, Tomek M, Nestor PJ. Abnormalities of connected speech in semantic dementia vs Alzheimer's disease. Aphasiology 2012 Jun;26(6):847-866. [doi: 10.1080/02687038.2012.654933]
23. Themistocleous C, Eckerström M, Kokkinakis D. Voice quality and speech fluency distinguish individuals with mild cognitive impairment from healthy controls. PLoS One 2020;15(7):e0236009 [FREE Full text] [doi: 10.1371/journal.pone.0236009] [Medline: 32658934]
24. Jarrold W, Peintner B, Wilkins D, Vergryi D, Richey C, Gorno-Tempini M. Aided diagnosis of dementia type through computer-based analysis of spontaneous speech. Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality; 2014 Presented at: Association for Computational Linguistics; June 27; Baltimore, Maryland, USA p. 27-37. [doi: 10.3115/v1/w14-3204]
25. Satt A, Hoory R, König A, Aalten P, Robert P. Speech-based automatic and robust detection of very early dementia. 2014 Presented at: Fifteenth Annual Conference of the International Speech Communication Association; Sep 14-18; Singapore.
26. Ahmed S, Haigh AF, de Jager CA, Garrard P. Connected speech as a marker of disease progression in autopsy-proven Alzheimer's disease. Brain 2013 Dec;136(Pt 12):3727-3737 [FREE Full text] [doi: 10.1093/brain/awt269] [Medline: 24142144]
27. Khodabakhsh A, Yesil F, Guner E, Demiroglu C. Evaluation of linguistic and prosodic features for detection of Alzheimer's disease in Turkish conversational speech. J Audio Speech Music Proc 2015 Mar 25;2015(1):1-15. [doi: 10.1186/s13636-015-0052-y]
28. Pastoriza-Dominguez P, Torre I, Dieguez-Vide F, Gomez-Ruiz I, Gelado S, Bello-Lopez J. Speech pause distribution as an early marker for Alzheimer's disease. medRxiv. Preprint posted online on January 4, 2021. [doi: 10.1101/2020.12.28.20248875]
29. Hernández-Domínguez L, Ratté S, Sierra-Martínez G, Roche-Bergua A. Computer-based evaluation of Alzheimer's disease and mild cognitive impairment patients during a picture description task. Alzheimers Dement (Amst) 2018;10:260-268 [FREE Full text] [doi: 10.1016/j.dadm.2018.02.004] [Medline: 29780871]
30. Aramaki E, Shikata S, Miyabe M, Kinoshita A. Vocabulary size in speech may be an early indicator of cognitive impairment. PLoS One 2016;11(5):e0155195 [FREE Full text] [doi: 10.1371/journal.pone.0155195] [Medline: 27176919]
31. Hoffmann I, Nemeth D, Dye CD, Pákáski M, Irinyi T, Kálmán J. Temporal parameters of spontaneous speech in Alzheimer's disease. Int J Speech Lang Pathol 2010 Feb;12(1):29-34. [doi: 10.3109/17549500903137256] [Medline: 20380247]
32. Kourtis LC, Regele OB, Wright JM, Jones GB. Digital biomarkers for Alzheimer's disease: the mobile/wearable devices opportunity. NPJ Digit Med 2019;2:9 [FREE Full text] [doi: 10.1038/s41746-019-0084-2] [Medline: 31119198]
33. Manfredi C, Lebacq J, Cantarella G, Schoentgen J, Orlandi S, Bandini A, et al. Smartphones offer new opportunities in clinical voice research. J Voice 2017 Jan;31(1):111.e1-111.e7. [doi: 10.1016/j.jvoice.2015.12.020] [Medline: 27068549]
34. Sezgin E, Huang Y, Ramtekkar U, Lin S. Readiness for voice assistants to support healthcare delivery during a health crisis and pandemic. NPJ Digit Med 2020;3:122 [FREE Full text] [doi: 10.1038/s41746-020-00332-0] [Medline: 33015374]
35. Konig A, Satt A, Sorin A, Hoory R, Derreumaux A, David R, et al. Use of speech analyses within a mobile application for the assessment of cognitive impairment in elderly people. Curr Alzheimer Res 2018;15(2):120-129. [doi: 10.2174/1567205014666170829111942] [Medline: 28847279]
36. Hall AO, Shinkawa K, Kosugi A, Takase T, Kobayashi M, Nishimura M, et al. Using tablet-based assessment to characterize speech for individuals with dementia and mild cognitive impairment: preliminary results. AMIA Jt Summits Transl Sci Proc 2019;2019:34-43 [FREE Full text] [Medline: 31258954]

37. Kobayashi M, Kosugi A, Takagi H, Nemoto M, Nemoto K, Arai T. Effects of age-related cognitive decline on elderly user interactions with voice-based dialogue systems. 2019 Presented at: IFIP Conference on Human Computer Interaction (Springer); September 2-6; Cyprus p. 53-74. [doi: 10.1007/978-3-030-29390-1_4]
38. Yamada Y, Shinkawa K, Shimmei K. Atypical repetition in daily conversation on different days for detecting Alzheimer disease: evaluation of phone-call data from regular monitoring service. JMIR Ment Health 2020 Jan 12;7(1):e16790 [FREE Full text] [doi: 10.2196/16790] [Medline: 31934870]
39. Matsuda H, Mizumura S, Nemoto K, Yamashita F, Imabayashi E, Sato N, et al. Automatic voxel-based morphometry of structural MRI by SPM8 plus diffeomorphic anatomic registration through exponentiated lie algebra improves the diagnosis of probable Alzheimer Disease. AJNR Am J Neuroradiol 2012 Jun;33(6):1109-1114 [FREE Full text] [doi: 10.3174/ajnr.A2935] [Medline: 22300935]
40. Procedures manual. Alzheimer's Disease Neuroimaging Initiative 2 (ADNI 2). 2008. URL: https://adni.loni.usc.edu/wp-content/uploads/2008/07/adni2-procedures-manual.pdf [accessed 2021-02-02]
41. Dahlbäck N, Jönsson A, Ahrenberg L. Wizard of Oz studies — why and how. Knowl Based Syst 1993 Dec;6(4):258-266. [doi: 10.1016/0950-7051(93)90017-n]
42. Alhanai T, Au R, Glass J. Spoken language biomarkers for detecting cognitive impairment. 2017 Presented at: Proc IEEE Workshop Autom Speech Recognit Underst; December 16-20; Okinawa, Japan p. 409-416. [doi: 10.1109/asru.2017.8268965]
43. Farrús M, Codina-Filbà J. Combining prosodic, voice quality and lexical features to automatically detect Alzheimer's disease. arXiv. Preprint posted online November 18, 2020 [FREE Full text]
44. Farrús M, Hernando J, Ejarque P. Jitter and shimmer measurements for speaker recognition. 2007 Presented at: Eighth annual conference of the international speech communication association; August 27-31; Antwerp, Belgium.
45. McFee B, Raffel C, Liang D, Ellis D, McVicar M, Battenberg E. librosa: Audio and music signal analysis in python. 2015 Presented at: Proceedings of the 14th python in science conference; July 6-12; Austin, Texas. [doi: 10.25080/majora-7b98e3ed-003]
46. Signal–Analysis. GitHub. URL: https://brookemosby.github.io/Signal_Analysis/ [accessed 2021-02-26]
47. Goldberger J, Hinton G, Roweis S, Salakhutdinov R. Neighbourhood components analysis. Adv Neural Inf Process Syst 2004;17:513-520.
48. Breiman L. Random forests. In: Boehmke B, Greenwell B, editors. Hands-On Machine Learning With R. Boca Raton: Chapman and Hall/CRC; 2001:5-32.
49. Boser B, Guyon I, Vapnik V. A training algorithm for optimal margin classifiers. In: Proceedings of the Fifth Annual Workshop on Computational Learning Theory. 1992 Presented at: Fifth Annual Workshop on Computational Learning Theory; July 27-29; Pittsburgh, Pennsylvania p. 144-152. [doi: 10.1145/130385.130401]
50. Stutts JC, Stewart JR, Martell C. Cognitive test performance and crash risk in an older driver population. Accid Anal Prev 1998 May;30(3):337-346. [doi: 10.1016/s0001-4575(97)00108-5] [Medline: 9663293]
51. McKnight AJ, McKnight AS. Multivariate analysis of age-related driver ability and performance deficits. Accid Anal Prev 1999 Sep;31(5):445-454. [doi: 10.1016/s0001-4575(98)00082-7] [Medline: 10440541]
52. Owsley C, Ball K, Sloane ME, Roenker DL, et al. Visual/cognitive correlates of vehicle accidents in older drivers. Psychol Aging 1991 Sep;6(3):403-415. [doi: 10.1037//0882-7974.6.3.403] [Medline: 1930757]
53. Payyanadan R, Gibson M, Chiou E, Ghazizadeh M, Lee J. Contextual design for driving: developing a trip-planning tool for older adults. Transp Res Part F Traffic Psychol Behav 2017 Apr;46:462-476. [doi: 10.1016/j.trf.2016.08.005]
54. Donohue MC, Sperling RA, Salmon DP, Rentz DM, Raman R, Thomas RG, et al. The preclinical Alzheimer cognitive composite: measuring amyloid-related decline. JAMA Neurol 2014 Aug;71(8):961-970 [FREE Full text] [doi: 10.1001/jamaneurol.2014.803] [Medline: 24886908]
55. Duke Han S, Nguyen CP, Stricker NH, Nation DA. Detectable neuropsychological differences in early preclinical Alzheimer's disease: a meta-analysis. Neuropsychol Rev 2017 Dec;27(4):305-325 [FREE Full text] [doi: 10.1007/s11065-017-9345-5] [Medline: 28497179]
56. Smith AL, Chaparro BS. Smartphone text input method performance, usability, and preference with younger and older adults. Hum Factors 2015 Sep;57(6):1015-1028. [doi: 10.1177/0018720815575644] [Medline: 25850116]
57. Liu Y, Chen C, Lin Y, Chen H, Irianti D, Jen T, et al. Design and usability evaluation of mobile voice-added food reporting for elderly people: randomized controlled trial. JMIR Mhealth Uhealth 2020 Sep 28;8(9):e20317 [FREE Full text] [doi: 10.2196/20317] [Medline: 32985999]
58. Stigall B, Waycott J, Baker S, Caine K. Older adults' perception and use of voice user interfaces: A preliminary review of the computing literature. In: Proceedings of the 31st Australian Conference on Human-Computer-Interaction.: Association for Computing Machinery; 2019 Presented at: 31st Australian Conference on Human-Computer-Interaction; December 3-5; Fremantle Australia p. 423-427. [doi: 10.1145/3369457.3369506]
59. Werner L, Huang G, Pitts B. Automated speech recognition systems and older adults: a literature review and synthesis. 2019 Presented at: Proc Hum Factors Ergon Soc Annu Meet; October 28-November 1; Seattle, Washington, USA p. 42-46.
60. Portet F, Vacher M, Golanski C, Roux C, Meillon B. Design and evaluation of a smart home voice interface for the elderly: acceptability and objection aspects. Pers Ubiquit Comput 2011 Oct 2;17(1):127-144. [doi: 10.1007/s00779-011-0470-5]

61. Knoefel F, Wallace B, Goubran R, Sabra I, Marshall S. Semi-autonomous vehicles as a cognitive assistive device for older adults. Geriatrics (Basel) 2019 Nov 16;4(4):63 [FREE Full text] [doi: 10.3390/geriatrics4040063] [Medline: 31744041]
62. Rovira E, McLaughlin AC, Pak R, High L. Looking for age differences in self-driving vehicles: examining the effects of automation reliability, driving risk, and physical impairment on trust. Front Psychol 2019;10:800 [FREE Full text] [doi: 10.3389/fpsyg.2019.00800] [Medline: 31105610]
63. Depestele S, Ross V, Verstraelen S, Brijs K, Brijs T, van Dun K, et al. The impact of cognitive functioning on driving performance of older persons in comparison to younger age groups: a systematic review. Transp Res Part F Traffic Psychol Behav 2020 Aug;73:433-452. [doi: 10.1016/j.trf.2020.07.009]
64. Rovira E, Pak R, McLaughlin A. Effects of individual differences in working memory on performance and trust with various degrees of automation. Theor Issues Ergon Sci 2016 Nov 21;18(6):573-591. [doi: 10.1080/1463922x.2016.1252806]
65. Pak R, McLaughlin AC, Leidheiser W, Rovira E. The effect of individual differences in working memory in older adults on performance with different degrees of automated technology. Ergonomics 2017 Apr;60(4):518-532. [doi: 10.1080/00140139.2016.1189599] [Medline: 27409279]

Abbreviations

MFCC: mel-frequency cepstral coefficients
MRI: magnetic resonance imaging

Edited by G Eysenbach; submitted 02.02.21; peer-reviewed by A Hall, V Stara; comments to author 25.02.21; revised version received 08.03.21; accepted 15.03.21; published 08.04.21.

Please cite as:
Yamada Y, Shinkawa K, Kobayashi M, Takagi H, Nemoto M, Nemoto K, Arai T
Using Speech Data From Interactions With a Voice Assistant to Predict the Risk of Future Accidents for Older Drivers: Prospective Cohort Study
J Med Internet Res 2021;23(4):e27667
URL: https://www.jmir.org/2021/4/e27667
doi:10.2196/27667
PMID:33830066

Original Paper

Voice-Controlled Intelligent Personal Assistants in Health Care: International Delphi Study

Alena Ermolina[1], MSc; Victor Tiberius[1], Prof Dr rer pol Dr phil, MBA

Faculty of Economics and Social Sciences, University of Potsdam, Potsdam, Germany

Corresponding Author:
Victor Tiberius, Prof Dr rer pol Dr phil, MBA
Faculty of Economics and Social Sciences
University of Potsdam
August-Bebel-Str 89
Potsdam, 14882
Germany
Phone: 49 331 977 ext 3593
Fax: 49 331 977 3404
Email: tiberius@uni-potsdam.de

Abstract

Background: Voice-controlled intelligent personal assistants (VIPAs), such as Amazon Echo and Google Home, involve artificial intelligence–powered algorithms designed to simulate humans. Their hands-free interface and growing capabilities have a wide range of applications in health care, covering off-clinic education, health monitoring, and communication. However, conflicting factors, such as patient safety and privacy concerns, make it difficult to foresee the further development of VIPAs in health care.

Objective: This study aimed to develop a plausible scenario for the further development of VIPAs in health care to support decision making regarding the procurement of VIPAs in health care organizations.

Methods: We conducted a two-stage Delphi study with an internationally recruited panel consisting of voice assistant experts, medical professionals, and representatives of academia, governmental health authorities, and nonprofit health associations having expertise with voice technology. Twenty projections were formulated and evaluated by the panelists. Descriptive statistics were used to derive the desired scenario.

Results: The panelists expect VIPAs to be able to provide solid medical advice based on patients' personal health information and to have human-like conversations. However, in the short term, voice assistants might neither provide frustration-free user experience nor outperform or replace humans in health care. With a high level of consensus, the experts agreed with the potential of VIPAs to support elderly people and be widely used as anamnesis, informational, self-therapy, and communication tools by patients and health care professionals. Although users' and governments' privacy concerns are not expected to decrease in the near future, the panelists believe that strict regulations capable of preventing VIPAs from providing medical help services will not be imposed.

Conclusions: According to the surveyed experts, VIPAs will show notable technological development and gain more user trust in the near future, resulting in widespread application in health care. However, voice assistants are expected to solely support health care professionals in their daily operations and will not be able to outperform or replace medical staff.

(J Med Internet Res 2021;23(4):e25312) doi:10.2196/25312

KEYWORDS

Delphi study; medical informatics; voice-controlled intelligent personal assistants; internet of things; smart devices

Introduction

Overview

With their expanding capabilities, voice-controlled intelligent personal assistants (VIPAs), such as Amazon Echo and Google Home, profoundly change the way people interact with technology [1]. Using natural language processing (NLP) or natural language understanding (NLU), as well as cloud data storage, VIPAs can process nearly any human request in real time, mimicking natural human communication [2]. Following the recent *Dr Google* phenomenon [3], conversational agents

have quickly become a source of knowledge for health-related requests, providing information on various physical and mental health conditions [4]. Over the past years, advances in machine learning, particularly in neural networks, have enabled voice-based technologies to assist health professionals during medical consultation [5], provide diagnostic support [6], assist elderly people in their daily routines [7,8], and promote an overall healthier lifestyle [9]. Furthermore, intelligent assistants show potential in recognizing emotions and developing robot-human relationships with users [10,11], which can open a wide range of new applications for mental health patients, after-diagnosis support, and treatment of chronic conditions [12]. In particular, the current COVID-19 pandemic might trigger telemedical approaches involving VIPAs [13].

However, the potential to improve treatment efficiency and save costs comes with possible safety and privacy risks [5,14]. Incomplete or incorrect health-related information, such as first-aid instructions or medication recommendations, may lead to patient harm, especially when users blindly follow the instructions of VIPAs without understanding the limitations of the technology [15]. Serious privacy concerns about the use, disclosure, and protection of personal health data may also keep users from sharing health information and therefore have a large impact on customer acceptance of online health care applications [16].

Such conflicting influences on the adoption of VIPAs make it difficult to foresee their further development. At the same time, the disruptive potential of the technology in health care urges companies to innovate to be able to meet consumer requirements in the future [17]. The necessity to respond to market changes creates a need for an insightful technological forecast, which could support decision making within organizations, facilitate smooth innovation processes, and help companies to maintain or even advance their competitiveness [18]. Despite a wide selection of literature on intelligent voice assistants and their use in health care, the majority of studies have focused on specific therapy areas or functions, such as health coaching [19,20], disease detection [21], applications in psychotherapy [22,23], and others. The overall state and current progress of the field were recently examined by Laranjo et al and Montenegro et al [5,24]. However, studies using widely recognized foresight or forecasting methods, such as a Delphi study, that can shed light on the further development of conversational agents in health care do not exist to date. This paper aimed to fill this gap and present a comprehensive future scenario of the adoption of VIPAs in health care, based on a two-round expert online survey. In particular, the main technology trends, customer acceptance development, promising use cases, and possible regulation changes in the next 5 years have been examined.

With our findings, we contribute to ubiquitous health-related computing services (uHealth) and more specifically VIPAs by forecasting their further development in regard to technology, consumer acceptance, potential use cases, and privacy and data protection regulations, and by drawing practical conclusions for health care providers.

The remainder of the paper is organized as follows. The next section provides a brief technical background of VIPAs and covers the main factors affecting adoption in health care. The methodology section provides details on the Delphi technique and the main steps carried out during the study. Conclusively, the research findings are presented, followed by a discussion of research limitations and suggestions for future research.

Background

A voice assistant is an artificial intelligence–powered computer system that aims to imitate human intelligence while engaging in realistic conversations with users [25]. The current most popular examples are Amazon Alexa, Google Assistant, Apple Siri, Microsoft Cortana, and Samsung Bibxy [26]. As of 2020, VIPAs are integrated in numerous devices, like smartphones, speakers, smartwatches, smart televisions, cars, headphones, game consoles, and household appliances [26]. Unless the system is deactivated by the user, the software always listens for trigger keywords, such as “Alexa” and “Hey Siri,” and automatically starts audio recording when awakened [2]. The request is then transmitted to the cloud, processed using NLU, and assigned to a specific intention. Depending on the inquiry, the server will provide relevant information for the voice assistant to be presented to the user or execute tasks with numerous voice applications and connected devices [2].

Unlike previous intelligent systems, VIPAs can respond to much larger numbers of requests owing to the constant internet connection and access to rapidly growing amounts of services developed directly by VIPA providers or external companies like Uber, McDonalds, and Disney [2]. Because voice built-in capabilities (eg, “Alexa Skills” and “Google Actions”) are stored in the cloud, there is no need for users to download or install them in contrast to smartphone apps. Third-party voice extensions can be published for the general public in a voice app store like *Alexa Skills Store* or used exclusively within an organization, which provides a great opportunity for companies to use VIPAs for their specific needs [27].

Petrock highlights that 33.8% of people from the total US population are currently using a voice assistant at least monthly for various requests [26]. This number is predicted to increase to 36.6% in 2021 [26]. Constantly improving, VIPAs can execute over 100 different tasks, covering the ability to control home appliances, answer numerous questions, set reminders for medications or tasks, shop online, and others [7,28]. In the health care field, various research trials using virtual assistants have shown major advancements in physical training, diet adjustments, accessibility to health information, etc [5,29,30]. However, according to the number of applications already available on the market, the use of voice assistants in the health care context is in its early days in comparison to entertainment, information search, navigation, and calling [26]. Several innovative health care providers, like Mayo Clinic, Boston Children’s Hospital, Atrium Health, and Deloitte, have introduced their voice solutions for Alexa ranging from first aid help and home care treatment to medical appointment scheduling and patient communication [31]. Still, most health care companies are hesitant to adopt VIPAs, primarily due to legal compliance and privacy concerns. At present, given that

private health care data may be shared by a patient using a virtual assistant, health care providers have the option of getting access to Amazon's Health Insurance Portability and Accountability Act–compliant solution for US-based Alexa Skills or building their own voice assistant [31].

Another important barrier for the adoption of VIPAs in health care is privacy concerns. Amazon, Apple, Google, Facebook, and Microsoft have all been using third-party human contractors who listen to user audio recordings to improve the quality of NLU [31]. Since private consumer data had been shared without user consensus, major data privacy concerns can arise from the user side.

Methods

Delphi Study

To generate a plausible scenario for the use of VIPAs in health care, a two-stage Delphi study was conducted online. The Delphi method is a forecasting technique that relies on experts in a particular field to identify technology developments and trends [32]. According to research guidelines, experts anonymously provide their answers to a standardized questionnaire over at least two rounds [33]. The interim results of each round are summarized and fed back to the experts during the next stage to narrow the statistical spread and facilitate concurrence among the participants [34-38].

As Wright et al described, a Delphi study is especially suitable in cases requiring human judgmental input owing to missing historical or technical data, such as forecasting the development of emerging technologies [39]. In contrast to causal-deterministic natural development processes, such as the weather, societal changes are based on human intentions, social interactions, and coincidence [40]. Therefore, societal forecasts can be deduced from subjective expert knowledge and experience-based assessments [41]. The Delphi method surpasses similar interactive group techniques regarding accuracy and efficacy [42-44], and has been frequently used in various fields. In their bibliometric analysis, Flostrand et al found 175 Delphi-related papers in the business field and 1462 papers in the health care field in the period between 1975 and 2017 [45]. Although health care research often aims to find a consensus in general, without a future focus, it has also been applied to examine future developments and tendencies [46-49]. Therefore, the Delphi technique is appropriate to support health care providers' strategic decision making in regard to VIPA implementation.

Formulation of Projections

The topic under investigation is the future of VIPAs in health care within the next 5 years. This timeframe was used owing to the rapid development of the technology, which complicates forecasting the further development regarding VIPAs within a wider time span. Furthermore, the given time period serves the purpose of this study to provide health care organizations with clear guidance regarding the necessity and application of VIPAs in the near future.

According to our knowledge, the future of VIPAs in health care has not been examined in other studies yet. Hence, an exploratory approach was applied to achieve a broad rather than deep scenario. We determined four thematic sections (technology, consumer acceptance, potential use cases, and privacy and data protection regulations) that were supposed to be investigated. As a consequence of this multisection approach, we omitted deeper investigations of each topic in greater detail to ensure an appropriate length of the questionnaire and thus to avoid a low response and high dropout rate from the experts. We considered 20 projections as appropriate for the questionnaire. An equal number of projections per section was not considered necessary.

The formulation of the projections followed a systematic process. First, we conducted desk research (based on academic journals; publications by relevant health associations, consultancies, and state authorities; news articles; and social media posts) and brainstormed to identify relevant issues for the four sections. Second, a small expert panel consisting of two information technology (IT) and two health care experts reviewed the list of issues, without adding or removing items. Third, the issues were transformed into projections. We aimed to formulate the projections in a plausible way such that they could be agreed to. However, implausible statements could also be incorporated into a Delphi study, since both the acceptance of a plausible projection and the rejection of an implausible projection represent the same general trend and can equally contribute to the resulting scenario [50]. Fourth, the list of projections was forwarded to the small expert panel again for review. As a consequence, some formulations became more precise and concise. The final list of projections is presented in Textbox 1.

Textbox 1. Delphi projections (within the next 5 years).

Cluster 1: Technology

1. Voice-controlled intelligent personal assistants (VIPAs) will be able to give solid medical advice based on personal medical information, instructions, and plans.

2. VIPAs will be able to hold a human-like conversation.

3. VIPAs will be able to take into account patients' emotions, moods, and traits.

4. VIPAs will outperform humans in the quality of anamnesis and the ability to have the complete medical history available.

Cluster 2: Consumer Acceptance

5. From a patient's perspective, VIPAs will provide a frustration-free user experience.

6. A high number of patients will regularly use VIPAs for health-related inquiries.

7. Owing to simple use, a high share of elderly people will use VIPAs.

8. The majority of patients will prefer human communication over VIPA communication when in need of psychological support.

9. The majority of patients will prefer VIPAs owing to time and cost savings.

10. Patients will consider trust in content providers (owners of voice applications, such as hospitals and pharmaceutical companies) as more important than trust in VIPA providers (eg, Amazon and Google).

Cluster 3: Potential Use Cases

11. VIPAs will be widely used as remote and in-house real-time anamnesis tools.

12. VIPAs will be widely used as diagnostic support tools for diseases that can be detected through speech characteristics.

13. VIPAs will be widely used as hands-free instruction tools for medical staff (eg, in sterile environments).

14. VIPAs will be widely used as communication tools between medical staff and patients.

15. VIPAs will be widely used as self-therapy tools for patients outside the clinic.

Cluster 4: Privacy and Data Protection Regulations

16. The majority of VIPAs will be compliant with the applicable regulations regarding protected health information.

17. The majority of VIPA providers will discontinue the use of subcontractors to increase trust in data security.

18. Government privacy concerns will greatly decrease.

19. The majority of customers will deliberately share sensitive health information with VIPAs.

20. Too strict regulations will inhibit VIPAs to provide reliable high-quality medical help services.

Panelists

Expert selection is essential for the quality of the resulting forecast [36,51-54]. Following a purposive sampling approach [55], the experts were chosen based on their expertise in the field [33,54,56]. For this study, we searched for the following expert groups: (1) managers from VIPA manufacturers or developers (with job titles such as business development manager and product manager), (2) IT experts from VIPA manufacturers and developers (with job titles such as data scientist, software development engineer, and UX designer), (3) managers from hospitals (with job titles such as executive director, chief innovation officer, and chief technology officer), (4) physicians from hospitals, (5) researchers (especially professors) working in the academic fields of health care management, health care economics, or health care law, (6) representatives of state authorities related to health care and nongovernmental health care associations, and (7) health care consultants. Eligible participants were identified through scholarly publications, conference presentations, and field-specific expert awards related to VIPAs, and job descriptions or posts on the professional social network *LinkedIn*.

Overall, 154 experts were invited to complete the survey. Of these, 35 experts participated in the first round and 27 remained in the second round. The dropout rate of 22.9% is below the 30% rate that is considered acceptable [57,58]. The majority of Delphi panels involve 15 to 35 participants [57], acknowledging seven [59] to 10 [58] as the minimum. Especially in regard to the infancy of the field, the panel selection in this study is considered adequate to achieve the research goal. The panel structure is presented in Table 1. Unfortunately, we were unable to recruit participants from state authorities related to health care or from nongovernmental health care associations, as well as health care consultants. The low proportion of female participants represents the gender distribution in the field of artificial intelligence. Similarly, the prevalence of experts from the North, Central, and South America (AMER) region mirrors the relatively high level of voice technology development and VIPA application in health care in the United States.

Table 1. Panel demographics.

Characteristic	First round (n=35), n (%)	Second round (n=27), n (%)
Gender		
Male	25 (71)	17 (63)
Female	9 (26)	9 (33)
Other	1 (3)	1 (4)
Age (years)		
<30	3 (9)	1 (4)
30-40	10 (29)	9 (33)
41-50	9 (26)	7 (26)
51-60	10 (29)	8 (30)
>60	3 (9)	2 (7)
Affiliation		
Information technology expert (VIPA[a] firm)	5 (14)	4 (15)
Manager (VIPA firm)	9 (26)	8 (30)
Manager (hospital)	8 (23)	5 (19)
Physician (hospital)	7 (20)	3 (11)
Researcher (academia)	6 (17)	7 (26)
Representative (authorities or association)	0 (0)	0 (0)
Health care consultant	0 (0)	0 (0)
Region		
AMER[b]	17 (49)	17 (67)
EMEA[c]	14 (40)	9 (30)
APAC[d]	4 (11)	1 (4)

[a]VIPA: voice-controlled intelligent personal assistant.

[b]AMER: North, Central, and South America.

[c]EMEA: Europe, Middle-East, and Africa.

[d]APAC: Asia Pacific.

Data Collection

The questionnaire covered the 20 projections presented in Textbox 1. The experts were asked to provide their level of agreement with the future statements, using a 4-point Likert scale ("do not agree," "somewhat disagree," "somewhat agree," and "agree"). An even-numbered Likert scale was used to avoid neutral responses [41]. Further, demographic data, such as gender, age, profession, and location, were collected from the participants (Table 1).

To minimize the possibility of misinterpretation of the statements and prevent technical problems with the survey tool, pretests were conducted prior to each round. The first survey round was executed between September 18 and 29, 2020, followed by the second round between October 7 and 14, 2020.

Results

Descriptive Statistics

According to Delphi study guidelines, the most probable future scenario is generated from the accumulation of the group consensus on each projection. The panel's agreement on each statement is concluded from the aggregation of individual assessments. The following numerical values were assigned to the response options to enable the calculation of statistical distribution: "do not agree," 1; "somewhat disagree," 2; "somewhat agree," 3; and "agree," 4.

Considering resistance toward outliers and the potential risk of statistical biases, the median is favored over the mean as the statistical average for Delphi studies [36,41]. Scattering of the responses is evaluated by IQR, which is the difference between the upper quartile ($x_{0.75}$) and the lower quartile ($x_{0.25}$). A low IQR indicates a high level of agreement, whereas a high IQR reflects a high level of discord [41].

Table 2 depicts the detailed results for both rounds. In the second round, no projection received strong agreement (median 4). The experts somewhat agreed (median 3) with most of the projections, namely projections 1, 2, 6, 7, 8, 10, 11, 13, 14, 15, 16, and 19. They somewhat disagreed with projections 3, 4, 5, 9, 12, 17, and 20, and even strongly disagreed with projection 18.

During the second round, only slight changes occurred. The median did not change for the majority of the projections. For projections 3 and 5, the experts revised their assessment from partial agreement to partial disagreement. Further, the median changed from slight to strong disagreement for projection 18. The IQR did not rise for any projection except for projection 9, showing increased insecurity in the panel, but reduced (as aimed) for eight projections, namely projections 2, 4, 7, 10, 13, 16, 17, and 18. Generally, the IQR evaluation shows a high level of consensus among participants. Additionally, to measure the level of consensus, Kendall *W* coefficient of concordance was calculated for both rounds [60]. The value can vary from 0 to 1, but is somewhat difficult to interpret, as it does not depict a linear rise. However, closer proximity to 1 reflects stronger agreement [61,62]. Kendall *W* equaled 0.1188 (χ^2_{19}=78.9689, P<.001) for the first round and 0.2948 (χ^2_{19}=151.2496, P<.001) for the second round of the Delphi study. As expected, the level of consensus among panelists increased in the course of the study. In light of the level of consensus and only minor variations between the rounds, no further iterations were required.

Table 2. Descriptive statistics.

Cluster and projection	First round (n=35)				Second round (n=27)				Difference			
	$X_{0.25}$[a]	$X_{0.5}$[b]	$X_{0.75}$[c]	IQR	$X_{0.25}$	$X_{0.5}$	$X_{0.75}$	IQR	$X_{0.25}$	$X_{0.5}$	$X_{0.75}$	IQR
Cluster 1: Technology												
1	2	3	3	1	2	3	3	1	0	0	0	0
2	2	3	4	2	2	3	3	1	0	0	-1	-1
3	2	3	3	1	2	2	3	1	0	-1	0	0
4	1	2	3	2	1	2	2	1	0	0	-1	-1
Cluster 2: Consumer acceptance												
5	2	3	3	1	2	2	3	0	-1	0	0	0
6	3	3	4	1	3	3	4	1	0	0	0	0
7	2	3	4	2	2	3	3	1	0	0	-1	-1
8	2	3	4	2	2	3	4	2	0	0	0	0
9	2	2	3	1	1	2	3	2	-1	0	0	1
10	2	3	4	2	3	3	4	1	1	0	0	-1
Cluster 3: Potential use cases												
11	2	3	3	1	2	3	3	1	0	0	0	0
12	2	2	3	1	2	2	3	1	0	0	0	0
13	2	3	4	2	2	3	3	1	0	0	-1	-1
14	2	3	3	1	2	3	3	1	0	0	0	0
15	2	3	3	1	2	3	3	1	0	0	0	0
Cluster 4: Privacy and data protection regulations												
16	2	3	4	2	3	3	3	0	1	0	-1	-2
17	2	2	3	1	2	2	2	0	0	0	-1	-1
18	1	2	3	2	1	1	2	1	0	-1	-1	-1
19	2	3	3	1	2	3	3	1	0	0	0	0
20	2	2	3	1	2	2	3	1	0	0	0	0

[a]$X_{0.25}$: lower quartile.

[b]$X_{0.5}$: median quartile.

[c]$X_{0.75}$: upper quartile.

Scenario

According to the surveyed experts, VIPAs will show notable technological development and gain more trust among users in the next 5 years, resulting in their widespread utilization in health care. However, voice assistants are expected to only support health care professionals in their daily operations and will not be able to outperform or replace medical staff.

Within the next 5 years, the respondents expect a high percentage of patients to use VIPAs for health-related inquiries on a regular basis. Remarkably, a large proportion of elderly people will use voice assistants owing to their screen-free simple interface. Although experts agreed on the ability of VIPAs to hold a human-like conversation in the next 5 years, the given timeframe is still assessed as too short for the technology to be able to provide a frustration-free user experience.

Anticipating extensive use of voice assistants in health care, experts mentioned communicational, anamnesis, and self-therapy tools for patients and medical staff as the most promising use cases. Still, questioning the ability of VIPAs to recognize and take patients' emotions, moods, and traits into consideration, wide application of conversational assistants as diagnostic tools is not expected by 2025.

Pursuant to the survey outcome, the majority of voice assistants will be compliant with the applicable health information regulations within the time span of 5 years. Furthermore, data compliance together with trust in content providers, such as hospitals and pharmaceutical companies, might result in higher user willingness to share sensitive health information with VIPAs. While experts strongly rejected the possibility of a decrease in governmental privacy concerns, strict regulations capable of preventing VIPAs from providing medical help services are not expected.

Discussion

Discussion of the Results

In the technology section, the experts expressed discordant opinions about three of four projections. While the majority of responders agreed with the ability of VIPAs to provide solid medical advice based on patients' medical information, instructions, and plans within the next 5 years, 48% of the panelists rejected the statement. Hence, no clear group opinion could be derived for projection 1. Since not only technological development but also a comprehensive legal framework is required to enable voice assistants to provide medical help, some participants might reject the projection taking into account broad legal and privacy concerns. The presence of expert preoccupation with legal issues was confirmed in the fourth cluster of this study.

Likewise, projection 2 was accepted by the majority (52%) of panelists. Still, 48% of panelists expressed doubt about the potential of VIPAs to have a human-like conversation within the next 5 years. The group skepticism regarding the ability of VIPAs to accurately imitate human skills is reinforced in projections 3, 4, 5, 8, and 9. In detail, 81% of the interviewees believed voice assistants will not be able to outperform humans in the quality of anamnesis in the next 5 years. Projection 5, which stated that VIPAs can provide a frustration-free user experience, was also rejected. Furthermore, 66% of the panel expected patients to prefer humans over voice assistants when in need of psychological support.

Questioning the likelihood of the interchangeability between voice assistants and medical staff in the near future, the experts agreed with the widespread use of VIPAs as support tools. With a high level of consensus, 85% of the panel members anticipated a high number of patients to regularly use VIPAs for health-related inquires. Notably, no expert strongly disagreed with the projection.

Examining the most promising use cases, the panelists agreed on the application of VIPAs by medical staff (eg, in sterile environments, with a majority of 74%). Accordingly, 70% of the group expected extensive use of conversational agents among elderly people and 73% believed patients and health care professionals will use voice assistants for communication. Projection 15 suggesting the use of voice assistants as self-therapy tools was accepted by 59% of the experts. The fact that 41% of the responders expressed doubt about the given use case might reflect the general trend of expert skepticism toward the employment of voice assistants without the presence of health professionals.

The assumption of a potential decrease in the government's privacy concerns was strongly rejected. Only one expert somewhat agreed with the statement, with the rest of the group showing slight or strong disagreement. Furthermore, 85% of the panelists declared they do not expect VIPA providers to stop using subcontractors who listen to user audio files to improve the quality of NLU. Thus, data concerns might remain in place, requiring a new set of regulations. Further research is required to obtain a better understanding of the available tools and the amount of time it would take to address regulatory issues.

From the scenario, several practical implications for health care providers can be concluded, which have been summarized in Textbox 2.

Textbox 2. Practical implications for health care providers (within the next 5 years).

- Voice-controlled intelligent personal assistant (VIPA) technology will be mature.
- VIPAs will be accepted by patients, including elderly people.
- VIPAs will be used for regular anamneses, medical staff support, staff-patient communication, and self-therapy.
- Privacy and data protection issues will not harm the dissemination of VIPAs.

Limitations and Future Research

As this study involves early research on the future use of VIPAs in health care, several limitations apply to this study, which could also provide guidance for future research. First, the Delphi technique develops the most probable future scenario based on experts' present knowledge. Thus, although the method outperforms comparable interactive group forecasting techniques [42-44], the Delphi study cannot guarantee the exact realization of the forecast. Second, the range of addressed questions was limited to keep a sufficient length of the survey and avoid low response rates from the experts. Therefore, only existing features and use cases, which are currently in the early days of their implementation, were examined. The probability of the emergence of new technologies and applications for VIPAs was not investigated.

This study aimed to provide clear guidance for health care companies regarding the current necessity to focus their efforts on voice technology. Thus, a relatively short time frame of 5 years was chosen. Especially for the Europe, Middle-East, and Africa (EMEA) and Asia Pacific (APAC) regions, where the use of voice assistants in health care is in its infancy, the given time span could potentially interfere with the experts' future assumptions.

Currently, the use of voice assistants in health care varies widely across countries. The quality of NLP/NLU in English greatly outperforms other languages. Further factors, including the total user base, tools available for third-party developers, data protection regulations, and cultural differences, determine the adoption of VIPAs in general and in particular in health care. During this study, no large differences were detected comparing AMER, EMEA, and APAC expert assessments. Still, a study covering a specific region might provide more insightful forecasts of VIPA development in a particular geographical area.

Despite the widespread use of the Delphi technique, a number of claims have been made regarding its methodology. First, anonymity is one of the central characteristics of the Delphi technique, which aims to encourage experts to provide true and not socially likeable assessments. However, Sackman points out that anonymity can lead to hasty judgments as a result of experts' assurance that there is no necessity to defend their responses [63]. Second, the Delphi technique requires disclosure of the interim results of each round to generate a group opinion that can be claimed to be representative [64]. Some scholars argue that independent judgement is violated once the panelists know how others have evaluated each item [64]. Further, this disclosure can encourage outliers to revise their assessments owing to group pressure and not because of a changed opinion. Hence, a reduced IQR rate in Delphi studies can correspond to the problem of group thinking rather than higher consensus among experts [36,64,65].

Conclusion

We conducted an international Delphi study and derived a plausible scenario of the future of voice assistants in health care within the next 5 years. Twenty projections were designed and evaluated by an internationally recruited panel consisting of voice experts, as well as medical professionals and representatives of academia, health authorities, and nongovernmental health associations having broad experience with voice technology.

With a high level of consensus, the experts anticipate widespread application of VIPAs in various health care domains in the next 5 years. Although conversational assistants are not expected to replace medical workers, their use as operational supporting tools for health care professionals has strong potential in the industry. In detail, the panelists agreed with the capability of VIPAs to support elderly people and to be widely used as anamnesis, informational, self-therapy, and communicational tools by patients and health care professionals. Although users' and governments' privacy concerns are not expected to decrease in the near future, the panel members believe that strict regulations capable of preventing VIPAs from providing medical help services will not be imposed. To be able to meet consumer expectations and withstand competition within the next 5 years, health care companies are advised to carefully observe current research and development activities in the field of conversational artificial intelligence and allocate resources to optimize business processes using VIPAs.

This was an exploratory study on the future of voice assistants in health care. Hence, a broad rather than deep scenario approach was applied. Further studies might take a deeper look at one of the four clusters examined in this study or focus on a specific geographical area to provide more detailed insights.

Conflicts of Interest

None declared.

References

1. McLean G, Osei-Frimpong K. Hey Alexa … examine the variables influencing the use of artificial intelligent in-home voice assistants. Computers in Human Behavior 2019 Oct;99:28-37. [doi: 10.1016/j.chb.2019.05.009]
2. Hoy MB. Alexa, Siri, Cortana, and More: An Introduction to Voice Assistants. Med Ref Serv Q 2018;37(1):81-88. [doi: 10.1080/02763869.2018.1404391] [Medline: 29327988]
3. Cocco AM, Zordan R, Taylor DM, Weiland TJ, Dilley SJ, Kant J, et al. Dr Google in the ED: searching for online health information by adult emergency department patients. Med J Aust 2018 Oct 15;209(8):342-347. [doi: 10.5694/mja17.00889] [Medline: 30107763]

4. Miner AS, Milstein A, Schueller S, Hegde R, Mangurian C, Linos E. Smartphone-Based Conversational Agents and Responses to Questions About Mental Health, Interpersonal Violence, and Physical Health. JAMA Intern Med 2016 May 01;176(5):619-625 [FREE Full text] [doi: 10.1001/jamainternmed.2016.0400] [Medline: 26974260]
5. Laranjo L, Dunn AG, Tong HL, Kocaballi AB, Chen J, Bashir R, et al. Conversational agents in healthcare: a systematic review. J Am Med Inform Assoc 2018 Sep 01;25(9):1248-1258 [FREE Full text] [doi: 10.1093/jamia/ocy072] [Medline: 30010941]
6. Philip P, Micoulaud-Franchi J, Sagaspe P, Sevin ED, Olive J, Bioulac S, et al. Virtual human as a new diagnostic tool, a proof of concept study in the field of major depressive disorders. Sci Rep 2017 Feb 16;7:42656 [FREE Full text] [doi: 10.1038/srep42656] [Medline: 28205601]
7. O'Brien K, Liggett A, Ramirez-Zohfeld V, Sunkara P, Lindquist LA. Voice-Controlled Intelligent Personal Assistants to Support Aging in Place. J Am Geriatr Soc 2020 Jan 16;68(1):176-179. [doi: 10.1111/jgs.16217] [Medline: 31617581]
8. Nallam P, Bhandari S, Sanders J, Martin-Hammond A. A Question of Access: Exploring the Perceived Benefits and Barriers of Intelligent Voice Assistants for Improving Access to Consumer Health Resources Among Low-Income Older Adults. Gerontol Geriatr Med 2020 Dec 29;6:2333721420985975-2333721420985979 [FREE Full text] [doi: 10.1177/2333721420985975] [Medline: 33457459]
9. Sezgin E, Militello L, Huang Y, Lin S. A Scoping Review of Patient-Facing, Behavioral Health Interventions with Voice Assistant Technology Targeting Self-management and Healthy Lifestyle Behaviors. SSRN. 2019. URL: https://ssrn.com/abstract=3381183 [accessed 2021-03-05]
10. Zhou S, Jia J, Wang Y, Chen W, Meng F, Li Y, et al. Emotion Inferring from Large-scale Internet Voice Data: A Multimodal Deep Learning Approach. 2018 Presented at: 2018 First Asian Conference on Affective Computing and Intelligent Interaction (ACII Asia); May 20-22, 2018; Beijing, China. [doi: 10.1109/ACIIAsia.2018.8470311]
11. Schweitzer F, Belk R, Jordan W, Ortner M. Servant, friend or master? The relationships users build with voice-controlled smart devices. Journal of Marketing Management 2019 Apr 05;35(7-8):693-715. [doi: 10.1080/0267257x.2019.1596970]
12. Wolters MK, Kelly F, Kilgour J. Designing a spoken dialogue interface to an intelligent cognitive assistant for people with dementia. Health Informatics J 2016 Dec;22(4):854-866 [FREE Full text] [doi: 10.1177/1460458215593329] [Medline: 26276794]
13. Sezgin E, Huang Y, Ramtekkar U, Lin S. Readiness for voice assistants to support healthcare delivery during a health crisis and pandemic. NPJ Digit Med 2020 Sep 16;3:122 [FREE Full text] [doi: 10.1038/s41746-020-00332-0] [Medline: 33015374]
14. Guarda P. "Ok Google, am I sick?": artificial intelligence, e-health, and data protection regulation. BioLaw Journal 2019;15(1):359-375 [FREE Full text]
15. Bickmore TW, Trinh H, Olafsson S, O'Leary TK, Asadi R, Rickles NM, et al. Patient and Consumer Safety Risks When Using Conversational Assistants for Medical Information: An Observational Study of Siri, Alexa, and Google Assistant. J Med Internet Res 2018 Sep 04;20(9):e11510 [FREE Full text] [doi: 10.2196/11510] [Medline: 30181110]
16. Bansal G, Zahedi FM, Gefen D. The impact of personal dispositions on information sensitivity, privacy concern and trust in disclosing health information online. Decis Support Syst 2010 May;49(2):138-150. [doi: 10.1016/j.dss.2010.01.010]
17. Khan R. Standardized Architecture for Conversational Agents a.k.a. ChatBots. IJCTT 2017 Aug 25;50(2):114-121. [doi: 10.14445/22312803/ijctt-v50p120]
18. Rohrbeck R, Schwarz JO. The value contribution of strategic foresight: Insights from an empirical study of large European companies. Technological Forecasting and Social Change 2013 Oct;80(8):1593-1606. [doi: 10.1016/j.techfore.2013.01.004]
19. Fadhil A, Wang Y, Reiterer H. Assistive Conversational Agent for Health Coaching: A Validation Study. Methods Inf Med 2019 Jun;58(1):9-23. [doi: 10.1055/s-0039-1688757] [Medline: 31117129]
20. Tanaka H, Negoro H, Iwasaka H, Nakamura S. Embodied conversational agents for multimodal automated social skills training in people with autism spectrum disorders. PLoS One 2017;12(8):e0182151 [FREE Full text] [doi: 10.1371/journal.pone.0182151] [Medline: 28796781]
21. Moro-Velázquez L, Gómez-García JA, Godino-Llorente JI, Villalba J, Orozco-Arroyave JR, Dehak N. Analysis of speaker recognition methodologies and the influence of kinetic changes to automatically detect Parkinson's Disease. Applied Soft Computing 2018 Jan;62:649-666. [doi: 10.1016/j.asoc.2017.11.001]
22. Miner AS, Shah N, Bullock KD, Arnow BA, Bailenson J, Hancock J. Key Considerations for Incorporating Conversational AI in Psychotherapy. Front Psychiatry 2019;10:746 [FREE Full text] [doi: 10.3389/fpsyt.2019.00746] [Medline: 31681047]
23. Vaidyam AN, Wisniewski H, Halamka JD, Kashavan MS, Torous JB. Chatbots and Conversational Agents in Mental Health: A Review of the Psychiatric Landscape. Can J Psychiatry 2019 Jul;64(7):456-464 [FREE Full text] [doi: 10.1177/0706743719828977] [Medline: 30897957]
24. Montenegro JLZ, da Costa CA, da Rosa Righi R. Survey of conversational agents in health. Expert Systems with Applications 2019 Sep;129:56-67. [doi: 10.1016/j.eswa.2019.03.054]
25. Bibault J, Chaix B, Nectoux P, Pienkowsky A, Guillemasse A, Brouard B. Healthcare ex Machina: Are conversational agents ready for prime time in oncology? Clin Transl Radiat Oncol 2019 May;16:55-59 [FREE Full text] [doi: 10.1016/j.ctro.2019.04.002] [Medline: 31008379]
26. Petrock V. US Voice Assistant Users 2019: Who, What, When, Where and Why. eMarketer. 2019. URL: https://www.emarketer.com/content/us-voice-assistant-users-2019 [accessed 2021-03-05]

27. Alexa for Business. Amazon. URL: https://aws.amazon.com/alexaforbusiness/ [accessed 2020-06-15]
28. Rhee CE, Choi J. Effects of personalization and social role in voice shopping: An experimental study on product recommendation by a conversational voice agent. Computers in Human Behavior 2020 Aug;109:106359. [doi: 10.1016/j.chb.2020.106359]
29. Bickmore TW, Schulman D, Sidner C. Automated interventions for multiple health behaviors using conversational agents. Patient Educ Couns 2013 Aug;92(2):142-148 [FREE Full text] [doi: 10.1016/j.pec.2013.05.011] [Medline: 23763983]
30. Edwards RA, Bickmore T, Jenkins L, Foley M, Manjourides J. Use of an interactive computer agent to support breastfeeding. Matern Child Health J 2013 Dec;17(10):1961-1968. [doi: 10.1007/s10995-013-1222-0] [Medline: 23329167]
31. Voice Assistant Consumer Adoption Report for Healthcare 2019. Voicebot AI. 2019. URL: https://voicebot.ai/voice-assistant-consumer-adoption-report-for-healthcare-2019/ [accessed 2021-03-05]
32. Jiang R, Kleer R, Piller FT. Predicting the future of additive manufacturing: A Delphi study on economic and societal implications of 3D printing for 2030. Technological Forecasting and Social Change 2017 Apr;117:84-97. [doi: 10.1016/j.techfore.2017.01.006]
33. Hasson F, Keeney S, McKenna H. Research guidelines for the Delphi survey technique. J Adv Nurs 2000;32(4):1008-1015. [doi: 10.1046/j.1365-2648.2000.t01-1-01567.x]
34. Dalkey N, Helmer O. An Experimental Application of the DELPHI Method to the Use of Experts. Management Science 1963 Apr;9(3):458-467. [doi: 10.1287/mnsc.9.3.458]
35. Melander L, Dubois A, Hedvall K, Lind F. Future goods transport in Sweden 2050: Using a Delphi-based scenario analysis. Technological Forecasting and Social Change 2019 Jan;138:178-189. [doi: 10.1016/j.techfore.2018.08.019]
36. Rowe G, Wright G. The Delphi technique as a forecasting tool: issues and analysis. International Journal of Forecasting 1999 Oct;15(4):353-375. [doi: 10.1016/S0169-2070(99)00018-7]
37. Tiberius V, Borning J, Seeler S. Setting the table for meat consumers: an international Delphi study on in vitro meat. NPJ Sci Food 2019 Jun 4;3:10 [FREE Full text] [doi: 10.1038/s41538-019-0041-0] [Medline: 31304282]
38. Woudenberg F. An evaluation of Delphi. Technological Forecasting and Social Change 1991 Sep;40(2):131-150. [doi: 10.1016/0040-1625(91)90002-w]
39. Wright G, Lawrence MJ, Collopy F. The role and validity of judgment in forecasting. International Journal of Forecasting 1996 Mar;12(1):1-8. [doi: 10.1016/0169-2070(96)00674-7]
40. Tiberius V. Towards a "Planned Path Emergence" View on Future Genesis. Journal of Futures Studies 2011;15(4):9-24 [FREE Full text]
41. Tiberius V, Hirth S. Impacts of digitization on auditing: A Delphi study for Germany. Journal of International Accounting, Auditing and Taxation 2019 Dec;37:100288 [FREE Full text] [doi: 10.1016/j.intaccaudtax.2019.100288]
42. Graefe A, Armstrong JS. Comparing face-to-face meetings, nominal groups, Delphi and prediction markets on an estimation task. International Journal of Forecasting 2011 Jan;27(1):183-195. [doi: 10.1016/j.ijforecast.2010.05.004]
43. Parente R, Anderson-Parente J. A case study of long-term Delphi accuracy. Technological Forecasting and Social Change 2011 Nov;78(9):1705-1711. [doi: 10.1016/j.techfore.2011.07.005]
44. Rowe G, Wright G, McColl A. Judgment change during Delphi-like procedures: The role of majority influence, expertise, and confidence. Technological Forecasting and Social Change 2005 May;72(4):377-399. [doi: 10.1016/j.techfore.2004.03.004]
45. Flostrand A, Pitt L, Bridson S. The Delphi technique in forecasting– A 42-year bibliographic analysis (1975–2017). Technological Forecasting and Social Change 2020 Jan;150:119773. [doi: 10.1016/j.techfore.2019.119773]
46. Haluza D, Jungwirth D. ICT and the future of health care: aspects of doctor-patient communication. Int J Technol Assess Health Care 2014 Jul;30(3):298-305. [doi: 10.1017/S0266462314000294] [Medline: 25308693]
47. Haluza D, Jungwirth D. ICT and the future of health care: aspects of health promotion. Int J Med Inform 2015 Jan;84(1):48-57. [doi: 10.1016/j.ijmedinf.2014.09.005] [Medline: 25293532]
48. Norcross JC, Hedges M, Prochaska JO. The face of 2010: A Delphi poll on the future of psychotherapy. Professional Psychology: Research and Practice 2002;33(3):316-322. [doi: 10.1037/0735-7028.33.3.316]
49. Starkweather DB, Gelwicks L, Newcomer R. Delphi forecasting of health care organization. Inquiry 1975 Mar;12(1):37-46. [Medline: 123221]
50. Höhne S, Tiberius V. Powered by blockchain: forecasting blockchain use in the electricity market. IJESM 2020 May 11;14(6):1221-1238 [FREE Full text] [doi: 10.1108/ijesm-10-2019-0002]
51. Møldrup C, Morgall JM. Risks of Future Drugs. Technological Forecasting and Social Change 2001 Jun;67(2-3):273-289. [doi: 10.1016/s0040-1625(99)00061-x]
52. Tersine RJ, Riggs WE. The delphi technique: A long-range planning tool. Business Horizons 1976 Apr;19(2):51-56. [doi: 10.1016/0007-6813(76)90081-1]
53. Webler T, Levine D, Rakel H, Renn O. A novel approach to reducing uncertainty: The group Delphi. Technological Forecasting and Social Change 1991 May;39(3):253-263. [doi: 10.1016/0040-1625(91)90040-m]
54. Welty G. Communications. Problems of selecting experts for Delphi exercises. Academy of Management Journal 1972 Mar 01;15(1):121-124. [doi: 10.2307/254805]
55. Guest G, Bunce A, Johnson L. How Many Interviews Are Enough? Field Methods 2016 Jul 21;18(1):59-82. [doi: 10.1177/1525822X05279903]

56. Winkler J, Moser R. Biases in future-oriented Delphi studies: A cognitive perspective. Technological Forecasting and Social Change 2016 Apr;105:63-76. [doi: 10.1016/j.techfore.2016.01.021]
57. Gordon TJ. The Delphi method. Futures Research Methodology 1994:1-30 [FREE Full text]
58. Häder M. Delphi-Befragungen: Ein Arbeitsbuch. Wiesbaden, Germany: Springer; 2014.
59. Dalkey N, Brown B, Cochran S. Use of self-ratings to improve group estimates. Technological Forecasting 1970 Mar;1(3):283-291. [doi: 10.1016/0099-3964(70)90029-3]
60. Meijering J, Kampen J, Tobi H. Quantifying the development of agreement among experts in Delphi studies. Technological Forecasting and Social Change 2013 Oct;80(8):1607-1614. [doi: 10.1016/j.techfore.2013.01.003]
61. Schmidt RC. Managing Delphi Surveys Using Nonparametric Statistical Techniques. Decision Sciences 1997 Jul;28(3):763-774. [doi: 10.1111/j.1540-5915.1997.tb01330.x]
62. von der Gracht HA. Consensus measurement in Delphi studies. Technological Forecasting and Social Change 2012 Oct;79(8):1525-1536. [doi: 10.1016/j.techfore.2012.04.013]
63. Sackman H. Summary Evaluation of Delphi. Policy Analysis 1975;1(4):693-718 [FREE Full text]
64. Goodman CM. The Delphi technique: a critique. J Adv Nurs 1987 Nov;12(6):729-734. [doi: 10.1111/j.1365-2648.1987.tb01376.x] [Medline: 3320139]
65. Bardecki MJ. Participants' response to the Delphi method: An attitudinal perspective. Technological Forecasting and Social Change 1984 May;25(3):281-292. [doi: 10.1016/0040-1625(84)90006-4]

Abbreviations

AMER: North, Central, and South America
APAC: Asia Pacific
EMEA: Europe, Middle-East, and Africa
IT: information technology
NLP: natural language processing
NLU: natural language understanding
VIPA: voice-controlled intelligent personal assistant

Edited by R Kukafka; submitted 27.10.20; peer-reviewed by B Chaudhry, P Bican, R De Carvalho, T O'Leary; comments to author 14.01.21; revised version received 05.03.21; accepted 16.03.21; published 09.04.21.

Please cite as:
Ermolina A, Tiberius V
Voice-Controlled Intelligent Personal Assistants in Health Care: International Delphi Study
J Med Internet Res 2021;23(4):e25312
URL: https://www.jmir.org/2021/4/e25312
doi:10.2196/25312
PMID:33835032

Original Paper

Use of Self-Reported Computerized Medical History Taking for Acute Chest Pain in the Emergency Department – the Clinical Expert Operating System Chest Pain Danderyd Study (CLEOS-CPDS): Prospective Cohort Study

Helge Brandberg[1], MD; Carl Johan Sundberg[2,3], MD, PhD; Jonas Spaak[1], MD, PhD; Sabine Koch[2], MSc, PhD; David Zakim[2], MD; Thomas Kahan[1], MD, PhD

[1]Division of Cardiovascular Medicine, Department of Clinical Sciences, Danderyd Hospital, SE-182 88 Stockholm, Sweden

[2]Medical Management Centre and Health Informatics Centre, Department of Learning, Informatics, Management and Ethics, Karolinska Institutet, Stockholm, Sweden

[3]Department of Physiology & Pharmacology, Karolinska Institutet, Stockholm, Sweden

Corresponding Author:
Helge Brandberg, MD
Division of Cardiovascular Medicine
Department of Clinical Sciences, Danderyd Hospital
Karolinska Institutet
SE-182 88 Stockholm
Sweden
Phone: +46 8 123 58347
Email: helge.brandberg@ki.se

Abstract

Background: Chest pain is one of the most common chief complaints in emergency departments (EDs). Collecting an adequate medical history is challenging but essential in order to use recommended risk scores such as the HEART score (based on history, electrocardiogram, age, risk factors, and troponin). Self-reported computerized history taking (CHT) is a novel method to collect structured medical history data directly from the patient through a digital device. CHT is rarely used in clinical practice, and there is a lack of evidence for utility in an acute setting.

Objective: This substudy of the Clinical Expert Operating System Chest Pain Danderyd Study (CLEOS-CPDS) aimed to evaluate whether patients with acute chest pain can interact effectively with CHT in the ED.

Methods: Prospective cohort study on self-reported medical histories collected from acute chest pain patients using a CHT program on a tablet. Clinically stable patients aged 18 years and older with a chief complaint of chest pain, fluency in Swedish, and a nondiagnostic electrocardiogram or serum markers for acute coronary syndrome were eligible for inclusion. Patients unable to carry out an interview with CHT (eg, inadequate eyesight, confusion or agitation) were excluded. Effectiveness was assessed as the proportion of patients completing the interview and the time required in order to collect a medical history sufficient for cardiovascular risk stratification according to HEART score.

Results: During 2017-2018, 500 participants were consecutively enrolled. The age and sex distribution (mean 54.3, SD 17.0 years; 213/500, 42.6% women) was similar to that of the general chest pain population (mean 57.5, SD 19.2 years; 49.6% women). Common reasons for noninclusion were language issues (182/1000, 18.2%), fatigue (158/1000, 15.8%), and inability to use a tablet (152/1000, 15.2%). Sufficient data to calculate HEART score were collected in 70.4% (352/500) of the patients. Key modules for chief complaint, cardiovascular history, and respiratory history were completed by 408 (81.6%), 339 (67.8%), and 291 (58.2%) of the 500 participants, respectively, while 148 (29.6%) completed the entire interview (in all 14 modules). Factors associated with completeness were age 18-69 years (all key modules: *P*s<.001), male sex (cardiovascular: *P*=.04), active workers (all key modules: *P*s<.005), not arriving by ambulance (chief complaint: *P*=.03; cardiovascular: *P*=.045), and ongoing chest pain (complete interview: *P*=.002). The median time to collect HEART score data was 23 (IQR 18-31) minutes and to complete an interview was 64 (IQR 53-77) minutes. The main reasons for discontinuing the interview prior to completion (n=352) were discharge from the ED (101, 28.7%) and tiredness (95, 27.0%).

Conclusions: A majority of patients with acute chest pain can interact effectively with CHT on a tablet in the ED to provide sufficient data for risk stratification with a well-established risk score. The utility was somewhat lower in patients 70 years and older, in patients arriving by ambulance, and in patients without ongoing chest pain. Further studies are warranted to assess whether CHT can contribute to improved management and prognosis in this large patient group.

Trial Registration: ClinicalTrials.gov NCT03439449; https://clinicaltrials.gov/ct2/show/NCT03439449

International Registered Report Identifier (IRRID): RR2-10.1136/bmjopen-2019-031871

(J Med Internet Res 2021;23(4):e25493) doi:10.2196/25493

KEYWORDS

chest pain; computerized history taking; coronary artery disease; eHealth; emergency department; health informatics; medical history; risk management

Introduction

Chest pain is one of the most common chief complaints in emergency departments (EDs) worldwide [1,2]. Aside from electrocardiogram (ECG) and cardiac biomarkers, the medical history is regarded as central for management [3,4]. However, collecting an adequate medical history is a challenge for the physician due to limited time and is seldom done in a systematic, standardized way [5]. To improve chest pain management, emphasis has been put on developing new algorithms and advanced examinations [6-11].

Self-reported computerized history taking (CHT) is a method to collect a structured medical history by direct interaction between patients and a digital device. The concept of standardized history taking with structured paper questionnaires had already appeared in the 1940s [12]. The first software for CHT emerged in the 1960s [13]. Numerous CHT software programs have been developed and shown to collect more detailed data, as compared with conventional questionnaires [14]. CHT has the benefit of being reliable, as it never forgets to pose a question or diverges from what it is programmed to do [5]. As well, it can interpret the data instantly [15], which could aid the physician with complex information processing in a hectic environment (eg, triage in the ED). In several studies, CHT software collected more documented information than the physician (eg, in psychiatric history taking [16], outpatients with gastrointestinal symptoms [17] or dyslipidemia [18]). For the patient, highlighted benefits are that there is good acceptance of the software [14,16]; the patient is more likely to share sensitive information [14]; and consultation can be focused on identifying concerns and problems, rather than history taking [19]. The main disadvantages raised are irrelevant questioning, technical issues, and the programs' lack of empathy and inability to interpret body language [5,19].

Despite promising results, CHT is rarely used in clinical practice [5]. In 2007, a small feasibility study [20] including 64 patients showed that CHT was well accepted, that it collected an appropriate medical history of the various ED chief complaints, and that the concept could successfully be integrated with the process. However, there are only occasional studies on CHT in the acute cardiology setting or for ED patients with an acute complaint [20,21]. Indeed, the authors of a recent review for CHT in the management of cardiovascular disease concluded that there is a need to develop an evidence base for the use of CHT in this area of practice [22].

The overall aim of the Clinical Expert Operating System Chest Pain Danderyd Study (CLEOS-CPDS; ClinicalTrials.gov identifier: NCT03439449) is to determine the value of self-reported CHT for acute chest pain management [23]. This substudy is a utility study among the first 500 patients included, aimed to evaluate whether chest pain patients can effectively interact with CHT in the ED. Effectiveness was assessed as the proportion of patients completing the CHT interview and the time required to collect a medical history sufficient for cardiovascular risk stratification with an established risk score (ie, HEART [history, ECG, age, risk factors, and troponin] score; see below).

Methods

Setting

The CLEOS-CPDS study is an ongoing prospective cohort study recruiting consecutive patients presenting at the ED at Danderyd University Hospital (Stockholm, Sweden) from October 1, 2017, to December 31, 2023 (preliminary). The study has been described elsewhere [23] and has been approved by the Stockholm Regional Ethical Committee (now Swedish Ethical Review Authority) (reference number 2015/1955-1).

Study Population

Clinically stable women and men (Rapid Emergency Triage and Treatment System [RETTS] level orange, yellow, green, and blue [24]) aged 18 years and older with a chief complaint of chest pain, fluency in Swedish, and a nondiagnostic first ECG or serum markers for acute coronary syndrome (ACS) were eligible for inclusion after providing informed consent. Patients unable to carry out an interview with CHT (eg, inadequate eyesight, confusion, or agitation) were excluded.

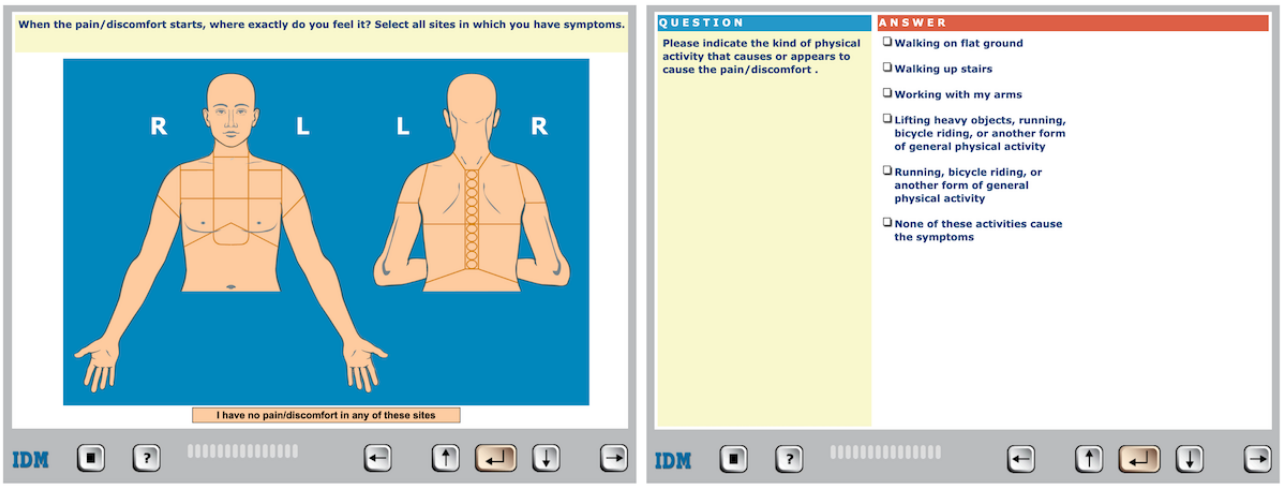
This study included the first 500 consecutive patients recruited (from October 1, 2017, to December 2, 2018). Danderyd University Hospital serves a population of approximately 600,000, and the ED had approximately 100,000 annual visits at the time of the study. The cardiology unit manages about 20% of the acute visits with about two-thirds walk-in patients and one-third patients arriving by ambulance. The average time spent in our ED for patients with a chief complaint of chest pain and RETTS level orange, yellow, green, or blue is 4 hours and

10 minutes, based on all 6920 visits from January through November 2018.

Interventions

Self-reported CHT was conducted with the CLEOS software. CLEOS has been described in detail elsewhere [18,23,25]. In brief, the patient interacts with CLEOS on tablets (iPad, Apple Inc) by answering sets of questions for different medical modules starting with the patient's chief complaint. Questions are mainly given in structured text format, such as yes/no or multiple-choice questions (one or many answers possible), but many questions display images, for instance asking the patient to click on an image of the upper body to indicate where pain is located (Figure 1).

Figure 1. User interface of the Clinical Expert Operating System software with examples of a clickable image and multiple-choice question.



The CLEOS software collects an in-depth history including demographics, present illness, organ systems review, medical history, prescription and over-the-counter medications, socioeconomic status, lifestyle, and family history. First, the software collects information on demographics and then reviews the major medical modules (Table 1). As this study specifically concerns chest pain management, questions regarding established risk factors for ACS were asked in the very first part of the interview. The interview is individually tailored by the software, where each question is determined on the basis of prior questions and a set of rules that interpret the clinical significance of prior answers. In total, the software has >17,000 decision nodes and can collect >40,000 data elements. The interview can be paused at any time when needed (eg, for physician encounter, lab test, or diagnostic imaging) and can be resumed whenever the patient has the opportunity.

Table 1. Consecutive order of medical modules in the interview.

Order	Module
1.	Chief complaint
2.	Cardiovascular
3.	Respiratory
4.	Immunology/rheumatology
5.	Endocrinology
6.	Gastroenterology/gastrointestinal surgery
7.	Hepatology
8.	Nephrology and urology
9.	Obstetrics and gynecology
10.	Neurology
11.	Hematology/oncology
12.	Mental health
13.	History of medical/surgical events
14.	Family history

Data Collection

All patients presenting to the ED with a suspected cardiac condition were triaged by a cardiology consultant or senior resident (office hours) or by a trained nurse (out-of-office hours) using the triage protocol RETTS, where a targeted medical history is included. For chest pain patients, ECG and biomarkers were collected before admission to the cardiology unit or to the inpatient day-care unit. If further workup was not indicated, the patients could also be sent home directly from the triage. Less than 0.5% of the patients were sent home directly after triage. If there were signs of ST-elevation myocardial infarction on ECG or if the patient was clinically unstable, the patient was immediately admitted and not included in the study. For patients with a nondiagnostic first ECG, the physician in the cardiology unit or the inpatient day-care unit conducted a more thorough examination and standard history taking. For risk stratification, a combination of a modified HEART score [26], high-sensitivity cardiac troponin assays [10], and the 0/1 hour rule-in and rule-out algorithm [9] is recommended, according to regional guidelines. In the original HEART score, the History component is based on the physician's subjective assessment. In this study, as well as recommended by regional guidelines, the traditional clinical classification of suspected anginal symptoms was used, that is, (1) central chest pain, (2) precipitated by physical or emotional exertion, and (3) relieved by rest or nitrates [27]. Depending on the number of characteristics met, the history was classified as highly (three characteristics met), moderately (two characteristics met), or slightly suspicious (none or one characteristic met) for angina pectoris.

The patients were offered the choice to participate in the CLEOS-CPDS study by a member of the research staff. Standardized oral and written information regarding the study was given, and the patients were given opportunity to ask questions before giving their informed consent by signing a consent form, all according to the procedures approved by the appropriate ethical committee (Multimedia Appendix 1). To ensure that the patient could navigate the CHT software, the research staff supervised the patient as the first page on demographics was answered. If the patient could not navigate the CHT software, the patient was not included. CHT was only performed during waiting times and did not interfere with routine work flow or care in the ED. CHT could occur before, after, or both before and after being seen by a physician, and the staff at the ED was not aware of the information collected by CLEOS at any time. The interview was discontinued either when it was fully completed, if the patient chose to stop for any reason, or if the patient was discharged from the ED or admitted for in-hospital evaluation and treatment. Reasons for not including patients who were considered eligible, the cause for noninclusion, and the cause of discontinuation were registered by the research staff.

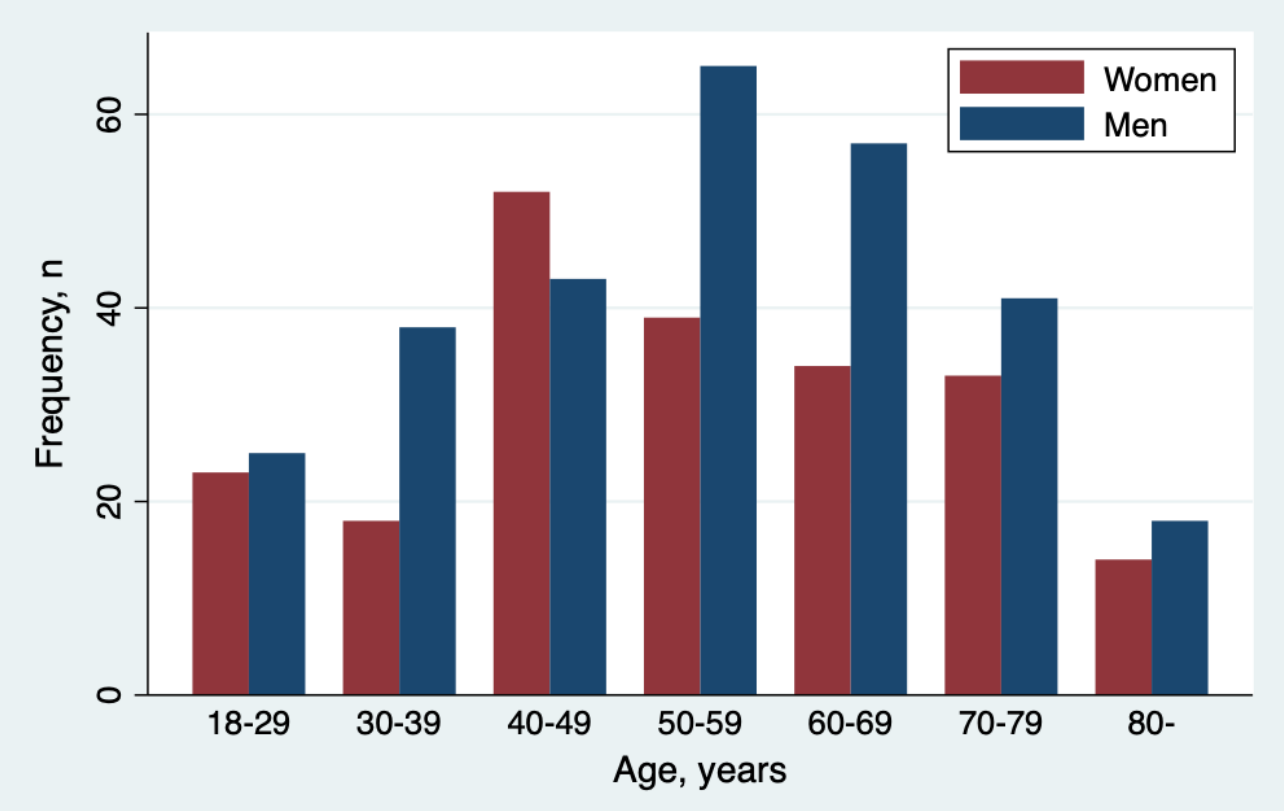
Self-reported descriptive data, medical history data and demographics, and time stamps for completion of each question were collected from the CHT software. All pauses lasting more than 2 minutes were assumed to be interruptions in the interview and excluded when calculating durations. Data on arrival type, arrival time, and admission for the study population were extracted manually by research staff from the electronic health record (TakeCare, CompuGroup Medical Sweden AB). Demographic data and time spent in the ED for the general ED chest pain population during the study period were collected with QlikView, Version 12.10 (QlikTech International AB).

Statistical Analysis

Study outcomes were (1) representativeness of the study population actually included (ie, age and gender of the study participants as compared with the general ED chest pain population), (2) the extent of interview completeness, overall and with regard to demographics, (3) the duration of interview segments, including completed modules, completed interview, and pauses, and (4) effectiveness, assessed as the proportion of patients completing the CHT interview to collect medical history sufficient for cardiovascular risk stratification with the established HEART score.

Descriptive statistics (mean values and standard deviations, median values and IQRs, or proportions, as appropriate) were used for patients' baseline characteristics and to summarize completion and duration of key modules, HEART score data, and completed interview. Pearson χ^2 tests (with 2 degrees of freedom) were used to compare the extent of completeness for binary variables. Wilcoxon rank sum tests were used to compare the median duration for completing the modules. Patients were stratified into 7 age groups (Figure 2), and time of arrival was grouped as morning (7 AM–noon), afternoon (noon–5 PM), evening (5 PM–10 PM), and night (10 PM–7 AM). To test for differences in completion for categorical variables with more than two groups (age groups, occupational status, and time of arrival), the Kruskal-Wallis test with a Dunn pairwise comparison with Bonferroni adjustments as post hoc analysis was used. All statistical analyses were performed using Stata, release 14 (StataCorp).

Figure 2. Age and sex distribution of the study population (n=500). Total number of patients in each age group—18-29: 48, 30-39: 56, 40-49: 95, 50-59: 104, 60-69: 91, 70-79: 74, 80 years or older: 32.



Results

Study Population and Their Characteristics

A total of 9532 patients presented at the ED with a chief complaint of chest pain during the entire study period. During the periods with research staff on duty (ie, when active inclusion was performed), 500 patients who met all inclusion criteria but no exclusion criteria were consecutively enrolled in the study.

The study population (Table 2) had a similar age and sex distribution (mean 54.3, SD 17.0 years, and 213/500, 42.6% women, respectively) as compared with the general chest pain patient population (mean 57.5, SD 19.2 years, and 49.6% women). For patients who were considered eligible for the study but were not included, the most common causes were that they had language issues (182, 18.2%, mostly nonfluent in Swedish), that they felt too tired (158, 15.8%), or that they were unable to use a tablet (152, 15.2%) (Multimedia Appendix 2).

Table 2. Patient baseline characteristics (self-reported).[a]

Characteristic	Value	Responses
Age (years), mean (SD)	54.3 (16.7)	500
Women, n (%)	213 (42.6)	500
Body mass index (kg/m^2), mean (SD)	26.4 (4.4)	500
Diabetes mellitus type 1 or 2, n (%)	27 (6.6)	412
Intake of lipid-lowering medication, n (%)	69 (19.6)	352
Hypertension, n (%)	163 (40.1)	406
Family history of coronary artery disease, n (%)	231 (57.0)	405
Known coronary artery disease, n (%)	74 (16.9)	437
Angina pectoris	45 (10.3)	437
History of myocardial infarction	47 (10.8)	437
History of CABG[b] or PCI[c]	53 (12.3)	432
No cardiovascular disease or diabetes, n (%)	322 (78.2)	412
Current smoker, n (%)	29 (7.0)	414
Previous smoker, n (%)	160 (38.6)	414
Region of birth, n (%)		500
Nordic countries	415 (83.0)	
Europe (outside the Nordic countries)	22 (4.4)	
Outside Europe	63 (12.6)	
Occupational status, n (%)		500
Active worker (employed, student)	320 (64.0)	
Not at work (unemployed, on sick leave)	38 (7.6)	
Retired	142 (28.4)	
Arrived at ED[d] by ambulance, n (%)	92 (19.9)	463
Arrival time, n (%)		495
Morning (7 AM–noon)	244 (49.3)	
Afternoon (noon–5 PM)	180 (36.4)	
Evening (5 PM–10 PM)	47 (9.5)	
Night (10 PM–7 AM)	24 (4.8)	
Reporting any ongoing chest discomfort/pain, n (%)	264 (58.5)	451
Admitted (to the ward or day-care unit), n (%)	225 (46.0)	489

[a]Data from 500 patients are presented as mean values (SD) or n (%), as appropriate.

[b]CABG: coronary artery bypass grafting.

[c]PCI: percutaneous coronary intervention.

[d]ED: emergency department.

Age and sex distributions of the study population are presented in Figure 2. Self-reported patient characteristics and ED data are presented in Table 2. Mean age was about the same for women as for men (53.8 vs 54.6 years, respectively). Self-reported key clinical characteristics included known coronary artery disease (ie, angina pectoris, history of myocardial infarction, or history of coronary artery bypass grafting or percutaneous coronary intervention) in 16.9% (74/437), diabetes mellitus in 6.6% (27/412), hypertension in 40.1% (163/406), and lipid-lowering medication in 19.6% (69/352) (Table 2).

About one-fifth of the participants arrived by ambulance. Nearly half of the population presented to the ED during the morning (7 AM–noon) and about one-third in the afternoon (noon–5 PM). A majority of the patients (264/451, 58.5%) reported ongoing chest pain. Nearly half of the participants (225/489, 46.0%) were eventually admitted, either to a ward or an inpatient day-care unit (Table 2).

Extent of Completeness

The number of participants who carried on with the interview decreased during the course of the interview (Table 3; Multimedia Appendix 3). Sufficient data to calculate HEART score (ie, clinical presentation and risk factors derived from a complete chief complaint module and the initial part of the cardiovascular module) were collected in 352 (70.4%) of the patients (Table 3). Of the 500 participants, the chief complaint (CC) module was completed by 408 (81.6%), the cardiovascular (CV) module by 339 (67.8%), and the Respiratory module by 291 (58.2%), while 148 (29.6%) completed the entire interview (Figure 3 and Table 3). Men completed the CV module and provided sufficient data to calculate the HEART score to a slightly greater extent than women (71.4% vs 62.9%, *P*=.04; 73.9% vs 65.7%, *P*=.049).

Table 3. Summary table for completion and duration of key modules, HEART score, and completed interview.[a,b]

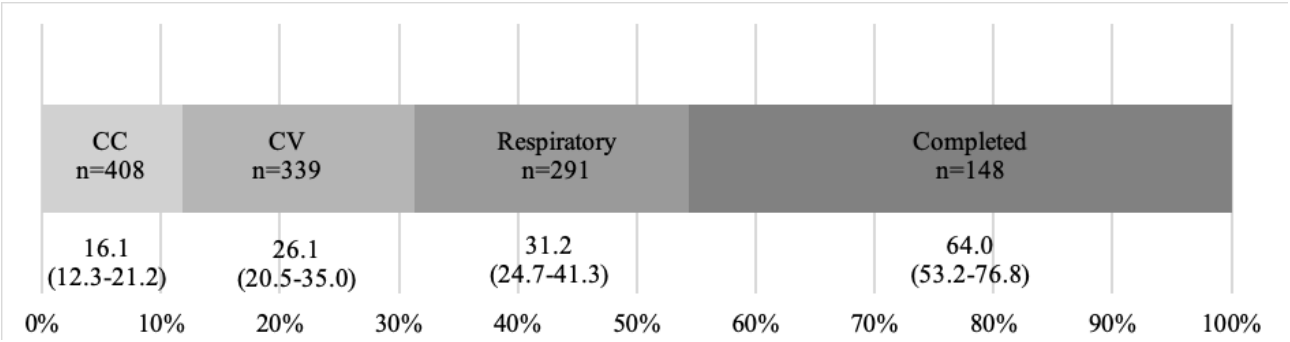
Characteristic by sex	Chief complaint	Cardiovascular	Respiratory	Completed	HEART[c] score
Women (n=213)					
Completers, n (%)	170 (79.8)	134 (62.9)	117 (54.9)	54 (25.4)	140 (65.7)
Duration (min), median (IQR)	15 (12-20)	24 (20-35)	30 (24-39)	66 (53-75)	21 (18-30)
Duration (min), range	1-42	12-82	15-93	41-182	2-66
Men (n=287)					
Completers, n (%)	238 (82.9)	205 (71.4)	174 (60.6)	94 (32.8)	212 (73.9)
Duration (min), median (IQR)	17 (13-23)	27 (21-35)	32 (25-42)	63 (54-78)	24 (19-32)
Duration (min), range	1-77	11-88	14-97	32-121	2-85
All (n=500)					
Completers, n (%)	408 (81.6)	339 (67.8)	291 (58.2)	148 (29.6)	352 (70.4)
P value (completion)	.37	.04	.18	.07	.049
Duration (min), median (IQR)	16 (12-21)	26 (21-35)	31 (25-41)	64 (53-77)	23 (18-31)
Duration (min), range	1-77	11-88	14-97	32-182	2-85
P value (duration)	.12	.15	.43	.25	.08

[a]Number (n) and proportions (%) of participants presented for completed modules, completed interview, and data sufficient for calculating HEART score before discontinuing the interview.

[b]No significant difference between sexes for duration, but significant difference for completion of cardiovascular module and complete HEART score.

[c]HEART: history, electrocardiogram, age, risk factors, and troponin.

Figure 3. Median durations with IQRs (in minutes) for completed modules and complete interview, excluding pauses >2 minutes. CC: chief complaint module; Completed: completed interview; CV: cardiovascular module; Respiratory: respiratory module.



The proportion of the participants who completed the key modules in this context (CC, CV, and Respiratory modules) and the complete interview was lower in the age groups 70-79 years and 80 years or older, as compared with younger age groups (*P*s<.001 for all key modules and completed interview; Figure 4). No other significant differences in rates of completion were found between the age groups. Active workers completed the key modules and the complete interview more often than retired participants (CC: *P*=.004, CV: *P*=.002, Respiratory: *P*<.001, and complete interview: *P*<.001) (Figure 5). Patients not at work completed the Respiratory module to a slightly greater degree than retired participants (*P*=.03). No other significant differences were found by occupational status.

Figure 4. Fractions (with 95% CIs) of completed key modules and completed interviews, stratified by age.

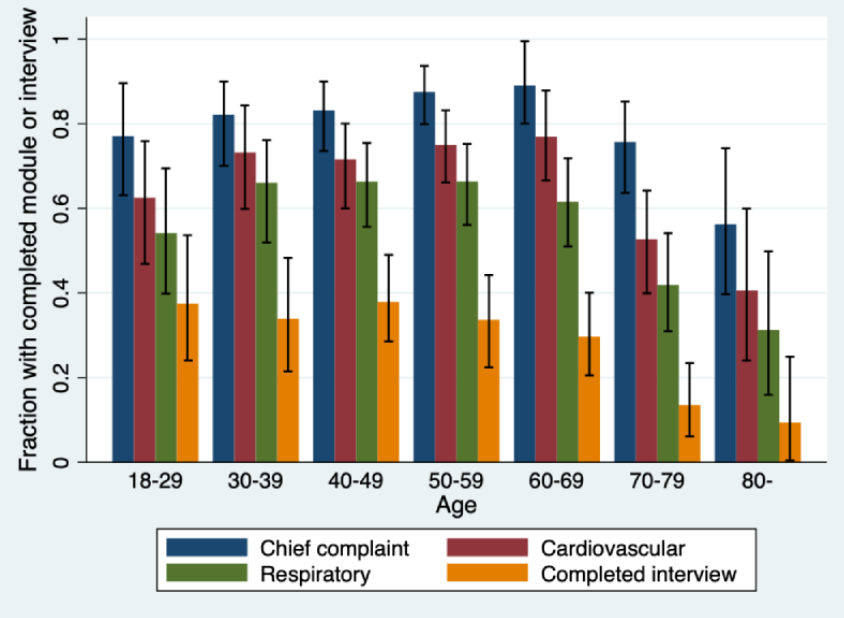
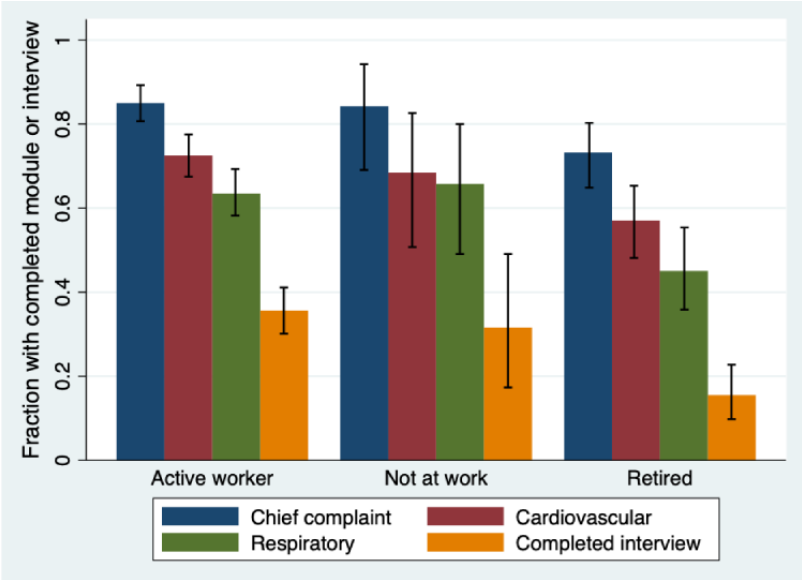
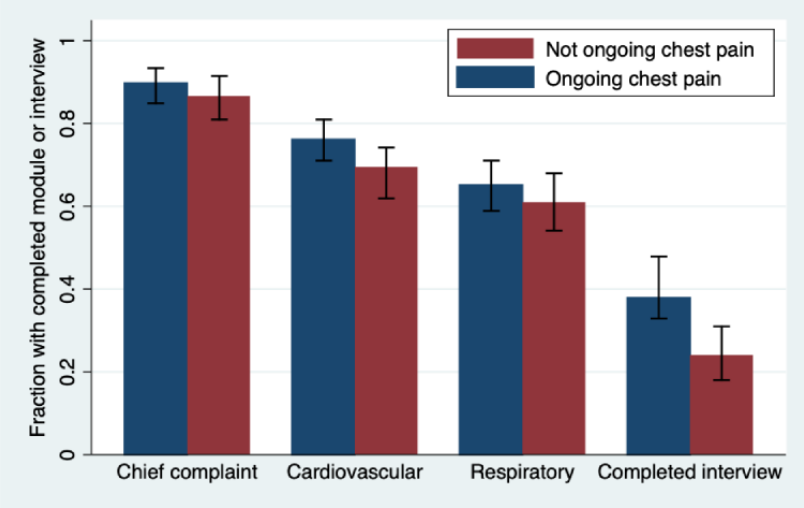


Figure 5. Fractions (with 95% CIs) of completed key modules and completed interviews, stratified by occupational status (320 active workers, 38 not at work, 142 retired). Active worker: employed or student; not at work: unemployed or on sick leave.



Patients arriving by ambulance completed the CC and CV modules to a slightly lesser extent compared with those not arriving by ambulance (79% vs 88%, *P*=.03 and 63% vs 74%, *P*=.045, respectively). Participants who reported ongoing chest pain completed the interview to a greater extent than those not reporting ongoing chest pain (38% vs 24%, *P*=.002; Figure 6). No other significant differences in the completion of modules, for example, relation to the time of day the participant presented to the ED or by admission (hospital admission vs discharged home), were found.

Figure 6. Fractions (with 95% CIs) of completed key modules and completed interviews, stratified by ongoing chest pain or not (264 with ongoing chest pain and 187 without).



Duration of CHT Session

The median duration (excluding pauses longer than 2 minutes) to collect HEART score data was 23 (IQR 18-31) minutes, to complete the CC module 16 (IQR 12-21) minutes, to complete the CV module 26 (IQR 20-35) minutes, to complete the Respiratory module 31 (IQR 25-41) minutes, and to complete an entire interview 64 (IQR 53-77) minutes (Figure 3 and Table 3). No difference for duration by sex was found. The number and proportions of patients who ended the interview within a certain time and mean pauses stratified by interview duration are presented as supplementary material (Multimedia Appendices 4 and 5).

In the group of 352 participants who did not complete the full interview, the main reasons for discontinuing were discharge from ED (101, 28.7%) and that the participant felt tired (95, 27.0%) (Table 4). When comparing the age groups 18-69 years (n=259) and ≥70 years (n=93) to participants of all ages, discharge from ED was more frequent in the first group (28.7% vs 12.9%, *P*<.001), and difficulty using the tablet was reported more often in the second group (4.5% vs 14.0%, *P*<.001) (Table 4).

Table 4. Reasons for discontinuing the interview.[a]

Reasons	Age groups (years), n (%)			*P* value
	All	18-69	≥70	
Discharge from ED[b]	101 (28.7)	89 (34.4)	12 (12.9)	<.001
Tired	95 (27.0)	69 (26.6)	26 (28.0)	.81
Missing/not stated	61 (17.3)	48 (18.5)	13 (14.0)	.32
Admission/transfer	22 (6.3)	13 (5.0)	9 (9.7)	.11
Difficulty to use tablet	16 (4.5)	3 (1.2)	13 (14.0)	<.001
Clinical examination	15 (4.3)	11 (4.2)	4 (4.3)	.98
Technical issues	11 (3.1)	8 (3.1)	3 (3.2)	.94
End of research staff work shift	10 (2.8)	6 (2.3)	4 (4.3)	.32
Perceived not relevant/too many questions	9 (2.6)	5 (1.9)	4 (4.3)	.21
Acute medical condition/measure	7 (2.0)	3 (1.2)	4 (4.3)	.06
Other	5 (1.4)	4 (1.5)	1 (1.1)	.75

[a]Number (n) and proportions (%) of all 352 participants who did not complete the full interview, according to age group.

[b]ED: emergency department.

Discussion

Principal Results

Although the utility of CHT has been studied in primary care settings [28,29] and general acute settings [20,21], this appears to be one of the first studies of CHT in an acute cardiology setting. We show that a majority (70.4%) of acute chest pain patients can interact with CHT to collect medical history adequately to provide a HEART score for chest pain management. Given the large proportion of people presenting to the ED with chest pain, our results suggest that CHT could potentially contribute to safer management with improved risk

stratification in this patient group, particularly during periods with high workload and crowding, which are associated with worse outcomes [30]. This could eventually reduce unnecessary, expensive, and potentially risky examinations. However, our study only shows the utility of this specific strategy for patient interview. Further studies are needed to validate the information provided by the CLEOS CHT program against information in the electronic health record obtained by an interview performed by a physician and to evaluate the results to prospective outcome data.

The interview was arranged so that the most important factors for assessing chest pain were asked in the very first part, in order to collect medical history sufficient for cardiovascular risk stratification in the ED setting. Subsequently, data was collected for all organ systems with, in the CLEOS developers' opinion, lesser significance for the assessment of chest pain the longer the interview went on. As expected, the proportion of patients who continued with the interview decreased the longer it went on. More importantly, however, the median duration for sufficient data to calculate the HEART score was only 23 minutes. This is comparable to the reported time for taking a standard history in an acute setting [31,32]. However, the CHT can make use of the waiting time in the ED and provide the patient with time to think through their answers more carefully, which could add to more reliable answers and improved diagnostic results.

Premature discontinuations were mainly due to patients being sent home or getting too tired to continue the interview. Only 30% went through a complete interview, with a median duration of 64 minutes, which is longer than a standard interview by a physician in an ED setting [31,32]. However, it is important to consider the context and the intention of data collection when assessing the duration of the interview. Reaching a fully completed interview is of importance to identify an unclear diagnosis or for research purposes, but for risk stratification in acute chest pain patients, it is more important to rapidly populate the elements included in an established risk score. Of note, in the current study of patients with mostly low-intermediate risk (ie, RETTS level orange, yellow, green, and blue), median time spent in the ED was about 4 hours. Thus, the somewhat longer time for the CHT would not prolong the time spent in the ED, as compared to using a standard history taking by the attending physician. Nevertheless, it would be of benefit in future development of CHT if the extent of an interview could be adjusted to the context of the visit and medical urgency.

The extent of completion was lower in the age group 70 years and older, a finding not previously observed in the few studies available [15,17,18,20]. Our data suggests that this was due to difficulties using tablets and to somewhat more frequent hospitalizations than in the younger (18-64 years) age group (Table 4). Patients reporting ongoing chest pain also completed the interview to a slightly greater extent, which raises the question of whether one is more inclined to complete the interview due to concerns about a present complaint. As well, the group of patients arriving by ambulance completed two key modules of the interview (CC and CV) to a lesser extent than walk-in patients, possibly due to a larger proportion of older people in this group.

There are some strengths of this study. First, this is a large sample of patients from a study population representative of a general chest pain population, CLEOS-CPDS, with a prospective cohort study design and a published study protocol. Second, a generic layout of the CHT software may allow the results to be generalized to other complaints and care settings. Finally, this is an academically initiated and driven study where the CHT software is owned by a public university. There are no commercial interests within the research project.

There are also important limitations to this study. First, although patients were recruited consecutively, this occurred mainly during office hours and evenings (due to research staff working hours) and when a sufficient number of tablets was available. This entails a risk of selection bias, and the results may not apply to patients presenting to the ED at other times of the day. However, the proportion of chest pain patients arriving at night was small, and the demographics were similar to the total ED chest pain population. Second, there may also be confounding by a selection of patients with good tablet skills. This potential confounding warrant further study. As well, a number of patients were not eligible; patients with language difficulties or inability to carry out CHT on the tablet were not included. These groups accounted for 18.2% and 15.8% of patients who were asked to participate but did not. However, this compares to results found by others, where no complete basic medical history could be obtained in the ED setting for 25% of the patients [31]. Thus, it is important in future CHT implementations to identify these patient groups and their characteristics, so they can be offered standard history taking. In addition, developing simpler and more user-friendly software in these patient groups is needed. Third, patients sent home directly from the physician triage (potentially healthier and younger) were not eligible for the study. Many EDs do not have physician triage, and the assessment may vary between physicians [33]. This may affect the generalizability of the study. However, a negligible fraction of patients were sent home directly from the medical triage, and we consider it unlikely to have affected the study results. Finally, our results may not be applied for critically ill patients, who should be seen by a physician immediately according to the RETTS triage protocol, as the CHT was not intended for use in these patients.

Conclusions

A majority of acute chest pain patients can interact effectively with CHT on a tablet in the ED to provide sufficient data for risk stratification with a well-established risk score (ie, HEART score). The utility was somewhat lower in patients 70 years and older, in patients arriving by ambulance, and in patients without ongoing chest pain. Further studies are warranted to assess whether CHT can contribute to improved management and better risk stratification for one of the most common chief complaints in the ED.

Acknowledgments

This work was funded by the Robert Bosch Stiftung (Stuttgart, Germany), grant number 11.5.1000.0258.0; Region Stockholm (ALF project; Stockholm, Sweden), grant number 20190593; Karolinska Institutet Research Foundation (Stockholm, Sweden); and Stiftelsen Hjärtat (Stockholm, Sweden).

Authors' Contributions

All authors contributed to the conception and design of the study. HB and TK drafted the manuscript. All authors revised the manuscript for intellectual content and approved the final text. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

Conflicts of Interest

DZ is the inventor on US patents for technology related to the CLEOS program. All patent rights and copyrights to technology, language, images, and knowledge content are assigned without royalty rights by DZ to Karolinska Institutet, Stockholm, Sweden, which is a public university. Apart from Karolinska Institutet and its subsidiaries, no individuals or companies may be owners or receive royalties or other revenue from use of CLEOS technology, language, images, or knowledge content or from clinical insights or computer algorithms generated from analysis of data acquired by the program. JS is speaker honoraria and on advisory board for AstraZeneca, Bayer, NovoNordisk.

Multimedia Appendix 1
Patient information sheet and consent form according to the standards of GCP-ICH applied in Sweden. These documents are reviewed and approved by the Stockholm Regional Ethical Committee (reference number 2015/1955-1).
[DOC File , 54 KB - jmir_v23i4e25493_app1.doc]

Multimedia Appendix 2
Reasons for not participating in the study (n=1000).
[DOCX File , 18 KB - jmir_v23i4e25493_app2.docx]

Multimedia Appendix 3
Number and proportion of participants who ended the interview within a certain time.
[DOCX File , 24 KB - jmir_v23i4e25493_app3.docx]

Multimedia Appendix 4
Fractions of completed modules for all participants. Only females could answer the Obstetrics/Gynecology module. Gastro: gastrointestinal.
[PNG File , 113 KB - jmir_v23i4e25493_app4.png]

Multimedia Appendix 5
Interview duration and number of pauses >2 minutes with mean pause duration stratified by interview duration.
[DOCX File , 18 KB - jmir_v23i4e25493_app5.docx]

References

1. Mockel M, Searle J, Muller R, Slagman A, Storchmann H, Oestereich P, et al. Chief complaints in medical emergencies: do they relate to underlying disease and outcome? The Charité Emergency Medicine Study (CHARITEM). Eur J Emerg Med 2013 Apr;20(2):103-108. [doi: 10.1097/MEJ.0b013e328351e609] [Medline: 22387754]
2. Rui P, Kang K. National Hospital Ambulatory Medical Care Survey: 2017 emergency department summary tables. URL: https://www.cdc.gov/nchs/data/nhamcs/web_tables/2017_ed_web_tables-508.pdf [accessed 2020-09-18]
3. Amsterdam EA, Wenger NK, Brindis RG, Casey DE, Ganiats TG, Holmes DR, ACC/AHA Task Force Members, Society for Cardiovascular AngiographyInterventionsthe Society of Thoracic Surgeons. 2014 AHA/ACC guideline for the management of patients with non-ST-elevation acute coronary syndromes: executive summary: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. Circulation 2014 Dec 23;130(25):2354-2394. [doi: 10.1161/CIR.0000000000000133] [Medline: 25249586]
4. Collet J, Thiele H, Barbato E, Barthélémy O, Bauersachs J, Bhatt DL, ESC Scientific Document Group. 2020 ESC Guidelines for the management of acute coronary syndromes in patients presenting without persistent ST-segment elevation. Eur Heart J 2021 Apr 07;42(14):1289-1367. [doi: 10.1093/eurheartj/ehaa575] [Medline: 32860058]

5. Pappas Y, Anandan C, Liu J, Car J, Sheikh A, Majeed A. Computer-assisted history-taking systems (CAHTS) in health care: benefits, risks and potential for further development. Inform Prim Care 2011;19(3):155-160 [FREE Full text] [doi: 10.14236/jhi.v19i3.808] [Medline: 22688224]
6. Antman EM, Cohen M, Bernink PJ, McCabe CH, Horacek T, Papuchis G, et al. The TIMI risk score for unstable angina/non-ST elevation MI: A method for prognostication and therapeutic decision making. JAMA 2000 Aug 16;284(7):835-842. [doi: 10.1001/jama.284.7.835] [Medline: 10938172]
7. Six AJ, Backus BE, Kelder JC. Chest pain in the emergency room: value of the HEART score. Neth Heart J 2008 Jun;16(6):191-196 [FREE Full text] [Medline: 18665203]
8. Fox KAA, Dabbous OH, Goldberg RJ, Pieper KS, Eagle KA, Van de Werf F, et al. Prediction of risk of death and myocardial infarction in the six months after presentation with acute coronary syndrome: prospective multinational observational study (GRACE). BMJ 2006 Nov 25;333(7578):1091 [FREE Full text] [doi: 10.1136/bmj.38985.646481.55] [Medline: 17032691]
9. Reichlin T, Schindler C, Drexler B, Twerenbold R, Reiter M, Zellweger C, et al. One-hour rule-out and rule-in of acute myocardial infarction using high-sensitivity cardiac troponin T. Arch Intern Med 2012 Sep 10;172(16):1211-1218. [doi: 10.1001/archinternmed.2012.3698] [Medline: 22892889]
10. Body R, Mueller C, Giannitsis E, Christ M, Ordonez-Llanos J, de Filippi CR, TRAPID-AMI Investigators. The Use of Very Low Concentrations of High-sensitivity Troponin T to Rule Out Acute Myocardial Infarction Using a Single Blood Test. Acad Emerg Med 2016 Sep;23(9):1004-1013 [FREE Full text] [doi: 10.1111/acem.13012] [Medline: 27178492]
11. Mueller C, Giannitsis E, Christ M, Ordóñez-Llanos J, deFilippi C, McCord J, TRAPID-AMI Investigators. Multicenter Evaluation of a 0-Hour/1-Hour Algorithm in the Diagnosis of Myocardial Infarction With High-Sensitivity Cardiac Troponin T. Ann Emerg Med 2016 Jul;68(1):76-87.e4 [FREE Full text] [doi: 10.1016/j.annemergmed.2015.11.013] [Medline: 26794254]
12. Brodman K, Erdmann AJ, Lorge I, Wolff HG, Broadbent TH. The Cornell medical index; a adjunct to medical interview. J Am Med Assoc 1949 Jun 11;140(6):530-534. [doi: 10.1001/jama.1949.02900410026007] [Medline: 18144531]
13. Slack WV, Hicks GP, Reed CE, Van Cura LJ. A computer-based medical-history system. N Engl J Med 1966 Jan 27;274(4):194-198. [doi: 10.1056/NEJM196601272740406] [Medline: 5902618]
14. Zakim D. Development and significance of automated history-taking software for clinical medicine, clinical research and basic medical science. J Intern Med 2016 Sep;280(3):287-299 [FREE Full text] [doi: 10.1111/joim.12509] [Medline: 27071980]
15. Wolford G, Rosenberg SD, Rosenberg HJ, Swartz MS, Butterfield MI, Swanson JW, et al. A clinical trial comparing interviewer and computer-assisted assessment among clients with severe mental illness. Psychiatr Serv 2008 Jul;59(7):769-775. [doi: 10.1176/ps.2008.59.7.769] [Medline: 18586994]
16. Taylor JL, Mistry M. An evaluation of automated, internet-based psychiatric history taking. Future Healthc J 2018 Jun;5(2):126-131 [FREE Full text] [doi: 10.7861/futurehosp.5-2-126] [Medline: 31098547]
17. Almario CV, Chey W, Kaung A, Whitman C, Fuller G, Reid M, et al. Computer-generated vs. physician-documented history of present illness (HPI): results of a blinded comparison. Am J Gastroenterol 2015 Jan;110(1):170-179 [FREE Full text] [doi: 10.1038/ajg.2014.356] [Medline: 25461620]
18. Zakim D, Fritz C, Braun N, Fritz P, Alscher MD. Computerized history-taking as a tool to manage dyslipidemia. Vasc Health Risk Manag 2010 Nov 12;6:1039-1046 [FREE Full text] [doi: 10.2147/VHRM.S14302] [Medline: 21127700]
19. Jimenez G, Tyagi S, Osman T, Spinazze P, van der Kleij R, Chavannes NH, et al. Improving the Primary Care Consultation for Diabetes and Depression Through Digital Medical Interview Assistant Systems: Narrative Review. J Med Internet Res 2020 Aug 28;22(8):e18109 [FREE Full text] [doi: 10.2196/18109] [Medline: 32663144]
20. Benaroia M, Elinson R, Zarnke K. Patient-directed intelligent and interactive computer medical history-gathering systems: a utility and feasibility study in the emergency department. Int J Med Inform 2007 Apr;76(4):283-288. [doi: 10.1016/j.ijmedinf.2006.01.006] [Medline: 16473548]
21. Arora S, Goldberg AD, Menchine M. Patient impression and satisfaction of a self-administered, automated medical history-taking device in the Emergency Department. West J Emerg Med 2014 Feb;15(1):35-40 [FREE Full text] [doi: 10.5811/westjem.2013.2.11498] [Medline: 24695871]
22. Pappas Y, Všetečková J, Poduval S, Tseng PC, Car J. Computer-Assisted versus Oral-and-Written History Taking for the Prevention and Management of Cardiovascular Disease: a Systematic Review of the Literature. Acta Medica (Hradec Kralove) 2017;60(3):97-107 [FREE Full text] [doi: 10.14712/18059694.2018.1] [Medline: 29439755]
23. Brandberg H, Kahan T, Spaak J, Sundberg K, Koch S, Adeli A, et al. A prospective cohort study of self-reported computerised medical history taking for acute chest pain: protocol of the CLEOS-Chest Pain Danderyd Study (CLEOS-CPDS). BMJ Open 2020 Jan 21;10(1):e031871 [FREE Full text] [doi: 10.1136/bmjopen-2019-031871] [Medline: 31969363]
24. Widgren BR, Jourak M. Medical Emergency Triage and Treatment System (METTS): a new protocol in primary triage and secondary priority decision in emergency medicine. J Emerg Med 2011 Jun;40(6):623-628. [doi: 10.1016/j.jemermed.2008.04.003] [Medline: 18930373]
25. Zakim D, Braun N, Fritz P, Alscher MD. Underutilization of information and knowledge in everyday medical practice: evaluation of a computer-based solution. BMC Med Inform Decis Mak 2008 Nov 05;8:50 [FREE Full text] [doi: 10.1186/1472-6947-8-50] [Medline: 18983684]

26. Six AJ, Cullen L, Backus BE, Greenslade J, Parsonage W, Aldous S, et al. The HEART score for the assessment of patients with chest pain in the emergency department: a multinational validation study. Crit Pathw Cardiol 2013 Sep;12(3):121-126. [doi: 10.1097/HPC.0b013e31828b327e] [Medline: 23892941]
27. Knuuti J, Wijns W, Saraste A, Capodanno D, Barbato E, Funck-Brentano C, ESC Scientific Document Group. 2019 ESC Guidelines for the diagnosis and management of chronic coronary syndromes. Eur Heart J 2020 Jan 14;41(3):407-477. [doi: 10.1093/eurheartj/ehz425] [Medline: 31504439]
28. Hess R, Santucci A, McTigue K, Fischer G, Kapoor W. Patient difficulty using tablet computers to screen in primary care. J Gen Intern Med 2008 Apr;23(4):476-480 [FREE Full text] [doi: 10.1007/s11606-007-0500-1] [Medline: 18373148]
29. Slack WV, Kowaloff HB, Davis RB, Delbanco T, Locke SE, Safran C, et al. Evaluation of computer-based medical histories taken by patients at home. J Am Med Inform Assoc 2012;19(4):545-548 [FREE Full text] [doi: 10.1136/amiajnl-2011-000580] [Medline: 22237866]
30. Af Ugglas B, Djärv T, Ljungman PLS, Holzmann MJ. Emergency department crowding associated with increased 30-day mortality: a cohort study in Stockholm Region, Sweden, 2012 to 2016. J Am Coll Emerg Physicians Open 2020 Dec;1(6):1312-1319 [FREE Full text] [doi: 10.1002/emp2.12243] [Medline: 33392538]
31. Lindner T, Slagman A, Senkin A, Möckel M, Searle J. Medical History of Elderly Patients in the Emergency Setting: Not an Easy Point-of-Care Diagnostic Marker. Emerg Med Int 2015;2015:490947 [FREE Full text] [doi: 10.1155/2015/490947] [Medline: 26421190]
32. Sabin J, Khan W, Subbe CP, Franklin M, Abulela I, Khan A, et al. 'The time it takes…' How doctors spend their time admitting a patient during the acute medical take. Clin Med (Lond) 2016 Aug;16(4):320-324 [FREE Full text] [doi: 10.7861/clinmedicine.16-4-320] [Medline: 27481373]
33. Li C, Syue Y, Kung C, Hung S, Lee C, Wu K. Seniority of Emergency Physician, Patient Disposition and Outcome Following Disposition. Am J Med Sci 2016 Jun;351(6):582-588. [doi: 10.1016/j.amjms.2016.02.041] [Medline: 27238920]

Abbreviations

ACS: acute coronary syndrome
CC: chief complaint
CHT: computerized history taking
CLEOS: Clinical Expert Operating System
CLEOS-CPDS: CLEOS Chest Pain Danderyd Study
CV: cardiovascular
ED: emergency department
ECG: electrocardiogram
HEART: history, electrocardiogram, age, risk factors, and troponin
RETTS: Rapid Emergency Triage and Treatment System

Edited by R Kukafka; submitted 12.11.20; peer-reviewed by C Subbe, M Zahmatkeshan, I Mircheva, C Hudak; comments to author 22.12.20; revised version received 12.02.21; accepted 16.03.21; published 27.04.21.

Please cite as:
Brandberg H, Sundberg CJ, Spaak J, Koch S, Zakim D, Kahan T
Use of Self-Reported Computerized Medical History Taking for Acute Chest Pain in the Emergency Department – the Clinical Expert Operating System Chest Pain Danderyd Study (CLEOS-CPDS): Prospective Cohort Study
J Med Internet Res 2021;23(4):e25493
URL: https://www.jmir.org/2021/4/e25493
doi:10.2196/25493
PMID:33904821

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org